
U.S. DEPARTMENT OF ENERGY

# 2022 Genomic Sciences Program (GSP) Annual PI Meeting

# Abstract Book

Sponsored by the
U.S. Department of Energy
Office of Biological and Environmental Research
Biological Systems Science Division

**Department of Energy**
Washington, DC 20585

February 28, 2022

Dear Colleague:

Welcome to the 2022 Genomic Science Program (GSP) and Bioimaging Science Program (BSP) Annual PI Meetings. It is hard to believe that it has been almost two years since the pandemic first struck, confining our annual meetings to a virtual format. While this format has posed many challenges, it has also brought some opportunities. Importantly, it has given us the ability to broaden PI meeting participation and we look forward to welcoming many new attendees via a virtual setting who would otherwise not be able to attend. Despite the challenges of the past 2 years, our programs continue to thrive, and that is entirely due to your continued hard work and resilient efforts to maintain momentum under extraordinary circumstances. On behalf of DOE, thank you all.

As with last year's meeting, we have combined the GSP and BSP PI meetings, reflecting the growing synergy between these two portfolio elements. Appropriately, this year's meeting brings an emphasis on emerging technologies, facilities, and resources that enable research funded by the Biological Systems Science Division (BSSD). There have been profound changes in some of these areas and rapid technological development in others. Who would have thought only a few years ago that it would now be possible to plausibly predict the structures of a great many proteins from primary sequence alone? We are indeed witnessing rapid changes in the broader biological sciences while driving new developments within our mission space and it is very exciting to see.

We are also excited about our lineup of sessions and presentations on topics as far ranging as plant genomics, microbial ecology, synthetic biology, and quantum imaging. Notably, we are honored to have two Nobel Laureates join us this year. Our joint keynote speaker is Dr. Jennifer Doudna (UC-Berkeley and LBNL), who was awarded the 2021 Nobel Prize in Chemistry. Equally exciting, we welcome Dr. Joachim Frank (Columbia University) Nobel Laureate in Chemistry in 2017 as the keynote address at the opening of the BSP meeting. The accomplishments of both scientists are certainly well known throughout these two programs, and it is a privilege to have them join us this year.

BSSD's portfolio continues to evolve and grow. In the past year, we've seen significant developments in our current portfolio through the Bioenergy Research Centers and the DOE Lab programs (SFAs) as well as within our Academic-led portfolio. We will hear from many of these projects at this meeting. Additionally, we have many new projects on display at this meeting from several areas across the two portfolios, including new *in situ* imaging techniques for living plant and microbial systems, new microbial biofuels and bioproducts projects, and a new topic area on microbial upcycling of synthetic polymers (plastics). We have also continued to invest in our early career researcher program and in new 'omics analysis capabilities through several new KBase partnerships. We look forward to exploring the potential incorporation of AI/ML techniques into our research portfolio and what new discoveries these techniques could bring. The broader portfolio is at the nexus of several key and confluent developments in synthetic, computational, microbial, and plant biology that will help to shape a more sustainable bioeconomy and clean energy future. The core mission of BSSD is therefore powerfully aligned with DOE's and the Administration's priorities, and we look forward to a bright future for our programs.

We hope that this meeting will be an opportunity to connect with your fellow researchers as well as with DOE program staff. We are using Gather.town for the poster sessions this year which should be a vast improvement over last year's format and provide increased opportunities for conversations with your colleagues. Your insights and scientific talents are critical to driving the continued success of the GSP and BSP portfolios, and we look forward to having a stimulating three days of scientific exchange and "seeing" you at the meeting.

Sincerely,

Todd Anderson, Ph.D.
Director, Biological Systems Science Division, SC-33.2
Office of Biological and Environmental Research
Office of Science

# Table of Contents

(Click on Title to advance to the abstract)

**Paul Adams**: Investigating Microbial Ecophysiology of Aerobic and Anaerobic Microbes Through the ENIGMA Environmental Atlas
PRESENTER: Brandon Enalls

**Paul Adams**: Characterization of a Nitrate-Respiring Multi-Metal-Resistant Bacillus Species Highly Abundant in Heavily Contaminated ORR FRC Subsurface
PRESENTER: Jennifer Goff

**Paul Adams**: ENIGMA Long Read Sequencing and Assembly for Microbial Genomes: Current Capabilities and Future for Metagenomics and KBase Integration for Assembly
PRESENTER: Lauren Lui

**Paul Adams**: Mobile Genetic Elements In The ORFRC: Native Plasmids And Viral Genomes
PRESENTER: Aindrila Mukhopadhyay

**Paul Adams**: The ENIGMA Subsurface Observatory: A high resolution approach to studying a shallow contaminated groundwater system
PRESENTER: Andrew Putt

**Paul Adams**: From fast to furious phenotyping of microbes: advancing the understanding of microbial diversities and gene functions using droplet microfluidics and high throughput sequencing
PRESENTER: Fangchao Song

**Paul Adams**: Investigating the abiotic control of denitrification processes using synthetic communities and laboratory simulations

PRESENTER: Jacob Valenzuela

**Paul Adams**: Diurnal and Seasonal Fluctuations with the Subsurface: A 17-Week Survey of Groundwater and Sediment in 27 Contaminated Wells
PRESENTER: Katie Walker

**Paul Adams**: High-Throughput Genetic Analysis of Rhodanobacter Reveals Genes Important for Metal Tolerance
PRESENTER: Li Wen Wang

**Paul Adams**: A Field-Scale Omics-Enabled Groundwater Reactive Transport Model
PRESENTER: Yajiao Wang

**Francis Alexander**: Optimal Experimental Design (OED) of Biological Systems
PRESENTER: Puhua Niu

**Andrew Allen**: Genome-wide mapping of cis-regulatory elements and regulation of nitrate assimilation in Phaeodactylum tricornutum
PRESENTER: Andrew Allen

**Andrew Allen**: Unravelling xanthophyll pigment biosynthesis in algae
PRESENTER: Yu Bai

**Andrew Allen**: CRISPR interference (CRISPRi) for Dynamic Regulation of Gene Expression in Diatoms
PRESENTER: Tessema Kassaw

**Andrew Allen**: Modeling carbon metabolism of the diatom Phaeodactylum tricornutum during nitrogen starvation and during high light and low light conditions
PRESENTER: Amy Zheng

**Steven Allison**: Bacterial Community Response to Environmental Change Varies With Depth in the Surface Soil
PRESENTER: Kristin Barbour

**Steven Allison**: Litter Microbial Trait-Based Strategies in Response to Drought
PRESENTER: Brianna Finley

**Ana Alonso**: Identification of Pennycress (Thlaspi arvense L.) Proteins that Influence Lipid Droplet Formation and Modulate Neutral Lipid Accumulation
PRESENTER: Julius Ver Sagun

**Ana Alonso**: Using 13C-labeling to Unravel the Temporal and Spatial Production of Seed Oil in Developing Embryos of Pennycress a Promising Source for Sustainable Aviation Fuel
PRESENTER: Umesh Prasad Yadav

**Daniel Amador-Noguez**: Application of a Novel 2H Isotope Tracer Approach to Characterizing Metabolic Thermodynamics in C. thermocellum
PRESENTER: Melanie Callaghan

**Jean-Michel Ané**: Optimizing Biological Nitrogen Fixation on Sorghum Using Microbial Communities
PRESENTER: Claire Palmer

**Jean-Michel Ané**: A Systems Understanding of Nitrogen-Fixation on the Aerial Roots of Sorghum
PRESENTER: Saddie Vela

**Dionysios Antonopoulos**: Rapid Design and Engineering of Smart and Secure Microbiological Systems
PRESENTER: Dion Antonopoulos

**Dionysios Antonopoulos**: Developing “Smart” Single-strain Systems Capable of Detecting and Responding to Target Organisms in the Environment
PRESENTER: Gyorgy Babnigg

**Dionysios Antonopoulos**: Guiding Data-Driven Integrative Design of Secure Biological Systems with Artificial Intelligence Techniques
PRESENTER: Carla Mann

**Dionysios Antonopoulos**: CRISPR-Act: AI-guided Prediction of a CRISPR Kill-switch Across Physiological Contexts
PRESENTER: Rebecca Weinberg

**Ludmilla Aristilde**: Probing Lignin Deconstruction and Catabolism in Soil Pseudomonas species
PRESENTER: Ludmilla Aristilde

**Adam Arkin**: KBase: A case study illustrating the derivation and testing of mechanistic connections between geochemistry and the microbiome
PRESENTER: Paramvir Dehal

**Adam Arkin**: How KBase Supports Education
PRESENTER: Ellen Dow

**Adam Arkin**: KBase: Significant Improvements to the DOE Systems Biology Knowledgebase in 2021
PRESENTER: Christopher Henry

**Adam Arkin**: Science Focus Areas use KBase to share new tools and data
PRESENTER: Elisha Wood-Charlson

**Jonathan Arnold**: Systems analysis of the beneficial associations of sorghum with arbuscular mycorrhizal fungi studied with genetics genomics imaging and microbiomics
PRESENTER: Jeffrey Bennetzen

**Jose Avalos**: Monochromatic control of bacteria/yeast consortia for fuel and chemical production
PRESENTER: Jose Avalos

## B

**Ivan Baxter**: Synthetic Genetic Circuits to Engineer Water Use Efficiency Photosynthetic Efficiency and Biocontainment
PRESENTER: Jennifer Brophy

**Ivan Baxter**: Engineering enhanced photosynthesis and water use efficiency in Sorgh
PRESENTER: Asaph Cousins

**Ivan Baxter**: Advanced Phenomic Approaches to Improving Water-Use Efficiency in Bioenergy Grasses
PRESENTER: Jose Dinneny

**Ivan Baxter**: Improvement of Sorghum Transformation and Genome Editing for the Development of Stable Lines for the Analysis of Photosynthetic and Water Use Efficiencies
PRESENTER: Albert Kausch

**Ivan Baxter**: Leveraging multi-omics data and pan-genomic sequence variation in Setaria viridis and Sorghum bicolor for candidate gene discovery and nomination for crop improvement.
PRESENTER: Todd Mockler

**Jeff Bennetzen**: Adaptive GWAS using MINE for discovery of structural equations in the AMF-Sorghum Project
PRESENTER: Shufan Zhang

**Mike Betenbaugh**: Metabolic flux analysis of an engineered sucrose-secreting strain of the cyanobacterium Synechococcus elongatus
PRESENTER: Bo Wang

**Jennifer Bhatnagar**: Plant-Mycorrhizal-Decomposer Interactions and Their Impacts on Terrestrial Biogeochemistry
PRESENTER: Nahuel Policelli

**Crysten Blaby**: Understanding poplar and sorghum micronutrient stress by integrating functional genomics with molecular-level experimentation
PRESENTER: Crysten Blaby

**Crysten Blaby**: A New Structural Paradigm in Heme Binding – a Novel Family of Plant Heme Oxidases
PRESENTER: Nicolas Grosjean

**Crysten Blaby**: Cryo-EM Structure of a Zinc Uptake Transporter in the Closed State
PRESENTER: Qun Liu

**Crysten Blaby**: Time-series analysis of phenotypic and transcriptomic responses to nutrient stress in two bioenergy crops
PRESENTER: Timothy Paape

**Crysten Blaby**: Transcriptional regulatory mechanisms linking secondary cell wall biosynthesis and iron homeostasis in Populus
PRESENTER: Meng Xie

**Crysten Blaby**: Characterization of a Novel Zinc Chaperone in Arabidopsis
PRESENTER: LIFANG ZHANG

**Eduardo Blumwald**: The Combined Effect of Abiotic Stresses Reveals Unique Cell Type-Specific Molecular Changes in Poplar
PRESENTER: Vimal Kumar Balasubramanian

**Eduardo Blumwald**: Transgenic poplar analysis of native and synthetic inducible promoters for sensing abiotic stress and tissue specificity from poplar cis-regulatory elements.
PRESENTER: Yongil Yang

**Nicholas Bouskill**: Climate history dictates microbial metabolic response to drought stress: from semi-arid soils to tropical forest precipitation gradients
PRESENTER: Nick Bouskill

**Nanette Boyle**: Agent-Based Modeling of Algae Reveals Impact of Self-Shading on Metabolism
PRESENTER: Alexander Metcalf

**Stephen Burley**: RCSB Protein Data Bank: Connecting genes to structures to ecosystems
PRESENTER: Stephen Burley

**Kristin Burnum-Johnson**: Spatiotemporal Mapping of Perturbations in a Naturally Evolved Fungal Garden Microbial Consortium
PRESENTER: Kristin Burnum-Johnson

## C

**John Cahill**: Direct measure of amino acid distribution across Populus trichocarpa roots in a rhizosphere-on-a-chip habitat
PRESENTER: John Cahill

**William Cannon**: Principles of Fungal Metabolism Growth and Bacterial Interactions
PRESENTER: William Cannon

**Patrick Chain**: A Bioinformatic Pipeline to Identify and Classify Potential Microbial Signatures from Fungal Sequencing Data Integrated with a Searchable Database of Described Interactions
PRESENTER: Michal Babinski

**Patrick Chain**: Development of a model root-associated microbial consortia to investigate the roles of bacterial-fungal-plant interactions in the heat- and drought-tolerant grass Bouteloua gracilis (blue grama)
PRESENTER: Buck Hanson

**Patrick Chain**: Bacterial Communities in Association to Specific Tissues in Wild Morels
PRESENTER: Pilar Junier

**Patrick Chain**: Comparative Genomics of Intracellular Mollicutes-related Bacterial Endosymbionts of Fungi
PRESENTER: Reid Longley

**Patrick Chain**: Advanced Fluorescence Microscopy Techniques to Measure Bacterial:Fungal Interactions
PRESENTER: Demostheens Morales

**Patrick Chain**: Development of untargeted metabolomics approaches to study bacterial-fungal co-cultures
PRESENTER: Baltazar Zuniga

**Clint Chapple**: Identification of novel tryptophan-derived metabolites and associated genes of Arabidopsis by integrating PODIUM and mGWAS
PRESENTER: Zhiwei Luo

**George Church**: A workflow for the systems-level analysis design and engineering of genomically recoded organisms
PRESENTER: Anush Chiappino-Pepe

**George Church**: A rapid biosensor engineering platform by translatable fluorogenic amino acids
PRESENTER: Erkin Kuru

**George Church**: Profiling bacterial tRNA 3’-termini in-vivo: Establishment of a deep-sequencing approach and a riboswitch based bio-sensor approach.
PRESENTER: Kamesh Narasimhan

**George Church**: Construction of a Synthetic 57-Codon E. coli Chromosome and Tools for Microbial Genome-Scale Recoding
PRESENTER: Akos Nyerges

**George Church**: Genetic Tools for Photosynthetic Microbes; Toward Genome-scale Engineering
PRESENTER: Max Schubert

**John Coates**: The Global Prevalence and Biogeochemical Impact of Ancient Phosphorus-Oxidizing Bacteria
PRESENTER: Sophia Ewens

**Benjamin Cole**: Defining the Influence of Environmental Stress on Bioenergy Feedstocks at Single-Cell Resolution
PRESENTER: Benjamin Cole

**Melissa Cregger**: Understanding the Effects of Populus—Mycorrhizal Associations on Plant Productivity and Resistance to Abiotic Stress
PRESENTER: Melissa Cregger

**Hugo Cuevas**: Genetic mapping of sugarcane aphid resistance in sorghum line SC112-14
PRESENTER: HUGO  E CUEVAS

## D

**Brian Davison**: Applying new models to describe biomembrane structure and solvent partitioning in living cell membranes and membrane mimics
PRESENTER: Jonathan Nickels

**Brian Davison**: Noncellulosic biopolymer morphology and structural changes during real-time reaction studies.
PRESENTER: Sai Venkatesh Pingali

**Brian Davison**: Solid state NMR characterization of lipid membrane and organic solvent induced effects
PRESENTER: Yunqiao Pu

**Brian Davison**: Effect of Cyrene Pretreatment on Switchgrass Lignin Structure
PRESENTER: Yun-Yan Wang

**Patrik D'haeseleer**: Probabilistic Annotation and Ensemble Metabolic Modeling in KBase
PRESENTER: Patrik D'haeseleer

**José Dinneny**: Mining Innovative Strategies for Stress-resilient Growth in Extremophyte Relatives of Brassicaceae Oil Crops
PRESENTER: Dong-Ha Oh

**José Dinneny**: Cell type specific abiotic stress responses among closely related Brassicaceae species reveal how stress tolerance in extremophytes differ from stress-sensitive models
PRESENTER: Guannan Wang

**Mitchel Docktycz**: Plant-Microbe Interfaces: Development of an experimental approach to achieve spatially resolved plant root-associated metaproteomics using an agar-plate system
PRESENTER: Manasa Appidi

**Mitchel Docktycz**: Plant-Microbe Interfaces: Temporal dynamics of the Populus microbiome across scales
PRESENTER: Nicholas Dove

**Mitchel Docktycz**: Plant-Microbe Interfaces: Plant mechanisms shaping poplar root microbiota
PRESENTER: Felix Fracchia

**Mitchel Docktycz**: Plant-Microbe Interfaces: Network integration uncovers gene-targets involved in plant- microbial interactions
PRESENTER: Piet Jones

**Mitchel Docktycz**: Plant-Microbe Interfaces: A high-throughput bioassay to investigate bacterial-provided benefits to heat stress
PRESENTER: Jun Lee

**Mitchel Docktycz**: Plant-Microbe Interfaces: Determining the rate and consequences of horizontal gene transfer in the rhizosphere by simulating lateral spread of salicylate catabolism genes
PRESENTER: Josh Michener

**Mitchel Docktycz**: Plant-Microbe Interfaces: Investigating how interactions and local environments in the rhizosphere influence microbial responses and behavior
PRESENTER: Jennifer Morrell-Falvey

**Mitchel Docktycz**: Plant-Microbe Interfaces: Modulation of microbial community dynamics by contrasting regulation of salicylic acid and jasmomic acid-ethylene signaling pathways
PRESENTER: Wellington Muchero

**Mitchel Docktycz**: Plant-microbe interfaces: Interrogating poplar fungal microbiome interactions using meta- transcriptomics and constructed communities
PRESENTER: Jake Nash

**Mitchel Docktycz**: Plant-Microbe Interfaces: Quantification of Populus transcriptomic response to colonization by select bacterial symbionts
PRESENTER: Bryan Piatkowski

**Mitchel Docktycz**: Plant-Microbe Interfaces: Towards a rhizosphere on a chip for understanding physical and chemical transitions in multi-kingdom systems
PRESENTER: Scott Retterer

**Mitchel Docktycz**: Plant-Microbe Interfaces: Beyond symbiosis fungi can specifically perceive and use lipochitooligosaccharides to organize and modulate the development of microbial communities
PRESENTER: Tomas Rush

**Mitchel Docktycz**: Bio-Scales: Defining gene function and its connection to ecosystem processes
PRESENTER: Chris Schadt

**Mitchel Docktycz**: Plant-Microbe Interfaces: Identification of genes controlled by a plant-responsive transcription factor in the Populus endophyte Pseudomonas GM79
PRESENTER: Amy Schaefer

**Mitchel Docktycz**: Plant-Microbe Interfaces: Identification of PtLecRLK1-based signaling cascade in Laccaria bicolor root colonization
PRESENTER: Tao Yao

**Tim Donohue**: Improving Iron-Sulfur Cluster Stability in Zymononas mobilis to Increase Terpenoid Production
PRESENTER: Isabel Askenasy

**Tim Donohue**: A High Solids Field-To-Fuel Research Pipeline to Identify Interactions Between Feedstocks and Biofuel Production
PRESENTER: Meenaa Chandrasekar

**Tim Donohue**: Characterizing fungal inhibitors from drought-stressed switchgrass
PRESENTER: Sarvada Chipkar

**Tim Donohue**: Genome-scale CRISPRi Enables Functional Annotation of the Zymomomas mobilis Genome
PRESENTER: Amy Enright

**Tim Donohue**: Economic and Environmental Analysis of Bioenergy with Carbon Capture and Sequestration (BECCS) Systems
PRESENTER: Caleb Geissler

**Tim Donohue**: Multiple-approaches to engineer mixed-linkage glucan in sorghum
PRESENTER: Sang-Jin Kim

**Tim Donohue**: Building a Molecular Understanding of Biomass Deconstruction
PRESENTER: Nathaniel Kuch

**Tim Donohue**: Development and Function of Bioenergy Sorghum's Deep Roots
PRESENTER: Austin Lamb

**Tim Donohue**: Role of the Phyllosphere microbiome in Sorghum Resilience
PRESENTER: Marco Mechan Llontop

**Tim Donohue**: Can Perennial Bioenergy Cropping Systems Promote Negative N2O Fluxes?
PRESENTER: Ekrem Ozlu

**Tim Donohue**: Effects of Chloramphenicol Treatment on Cellular Storage Granules and Membrane Structures in Rhodobacter sphaeroides
PRESENTER: Daniel Parrell

**Tim Donohue**: Engineering Novosphingobium aromaticivorans to Stoichiometrically Convert S-type Aromatics into PDC
PRESENTER: Jose Perez

**Tim Donohue**: Lipid Membrane Remodeling during Ethanol Isobutanol and Lignocellulosic Hydrolysate Stress in Zymomonas mobilis
PRESENTER: Julio Rivera Vazquez

**Tim Donohue**: Crabtree/Warburg-like aerobic xylose fermentation by engineered Saccharomyces cerevisiae
PRESENTER: Trey Sato

**Tim Donohue**: Linking Microbial Funneling to Hydrogenolysis-Based Lignin Depolymerization to Produce 2-Pyrone-46-Dicarboxylic Acid from Phenolic Monomers
PRESENTER: Canan Sener

**Tim Donohue**: Analysis of SI Engine Alternative Fuels for Mixture Formation Behavior Related to Particulate Matter Formation with a Representative Gasoline Surrogate: Application to 2-Methyl-3-Buten-2-ol (Methyl Butenol MBO)
PRESENTER: Andrea Shen

**Sharon Doty**: Elucidation of the Roles of Diazotrophic Endophyte Communities in Promoting Productivity and Resilience of Populus through Systems Biology Approaches
PRESENTER: Sharon Doty

**John Dunbar**: Microbial regulation of soil water repellency to control soil degradation
PRESENTER: Emily Boak

**John Dunbar**: Constructing microbial networks from Genome Scale Metabolic Models.
PRESENTER: James Brunner

**John Dunbar**: Microbially-Drive Carbon Flow Persists During Surface Litter Decomposition
PRESENTER: Marie Kroeger

**John Dunbar**: Nitrogen Availability Strongly Affects Carbon Cycling by Sub-Surface Microbial Communities
PRESENTER: Sanna Sevanto

**John Dunbar**: Utilization of Lignocellulosic Biofuel Conversion Residue by Diverse Microorganisms
PRESENTER: Caryn Wadler

## E

**Robert Egbert**: Identifying Data-Driven Gene Targets to Control Bacterial Fitness
PRESENTER: Shara Balakrishnan

**Robert Egbert**: Catabolism of Methyl-3-(4-hydroxyphenyl)propionate (MHPP) a Model Substrate for Metabolic Addiction with Unexpected Implications for p-Coumaric Acid Catabolism in Pseudomonas flourescens
PRESENTER: Joshua Elmore

**Robert Egbert**: Simplified Microbial Communities Provide Diverse Complementation Potential for Genome-reduced Microorganisms
PRESENTER: Citlali Fonseca-Garcia

**Robert Egbert**: Spatially-resolved Multi-omics Analyses Reveal Key Taxa Responding to Root Depth in the Sorghum Rhizosphere
PRESENTER: Pubudu Handakumbura

**Robert Egbert**: Discovery of Bacterial Species and Molecular Mechanisms Driving Growth on the Plant Root Metabolite Sorgoleone
PRESENTER: Ryan McClure

**Emiley Eloe-Fadrosh**: Standardized Workflows for Microbiome Omics Data Analysis
PRESENTER: Bin Hu

**Joanne Emerson**: Evidence for Active Dynamic Viral Communities in Wet Soils across Habitats
PRESENTER: Joanne Emerson

**Andrea Eveland**: PhytoOracle: Modular Scalable Phenomic Data Processing Pipelines
PRESENTER: Emmanuel Gonzalez

**Andrea Eveland**: Using Molecular Genetics and Precision Phenotyping to Map Gene Function Contributing to Drought Resilience in Sorghum
PRESENTER: Yuguo Xiao

## F

**Mary Firestone**: Spatial Turnover of Soil Viral Communities and Genotypes Overlain by Cohesive Responses to Moisture in Grasslands
PRESENTER: Christian Santos-Medellin

**Brian Fox**: Additions to the BAHD Acyltransferase Toolbox
PRESENTER: Brian Fox

## G

**Nicholas Gauthier**: Optimizing enzymes for plastic upcycling using machine learning design and high throughput experiments
PRESENTER: Benjamin Fram

**Jean Gibert**: Does Predation by Protists Mediate the Effects of Temperature and Nutrient Additions on Microbial Food Webs?
PRESENTER: Katrina DeWitt

**Jean Gibert**: Predicting complex microbial temperature responses across scales
PRESENTER: Daniel Wieczynski

**Ryan Gill**: Enhancement of the Random Barcode Transposon Sequencing Applications within KBase
PRESENTER: Omree Gal-Oz

**Ryan Gill**: Genome-Scale CRISPRi in the Rapidly Growing Cyanobacterium Synechococcus sp. PCC 7002
PRESENTER: Andrew Hren

**Ryan Gill**: A S. cerevisiae "Marionette" strain to control metabolic pathways
PRESENTER: Jong Hyun Park

**Ryan Gill**: Design and Engineering of Synthetic Control Structures
PRESENTER: David Romero-Suarez

**Michael Guarnieri**: IMAGINE BioSecurity: Mesocosm based methods to evaluate biocontainment strategies and impact of industrial microbes upon native ecosystems.
PRESENTER: Kathleen Arnolds

**Michael Guarnieri**: IMAGINE BioSecurity: Biocontainment Efficacy of Toxin-Antitoxin Cassettes in Laboratory and Industrial S. Cerevisiae
PRESENTER: Gabriella Li

**Michael Guarnieri**: IMAGINE BioSecurity: Biocontainment of Genetically Engineered Cyanobacteria
PRESENTER: Jacob Sebesta

**Michael Guarnieri**: Combinatorial Biocontainment Design and DNA Barcode Genotyping
PRESENTER: Yo Suzuki

**Michael Guarnieri**: Modeling Bacterial Metabolism and Expression to Develop Biocontainment Strategies
PRESENTER: Juan D. Tibocha-Bonilla

**Michael Guarnieri**: IMAGINE BioSecurity: Metabolic Modeling-Enabled Biocontainment Redesign in Microbial Chasses
PRESENTER: Wei Xiong

## H

**Kirsten Hofmockel**: The Soil Lipidome is a Robust Indicator of the Microbial Community Response to Rewetting Following a Summer Drought
PRESENTER: Sneha Couvillion

**Kirsten Hofmockel**: Distribution of soil microbial necromass accumulation controlled by microbe-mineral interactions
PRESENTER: Kirsten Hofmockel

**Kirsten Hofmockel**: Profiling of the Microbiome Metabolic Response to Soil Rewetting
PRESENTER: Mary Lipton

**Kirsten Hofmockel**: Coordination of species roles during chitin decomposition in a model soil microbial consortium
PRESENTER: Ryan McClure

**Kirsten Hofmockel**: Three Feet Deep: Abiotic and Biotic Drivers of Organic and Mineral Soil Carbon Cycling
PRESENTER: Katherine Naasko

**Kirsten Hofmockel**: Visually Mapping Phenotypes and Community Interactions at the Microbial Scale
PRESENTER: Natalie Sadler

**Kirsten Hofmockel**: Omics-enabled global gapfilling (OMEGGA) for phenotype-consistent metabolic network reconstruction of microorganisms and communities
PRESENTER: Hyun-Seob Song

**Kirsten Hofmockel**: Functional and structural characterization of soil viral auxiliary metabolic genes
PRESENTER: Ruonan Wu

**Bruce Hungate**: Temperature sensitivity of soil bacterial networks from the Arctic to the Tropics
PRESENTER: Bram Stone

## J

**Tiffany Jamann**: Novel quantitative trait loci for leaf blight resistance in sorghum aids in understanding of
PRESENTER: Sarah Lipps

**Laura Jarboe**: Novel Systems Approach for Rational Engineering of Robust Microbial Metabolic Pathways
PRESENTER: Laura Jarboe

**Laura Jarboe**: Mutations for Improved Enzyme Functionality at High Temperatures and Low pH
PRESENTER: Ambuj Kumar

**Kolby Jardine**: The Scent of Senescence: Cell Wall Ester Modifications and Volatile Emission Signatures of Plant Responses to Abiotic Stress
PRESENTER: Kolby Jardine

**Michael Jewett**: Phylogenomics of Solvent-Producing Clostridium Species To Enable Carbon-Negative Production of Acetone and Isopropanol
PRESENTER: Steve Brown

**Michael Jewett**: Temperature drives substantial metabolic changes in the gas fermenting Clostridium autoethanogenum as revealed by multi-omics characterization
PRESENTER: Payal Chirania

**Michael Jewett**: Rapid Prototyping for Development of A Novel Gas-to-13-Butanediol Bioprocess
PRESENTER: Fungmin (Eric) Liew

**Michael Jewett**: Dynamic Kinetic Models Capture Cell-Free Metabolism for Improved Metabolic Engineering
PRESENTER: Jacob Martin

**Michael Jewett**: Unearthing Enzyme Promiscuity with Cheminformatics to Design Biosynthetic Pathways Towards Novel Biomolecules
PRESENTER: Zhuofu Ni

**Michael Jewett**: Mechanistic Insights into Cell-free Gene Expression from an Integrated -Omics Analysis of Extract Preparation Methods
PRESENTER: Blake Rasor

**Yongqin Jiao**: Sequence Entanglement with Post-Entanglement Modifications Enhances Functionality and Biosecurity of Entanglement Pairs
PRESENTER: Jennifer Chlebek

**Yongqin Jiao**: Characterization of Toxins and Their Cognate Inactivators as Kill Switch Actuators in Plant-Beneficial Pseudomonas fluorescens
PRESENTER: Tiffany Halvorsen

**Yongqin Jiao**: Learning Protein Fitness Models from Evolutionary and Experimental Data
PRESENTER: Chloe Hsu

**Yongqin Jiao**: CAMEOX: Enhanced Computational Design of Overlapping Genes to Prolong Synthetic Device Function and Limit Horizontal Gene Transfer
PRESENTER: Jose Manuel Martí

**Yongqin Jiao**: Redesigning the Escherichia coli genome with a 19-Amino Acid Alphabet
PRESENTER: Guillaume Urtecho

**Yongqin Jiao**: Programmed Lysis of Cells in Response to Electrogenetic Inputs
PRESENTER: Eric VanArsdale

**Martin Jonikas**: Transforming our understanding of chloroplast-associated genes through comprehensive characterization of protein localizations and protein-protein interactions
PRESENTER: LIANYONG WANG

**Thomas Juenger**: Modeling Environmental Influences on Biomass Composition of Diverse Switchgrass Genotypes
PRESENTER: Laura Bartley

**Thomas Juenger**: Exploring Switchgrass Genetic Diversity with Multiple Reference Genomes
PRESENTER: John Lovell

**Thomas Juenger**: A Generalist-Specialist Tradeoff between Switchgrass Cytotypes Impacts Climate Adaptation and Geographic Range
PRESENTER: Joseph Napier

**Thomas Juenger**: Enrichment of Soil Bacterial Taxa in Switchgrass (Panicum virgatum) Cropping Systems Across a Latitudinal Gradient
PRESENTER: Michael Ricketts

**Thomas Juenger**: Agrobacterium-mediated Transformation of P. hallii
PRESENTER: Anthony Trieu

## K

**Jay Keasling**: Carbon Conserving Redox Balanced Co-Utilization of Aromatics and Sugar by Engineered Pseudomonas putida
PRESENTER: Deepika Awasthi

**Jay Keasling**: Liquid chromatography-mass spectrometry analysis of lignin depolymerized products from plant hydrolysates
PRESENTER: Edward Baidoo

**Jay Keasling**: Multiproduct Cellulosic Biorefinery Enables Market-Competitive Gasoline and Jet Fuel Blendstocks
PRESENTER: Nawa Baral

**Jay Keasling**: One-pot deconstruction and conversion of dry and ensiled sorghum
PRESENTER: Julius Choi

**Jay Keasling**: Field testing of transgenic Sorghum variants overexpressing 4 types of aromatic compounds
PRESENTER: Jutta Dalton

**Jay Keasling**: Understand the Nanoarchitecture of Native and Engineered Plant Cell Wall via Multi-dimensional Solid-state NMR
PRESENTER: Yu Gao

**Jay Keasling**: Engineering and optimization of lignin catabolic pathways in Rhodosporidium toruloides
PRESENTER: Valentina Garcia

**Jay Keasling**: A media optimization pipeline for improving TRY with machine learning
PRESENTER: Matthew Incha

**Jay Keasling**: Multifaceted Adaptive Laboratory Evolution Approaches for Improving Host Phenotypes
PRESENTER: Hyun Gyu Lim

**Jay Keasling**: LigCHIP Technology for Bond-Specific Analysis of Lignocellulose Deconstructing Enzymes
PRESENTER: Jenny Onley

**Jay Keasling**: Microbiomes that metabolize lignin fragments obtained by chemical pretreatment of biomass
PRESENTER: Mee-Rye Park

**Jay Keasling**: Integrated chemical and biological catalysis for the valorization of ionic liquid-based biorefinery lignin
PRESENTER: LE THANH MAI PHAM

**Jay Keasling**: High-Throughput Screening of Wild Type and Engineered Plants using the JBEI Feedstocks-To-Fuels (F2F) Pipeline
PRESENTER: Venkataramana Pidatala

**Jay Keasling**: Structural characterization of lignin degrading enzyme PmdC involved in the synthesis of polymer precursor 2-pyrone 46 dicarboxylic acid (PDC)
PRESENTER: Andria Rodrigues

**Jay Keasling**: Comparison of Isoprenol Production from Sorghum Biomass Hydrolysates using Engineered Microbial Hosts
PRESENTER: Aparajitha Srinivasan

**Jay Keasling**: High Throughput Bioengineering Using a Microfluidic Platform
PRESENTER: Swarnagowri Vaidyanathan

**Jay Keasling**: Comparing In Planta and Microbial Production of Bioproducts
PRESENTER: Minliang Yang

**Jay Keasling**: Isoprenol production in P. putida KT2440
PRESENTER: Ian Yunus

**Jay Keasling**: Characterization and Stress Response of Monolignol p-Hydroxybenzoyltransferase in Poplar
PRESENTER: Yunjun Zhao

**Jay Keasling**: Deep Neural Network-Guided Design of Orthogonal Trans-Elements for Plant Synthetic Biology
PRESENTER: Andy Zhou

**Matias Kirst**: Two Shifts in Evolutionary Lability Explain Many Independent Origins of Nitrogen-Fixing Nodulation Symbiosis in A Single Clade of The Plant Tree of Life
PRESENTER: Heather Kates

**Matias Kirst**: Single-cell discovery of nodulation regulators in Medicago and evaluation of the functional role of homologs in Populus
PRESENTER: Matias Kirst

**Matias Kirst**: Dynamic change in chromatin accessibility predicts regulators of nodulation in Medicago truncatula
PRESENTER: Sara Knaack

**Kate Kucharzyk**: SynThetic BiolOgy Driven Approach to Repurpose PolyaMides (STORM)
PRESENTER: Kate Kucharzyk

## L

**Andrew Leakey**: Sustainable Production of Acrylic Acid via 3-Hydroxypropionic Acid from Lignocellulosic Biomass
PRESENTER: Sarang Bhagwat

**Andrew Leakey**: Refining Metabolic Engineering Strategies for Hyperaccumulation of Triacylglycerol in Oilcane
PRESENTER: VIET CAO

**Andrew Leakey**: High Throughput Screening of Mutant Libraries for Producing Medium Chain Fatty Acids Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Analysis of Microbial Colonies
PRESENTER: Kisurb Choe

**Andrew Leakey**: Analyzing Lipid Synthesis and Turnover Using Stable Isotopes
PRESENTER: Teresa Clark

**Andrew Leakey**: Automated high-throughput genome editing of TAG-related genes in plants
PRESENTER: jia dong

**Andrew Leakey**: Cell-type specific analysis reveals the spatial and temporal transcriptome of the Sorghum bicolor stem
PRESENTER: Jie Fu

**Andrew Leakey**: Impact of Sugarcane Cultivation on C and H2O Fluxes in Southeastern United States following Conversion from Grazed Pastures
PRESENTER: Nuria Gomez-Casanovas

**Andrew Leakey**: Leveraging Population Genomics Analysis of Issatchenkia orientalis for Engineering a Better Strain for the Production of Lignocellulosic-biomass-based Bioproducts
PRESENTER: Ping-Hung Hsieh

**Andrew Leakey**: A New Bioenergy Model that Simulates the Impacts of Plant-Microbial Interactions Soil Carbon Protection and Mechanistic Tillage on Soil Carbon Cycling
PRESENTER: Stephanie Juice

**Andrew Leakey**: Investigation of xylose metabolism in Rhodosporidium toruloides using a modular cloning kit (RT-EZ)
PRESENTER: HYUNGI KOH

**Andrew Leakey**: Metabolic Engineering of Triacylglycerols in Vegetative Tissues of Sorghum
PRESENTER: Kiyoul Park

**Andrew Leakey**: Using Cross-Scale Data to Constrain an Agro-Ecosystem Model to Produce Estimates of Miscanthus Production at a Field-Scale
PRESENTER: Bryan Petersen

**Andrew Leakey**: Optimization of Solvent-Free Enzymatic Esterification of Free Fatty Acids using Taguchi Design of Experiments
PRESENTER: Ramkrishna Singh

**Andrew Leakey**: Metabolic Engineering of Issatchenkia orientalis for succinic acid production
PRESENTER: Vinh Tran

**Andrew Leakey**: Metabolic Engineering Yarrowia lipolytica to Produce 3-acetyl-12-diacyl-sn-glycerol
PRESENTER: Qiang Yan

**Andrew Leakey**: Responses of Total (DNA) and Metabolically Active (RNA) Microbial Communities in Miscanthus x giganteus Cultivated Soil
PRESENTER: Jihoon Yang

**Andrew Leakey**: Gene duplication and single nucleotide polymorphisms (SNPs) via adaptive laboratory evolution (ALE) of engineered Yarrowia lipolytica enabled the efficient utilization of sugars in lignocellulosic hydrolysate
PRESENTER: Sangdo Yook

**Andrew Leakey**: Engineering Cyclopropane Fatty Acid Accumulation in Plant Vegetative Tissues
PRESENTER: Xiao-Hong Yu

**Andrew Leakey**: Determining Profit-Maximizing Dynamic Mix of Nitrogen Rate and Stand Age in Miscanthus and Switchgrass Production
PRESENTER: Na Zhang

**Andrew Leakey**: Genomics of Winter-hardiness and Yield in Diverse Miscanthus Germplasm
PRESENTER: Xuying Zheng

**Nina Lin**: Developing Understanding and Harnessing Modular Carbon/Nitrogen-Fixing Tripartite Microbial Consortia for Versatile Production of Biofuel and Platform Chemicals
PRESENTER: Xiaoxia Nina Lin

**Karen Lloyd**: Ten Pseudomonas spp. from Svalbard Active Layer are using amino acids for energy
PRESENTER: Raegan Paul

**Karen Lloyd**: Metagenome-assembled genomes from active layer in Ny Ålesund Svalbard (79°N) show population dynamics related to seasonal thawing and soil depth
PRESENTER: Katie Sipes

**Karen Lloyd**: High enzyme stability as a feature to enable slow microbial life in the subsurface
PRESENTER: Andrew Steen

**Karen Lloyd**: Comparative metagenomics of Arctic landscapes
PRESENTER: Tatiana Vishnivetskaya

**Stephen Long**: Enhancing Vegetative Oil Content through Optimized Lipogenic Factors
PRESENTER: Sanket Anaokar

**Stephen Long**: Sustainable Improvement of C4 Photosynthesis in Bioenergy Grasses
PRESENTER: Kher Xing Chan

**Stephen Long**: Improving Energycane by Metabolic Engineering for Hyperaccumulation of Lipids and RNAi Suppression of Flowering
PRESENTER: Baskaran Kannan

**Stephen Long**: Bridging the Gap between Academic and Commercial Biofuel Production: Pilot-scale Processing of Transgenic Energycane for Lipid and Sugar Recovery
PRESENTER: Shraddha Maitra

**Chaofu Lu**: Germination and Seed Size Variability in Camelina
PRESENTER: Jennifer Lachowiec

**Chaofu Lu**: Understanding Nitrogen Use Efficiency and Oilseed Traits in Camelina by High Resolution Genome Sequencing and Whole-genome Resequencing
PRESENTER: Charlemagne Lim

**Chaofu Lu**: Genetic Engineering of Camelina to Improve Seed Oil Yield
PRESENTER: Xiao-Hong Yu

**Chaofu Lu**: Camelina growth impacted by bacteria and nitrogen stress
PRESENTER: Qing Zheng

## M

**Hiroshi Maeda**: Development and usage of atomic mapping for estimation of nitrogen fluxes in plant metabolic networks
PRESENTER: Sebastian Huss

**Hiroshi Maeda**: Comprehensive Characterization of Multi-Substrate Specificity of Aminotransferase Family Enzymes
PRESENTER: Kaan Koper

**Robert Martienssen**: Biological Design of Lemnaceae Aquatic Plants for Biodiesel Production

PRESENTER: Evan Ernst

**Christopher Marx**: Converting Methoxy Groups on Lignin-Derived Aromatics from a Toxic Hurdle to a Useful Resource: A Systems-Driven Approach
PRESENTER: Christopher Marx

**Josh Michener**: Systems metabolic engineering of Novosphingobium aromaticivorans for lignin valorization
PRESENTER: Marco Allemann

**Christopher Miller**: Scalable Computational Tools For Inference Of Protein Annotation And Metabolic Models In Microbial Communities
PRESENTER: Janaka Edirisinghe

**Hosein Mohimani**: HypoNPAtlas: an Atlas of hypothetical natural product for mass spectrometry database search
PRESENTER: Liu Cao

**Tae Seok Moon**: Engineering novel microbes for upcycling waste plastic and solving climate crisis
PRESENTER: Tae Seok Moon

**James Moran**: Spatial plasticity in plant-microbe interactions in response to applied nutrient heterogeneity in soil
PRESENTER: James Moran

**Paula Mouser**: Environmental and Engineered Factors Influence Membrane Features of Shale Taxa
PRESENTER: Chika Ugwuodo

## N

**Ellen Neidle**: Design Detect Evolve: Engineering Syringol Degradation in ADP1
PRESENTER: Alyssa Baugh

**William Nelson**: Snekmer – A Tool for Protein Classification Using Amino Acid Recoding and Kmer Analysis
PRESENTER: William Nelson

**Krishna Niyogi**: Metabolite Excretion and Metabolic Flux Analysis in Chromochloris zofingiensis an Emerging Model Green Alga for Sustainable Fuel Production
PRESENTER: Yuntao Hu

**Krishna Niyogi**: Mineral Nutrition and Energy Pathways Regulate Lipid Accumulation and Photosynthesis
PRESENTER: Timothy Jeffers

**Krishna Niyogi**: Synthetic Biology in Oleaginous Green Algae
PRESENTER: Jeffrey Moseley

**Justin North**: Metabolic Modeling and Engineering of Enhanced Anaerobic Microbial Ethylene Synthesis

PRESENTER: William Cannon

**Justin North**: High Throughput Screening of Enzymes that Bolster Anaerobic Ethylene Synthesis
PRESENTER: Sarah Young

**Trent Northen**: The Development of Plant and Soil Fabricated Ecosystems (EcoFABs) for Standardized Microbiome Experiments
PRESENTER: Peter Andeer

**Trent Northen**: The Twin Ecosystems Project: A New Capability for Field and Laboratory Ecosystems Coupled by Sensor Networks and Autonomous Controls
PRESENTER: Javier A. Ceja-Navarro

**Trent Northen**: Rhizosphere Biogeography of Brachypodium distachyon and Microbial Plant-Growth Promoting Traits
PRESENTER: Romy Chakraborty

**Trent Northen**: Examining Molecular Mechanisms Selecting for Rhizosphere Bacteria
PRESENTER: Dawn Chiniquy

**Trent Northen**: Spatio-temporal genome-scale metabolic modeling of the rhizosphere microbiome
PRESENTER: Ilija Dukovski

**Trent Northen**: Evaluating Plant-Microbe Interactions Persistence and Movement of Microbial Communities Across Scales
PRESENTER: Albina Khasanova

**Trent Northen**: A rapid Brachypodium distachyon transformation method using leaf whorl explants
PRESENTER: Hsiao-Han Lin

**Trent Northen**: m-CAFEs Phage Engineering for Targeted Editing of Microbial Communities
PRESENTER: Avery Roberts

**Trent Northen**: Targeted DNA Editing Within Microbial Communities
PRESENTER: Benjamin Rubin

**Trent Northen**: The Filamentous Fungus Trichoderma atroviride as a Model System for Understanding Fungal Genetics the Plant-Fungal Symbiosis and Interactions with Diverse Bacteria
PRESENTER: Manuel Villalobos

# Table of Contents

(Click on Title to advance to the abstract)

**Jennifer Pett-Ridge**: Functional genomics of replicating microbes and viruses following rewetting of a Mediterranean grassland soil
PRESENTER: Ella Sieradzki

**Jennifer Pett-Ridge**: Soil habitats and water limitation shape microbial traits correlated with formation of mineral-associated organic matter
PRESENTER: Noah Sokol

**Jennifer Pett-Ridge**: Environmental conditions shape active viral community structure and virus-host dynamics in soil ecosystems
PRESENTER: Gareth Trubl

**Priya Ranjan**: Design and -omics exploration of synthetic communities
PRESENTER: Alexis Marsh

**Ruben Rellan Alvarez**: Improving candidate gene discovery by combining multiple genetic mapping datasets
PRESENTER: Ruben Rellan-Alvarez

## R

**Sue Rhee**: Nanoparticle-Mediated Transformation of Sorghum towards the Determination of a Subcellular Metabolic Network Map
PRESENTER: Destiny Davis

**Sue Rhee**: Systematic Identification of Subcellular Location of Sorghum Metabolic Enzymes
PRESENTER: Purva Karia

**Simon Roux**: Phage diversity and activity associated with seasonal changes in a model montane soil ecosystem
PRESENTER: Simon Roux

**Davinia Salvachua Rodriguez**: Cross-kingdom Comparative Genomics of Aromatic Catabolic Pathways in Fungi and Bacteria
PRESENTER: Davinia Salvachua Rodriguez

## S

**James Schnable**: Data driven analysis of nitrogen deficit responses across a sorghum diversity panel using hyperspectral leaf reflectance
PRESENTER: Michael Tross

**Danny Schnell**: The impact of LIP36 seed-specific expression on seed and oil yields in Camelina sativa and its associated transcriptome and metabolome changes
PRESENTER: Hesham Abdullah

**Danny Schnell**: Transcriptional regulation of lipid metabolism in Camelina sativa
PRESENTER: Suresh Kumar Gupta

**Danny Schnell**: High-Throughput Detection of T-DNA Insertion Sites for Multiple Transgenes in Complex Genomes

PRESENTER: Heike Sederoff

**Danny Schnell**: A mitochondrial carrier protein from Chlamydomonas alter the root architecture in Camelina.

PRESENTER: Rajneesh Singhal

**Joseph Schoeniger**: Expanding the Utility of Integrases for Genome Editing and Stabilizing Gene Modules in Target Bacteria

PRESENTER: Lauren Clark

**Joseph Schoeniger**: Repurposing Cas13 for Precise Translational Inhibition and Activation

PRESENTER: Brady Cress

**Joseph Schoeniger**: Phage Factory: Creating a phage for any bacterial species

PRESENTER: Catherine Mageeney

**Joseph Schoeniger**: Design delivery and expression of synthetic genetic elements in diverse microorganisms

PRESENTER: Jaymin Patel

**John Sedbrook**: The Impact of Waterlogging on Pennycress Morphology and Yield and Modeling Effective Water Availability of Pennycress Natural Populations

PRESENTER: Rachel Combs-Giroir

**John Sedbrook**: Root adaptive responses for improvement of abiotic stress tolerance in Pennycress

PRESENTER: Marcus Griffiths

**John Sedbrook**: Tolerance to Heat Stress in Natural Variants and CRISPR Gene-Edited High Oleic Acid Lines of the Oilseed Plant Pennycress (Thlaspi arvense).

PRESENTER: Nikhil Jaikumar

**John Sedbrook**: Pennycress as an Emerging Bioenergy Crop: How Does the Microbiome Impact Performance and Resilience Factors in the inland Pacific Northwest?

PRESENTER: Karen Sanguinet

**Jeremy Semrau**: Structural and redox requirements for dioxygen generation coupled to metal reduction by methanobactins: implications for greenhouse gas emissions

PRESENTER: Phillip Dershwitz

**Jeremy Semrau**: Microbial competition for copper can enhance greenhouse gas emissions

PRESENTER: Peng Peng

**Jeffrey Skolnick**: Genome-scale structural prediction of protein sequences and complexes with deep learning

PRESENTER: Mu Gao

**Claudia Solis-Lemus**: Identifying microbial drivers in biological phenotypes with a Bayesian Network Regression model
PRESENTER: Sam Ozminkowski

**Kevin Solomon**: Developing anaerobic fungal tools for efficient upgrading of lignocellulsic feedstocks
PRESENTER: Casey Hooker

**Kevin Solomon**: Plastic degradation and upcycling by the gut microbiome of yellow mealworms
PRESENTER: Lummy Monteiro

**Greg Stephanopoulos**: Improving Bioprocess Robustness by Cellular Noise Engineering
PRESENTER: Constantinos Katsimpouras

**Rhona Stuart**: Dynamic Phaeodactylum tricornutum Exometabolites Shape Surrounding Bacterial Communities
PRESENTER: Vanessa Brisson

**Rhona Stuart**: Siderophore-Producing Phycosphere Bacteria Alleviate Iron Limitation Stress in Phaeodactylum tricornutum
PRESENTER: Nicole Coffey

**Rhona Stuart**: Mycorrhizal Fungi Mediate Plant and Bacterial Response to Water Limitation in a Marginal Soil
PRESENTER: Rachel Hestrin

**Rhona Stuart**: Microscale Characterization Tools for Algal-Bacterial Interactions and Cell Sinking
PRESENTER: Hyungseok Kim

**Rhona Stuart**: A novel algicidal bacterium threatens diatom productivity and incorporates algal-derived carbon and nitrogen
PRESENTER: Megan Morris

**Rhona Stuart**: System-Level Analysis of Metabolism in a Novel Algicidal Bacterium
PRESENTER: Ali Navid

**Matthew Sullivan**: Nutrient Limitation Drives Dynamics of Host-virus Interactions
PRESENTER: Roya AminiTabrizi

**Matthew Sullivan**: Viral ecogenomics across a permafrost thaw gradient
PRESENTER: Akbar Adjie Pratama

## T

**Michiko Taga**: Corrinoids a Class of Model Metabolites Differentially Impact Isolation of Soil Bacteria
PRESENTER: Zoila Alvarez-Aponte

**Michiko Taga**: Impact of Corrinoids on Soil Community Assembly in a Grassland Soil
PRESENTER: Zachary Hallberg

**Neslihan Tas Baas**: Microbial controls on biogeochemical cycles in permafrost ecosystems
PRESENTER: Neslihan Tas

**Malak Tfaily**: MetaboDirect: An analytical pipeline for FT-ICR mass spectrometry data
PRESENTER: Malak Tfaily

**Cong Trinh**: Understanding Robustness of Yarrowia lipolytica for Undetoxified Biomass Hydrolysate Utilization
PRESENTER: Seunghyun Ryu

**Cong Trinh**: Controlling Selectivity of Modular Microbial Biosynthesis of Designer Acetate Esters through Proteome Reallocation
PRESENTER: Cong Trinh

**Danielle Tullman-Ercek**: Employing Bacterial Microcompartments To Create Privileged Redox Pools for Biofuel Production
PRESENTER: Carolyn Mills

**Danielle Tullman-Ercek**: Engineering bacterial microcompartments in Clostridium autoethanogenum to overcome bottlenecks in sustainable production of synthetic rubber
PRESENTER: Alex Mueller

**Gerald Tuskan**: Modeling and NMR approaches to probe spatial arrangements in biomass: Towards better and more predictive models of plant cell wall structures
PRESENTER: Bennett Addison

**Gerald Tuskan**: Techno-Economic Analysis and Life Cycle Assessment of a biorefinery utilizing reductive catalytic fractionation (RCF)
PRESENTER: Andrew Bartling

**Gerald Tuskan**: Deploying docking calculations alongside kinetic model parameterization to elucidate mechanisms controlling metabolism in Clostridium thermocellum
PRESENTER: Veda Boorla

**Gerald Tuskan**: Candidate genes for lignin structure identified through genome-wide association of naturally variant Populus
PRESENTER: Nathan Bryant

**Gerald Tuskan**: Transient delivery of Cas9 using Agrobacterium for genome editing
PRESENTER: Tim Chappell

**Gerald Tuskan**: Identification and validation of a key gene contributing to the differential flowering time of switchgrass ecotypes
PRESENTER: Soyeon Choi

**Gerald Tuskan**: Catalytic Upgrading of Bioderived Alcohols to Sustainable Aviation Fuel (SAF)
PRESENTER: Brian Davison

**Gerald Tuskan**: Dynamic Control of Aromatic Catabolism In Situ Efflux Pump Engineering and High-Throughput Functional Genomics in P. putdia KT2440 Enabled by CRISPR-Cas9  Dynamic control of the Beta-ketoadipate pathway In-situ efflux pump engineering and high-throughput
PRESENTER: Jacob Fenster

**Gerald Tuskan**: Rapid domestication of poplar using genomic selection and P. trichocarpa X P.deltoides hybrids Molecular breeding for improving poplar bioenergy feedstocks
PRESENTER: David Kainer

**Gerald Tuskan**: Discovery characterization and metabolic engineering of Rieske non-heme iron monooxygenases for guaiacol O-demethylation
PRESENTER: Eugene Kuatsjah

**Gerald Tuskan**: Declining Carbohydrate Solubilization with Increasing Solids Loading During Fermentation of Cellulosic Feedstocks by Clostridium thermocellum: Documentation and Diagnostic Tests Documenting and exploring potential effectors of carbohydrate solubilization
PRESENTER: Matt Kubis

**Gerald Tuskan**: Employing rapid accurate high-precision phenotyping in poplar switchgrass and bacteria
PRESENTER: John Lagergren

**Gerald Tuskan**: Drought stress alters plant-microbe interactions but is contingent on host genetic background.
PRESENTER: Spencer Roth

**Gerald Tuskan:** Integrating synthetic biology and polysaccharide synthesis for designer polymers with tunable properties
PRESENTER: Peter Smith

**Gerald Tuskan**: Systems biology approaches uncover distinct genes controlling switchgrass biomass and height Systems biology approaches uncover novel genetic variants conferring productivity traits in switchgrass
PRESENTER: Kyle Sullivan

**Gerald Tuskan**: Understanding the genetic basis of yield and drought tolerance In Populus trichocarpa
PRESENTER: Gail Taylor

**Gerald Tuskan**: Quantitative trait loci (QTL) mapping: a novel method for dissecting the genetic basis of complex phenotypes in bacteria
PRESENTER: Delyana Vasiliva

**Gerald Tuskan**: Genome-wide association studies of drought stress and water use efficiency related traits in switchgrass
PRESENTER: Yongqin Wang

**Gerald Tuskan**: Controlling hydrodeoxygenation of lignin to jet-range aromatic hydrocarbons
PRESENTER: Yaping Xu
PRESENTER: Matthew Webber

**Gerald Tuskan**: UAV- based sustainability traits modeling of field-grown switchgrass GWAS population

**Keith Tyo**: High-throughput Screening for Caboligase Activity in ThDP-dependent Enzymes
PRESENTER: Tracey Dinh

## U

**James Umen**: Curation and Characterization of Conserved Green Lineage Proteins
PRESENTER: James Umen

## V

**Kranthi Varala**: Infernet: Gene Function Inference By Leveraging Large Organ-Specific Expression Datasets And Validation Of Non-Redundant Regulators
PRESENTER: Rajeev Ranjan

## W

**Jin Wang**: Elucidating the Evolution of Interspecies Metabolic Interactions within a Methanotroph-Cyanobacteria Coculture using Dynamic Genome-scale Metabolic Modeling
PRESENTER: Kiumars Badr

**David Weston**: Microbiome transfer and synthetic community approaches for determining the genetic and environmental factors underlying mutualism within a Sphagnum peatmoss system
PRESENTER: Alyssa Carrell

**Jan Westpheling**: Development of emerging model microorganisms: Megasphaera elsdenii for biomass and organic acid upgrading to fuels and chemicals.
PRESENTER: Janet Westpheling

**Travis Wheeler**: Tools for Faster and More Sensitive Sequence Annotation and Visualization of Those Annotations
PRESENTER: Travis Wheeler

**Thea Whitman**: A mechanistic approach to parsing pyrophilous lifestyles of fungal and bacterial isolates
PRESENTER: Monika Fischer

**Thea Whitman**: A community-based approach to understanding fungal and bacterial responses to wildfire
PRESENTER: Nayela Zeba

**Mari Winkler**: N2O Formation and Organic Nitrogen Utilization in Wetland Microbial Communities
PRESENTER: Julie Johnston

**Mari Winkler**: Experimental models bridging single cell-to-ecosystem scales to evaluate climate-wetland feedback mechanisms
PRESENTER: Dongyu Wang

**Kelly Wrighton**: Genomes to Ecosystem Function: Targeting Critical Knowledge Gaps in Soil Methanogenesis and Translation to Updated Global Biogeochemical Models
PRESENTER: Angela Oliverio

## Y

**Todd Yeates**: Enabling Structure Determination of Challenging Samples with new cryo Electron Microscopy Methods
PRESENTER: Roger Castells Graells

**Todd Yeates**: Metabolism in Microbial Communities and the Associated Biochemistry of Polymer Deconstruction
PRESENTER: Robert Gunsalus

**Shinjae Yoo**: Automated knowledge harvesting from literature text tables and figures
PRESENTER: Shinjae Yoo

**Jamey Young**: Accelerating Strain Phenotyping with Desorption Electrospray Ionization-Mass Spectrometry Imaging and Untargeted Molecular Analysis of Intact Microbial Colonies
PRESENTER: Hawkins Shepard

**Jamey Young**: Systems biology to enable modular metabolic engineering of fatty acid production in cyanobacteria
PRESENTER: Jamey Young

## Z

**Alexandre Zanghellini**: Designing Novel Enzymes for Complete Degradation of Recalcitrant Polyamides
PRESENTER: Amandeep Sangha

**Karsten Zengler**: Experimental Approaches to Understand Rhizosphere Processes for Improved Bioenergy Crop Productivity
PRESENTER: Yuntao Hu

**Karsten Zengler**: Metabolic modeling of a rhizosphere microbial community
PRESENTER: Manish Kumar

**Karsten Zengler**: Reconstruction of a genome-scale community metabolic model of a microbial co-culture to enable next generation biochemical production
PRESENTER: Gustavo Lastiri-Pancardo

**Philipp Zerbe**: Improved Biofuel Production through Discovery and Engineering of Terpene Metabolism in Switchgrass.
PRESENTER: Kira Tiedge

**Jiwei Zhang**: A gene-editing system for large-scale fungal phenotyping in a model wood decomposer
PRESENTER: Jiwei Zhang

**Huimin Zhao**: Multi-omic investigation of lipid accumulation mechanisms in R. toruloides
PRESENTER: Anshu Deewan

**Huimin Zhao**: Population phenomics of Issathenkia orientalis for creation of a safer and more robust chassis for next-generation industrial biotechnology

PRESENTER: Yusuke Sasaki

**Huimin Zhao**: Examining organic acid production and model-driven strategies in Issatchenkia orientalis
PRESENTER: Patrick Suthers

**Title:** Developing detection and countermeasure tools for CRISPR-based genome engineering tools in plants

**Authors:** Yang Liu[1]* (Liuy4@ornl.gov), Guoliang Yuan,[1] Xiaohan Yang,[1] and **Paul E. Abraham**[1]

**Institutions:** [1] Bioscience Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

**Website URL:** https://seed-sfa.ornl.gov/

**Project Goals:** The Secure Ecosystem Engineering and Design (SEED) Science Focus Area (SFA), led by Oak Ridge National Laboratory, combines unique resources and expertise in the biochemistry, genetics, and ecology of plant-microbe interactions with new approaches for analysis and manipulation of complex biological systems. The long-term objective is to develop a foundational understanding of how non-native microorganisms establish, spread, and impact ecosystems critical to U.S. Department of Energy missions. This knowledge will guide biosystems design for ecosystem engineering while providing the baseline understanding needed for risk assessment and decision-making.

**Abstract Text:**

CRISPR/Cas-based gene editing tools have been widely applied in plants. However, genome engineering in plants is currently limited by the reliability and performance of these tools. At present, typically only one or a few genes can be edited at a time and most successful edits are imprecise end-joining of double-stranded breaks (DSBs). While these knockout mutations can be useful in some contexts, these are often inadequate to confer complex traits needed to improve DOE-model biofeedstocks. More sophisticated gene edits are now becoming available with advanced tools such as based and prime editors, yet these tools have only recently been implemented in plants and successful edits are often rare events. To fully realize the potential of CRISPR tools in plants, methods are needed to assess the performance of these tools faster, with less cost, and at scale. For this reason, we developed a biosensor system for real-time detection of active CRISPR/Cas tools *in planta*[1]. Specifically, several non-functional GFP genes which cannot emit fluorescence signals together with single guide RNAs (sgRNAs) were built as biosensors. In the presence of various CRIPSR tools such as Cas9 endonuclease, base editors (BEs), and prime editors (PEs), nonfunctional GFP mutants can be successfully converted into the functional GFP gene, generating green fluorescence.

Over the past several years, there has been a rapidly growing interest in using plant viruses as vectors to deliver and spread genome editing tools. This is because the use of viruses, to create gene edits through infection, has been recognized as a method to overcome current bottlenecks in the production of gene edited plants at scale. The consequent ability to edit plants through viral infection introduces concerning biosecurity risks. Also, there are increasing concerns about the biosafety issues caused by CRISPR tools, such as off-target effects of CRISPR/Cas systems,

CRISPR-based contaminating gene drive, and potential malicious genomic modifications mediated by CRISPR tools. For these reasons, in addition to continually working to optimize biosensors for the real-time detection of active CRISPR/Cas tools[1], we are developing countermeasure systems for inhibiting the activity of CRISPR/Cas systems in planta. Using the developed biosensor systems, we successfully demonstrated that AcrIIA4 and AcrIIA5 could inhibit the activities of CRISPR/Cas9 and one base editor in Arabidopsis protoplasts. The functionality of AcrIIA4 and AcrIIA5 will be further investigated in *Nicotiana benthamiana* (tobacco) and *Populus tremula* x *P. alba* 717-1B4.

**References/Publications**

1. Yuan, G. L.; Hassan, M. M.; Yao, T.; Lu, H. W.; Vergara, M. M.; Labbe, J. L.; Muchero, W.; Pan, C. T.; Chen, J. G.; Tuskan, G. A.; Qi, Y. P.; Abraham, P. E.; Yang, X. H., Plant-Based Biosensors for Detecting CRISPR-Mediated Genome Engineering. *Acs Synth Biol* **2021**, 10, 12, 3600–3603

**Funding Statement:** The Secure Ecosystem & Engineering Design Science Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research, under FWP ERKPA17. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

**Title:** Introductions of plant growth promoting bacterial strains differentially modify growth and health of *Populus* biofeedstocks

**Authors:** Jessica A. M. Moore[1]* (mooreja1@ornl.gov), Joshua K. Michener[1], Wellington Muchero[1], Melissa A. Cregger[1], and **Paul E. Abraham[1]**

**Institutions:** [1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

**Website URL:** https://seed-sfa.ornl.gov/

**Project Goals:** The Secure Ecosystem Engineering and Design (SEED) Science Focus Area (SFA), led by Oak Ridge National Laboratory, combines unique resources and expertise in the biochemistry, genetics, and ecology of plant-microbe interactions with new approaches for analysis and manipulation of complex biological systems. The long-term objective is to develop a foundational understanding of how non-native microorganisms establish, spread, and impact ecosystems critical to U.S. Department of Energy missions. This knowledge will guide biosystems design for ecosystem engineering while providing the baseline understanding needed for risk assessment and decision-making.

**Abstract text:**

Biofeedstock plants, such as poplar trees (*Populus sp.*), grow in unison with soil microbes that can aid plant productivity. Commercially available biostimulants and biopesticides containing bacterial species are marketed for enhancing plant growth and pest resistance. However, the establishment and spread of beneficial bacteria in biofeedstock plants may be constrained by a variety of factors including microbe and host plant gene-gene interactions, characteristics of the invading organisms or resident community, and the environment. We are investigating how non-model bacteria drive their establishment, persistence, and the productivity of biofeedstock perennial plants.

*Bacillus* species are ubiquitous bacteria in soils and many field trials have demonstrated their ability to stimulate plant growth and suppress plant pathogens. Here, we conducted a greenhouse experiment to examine how single or mixed *Bacillus* strain introductions establish and differentially affect several *Populus* genotypes. *Bacillus* strains were introduced to soils planted with either *Populus deltoides* (genotype 11347 or 16842), a F1 hybrid cross between *P. deltoides* and *P. trichocarpa* (7300 or 8360D), or a no-plant control. We introduced five strains of three *Bacillus* species: *B. velezensis* (GB03 or FZB42), *B. subtilis* (3610 or RO-NN-1), *B. amyloliquefaciens* (DSM7) or a mixture of all five genotypes. Plant growth parameters – photosynthetic rate, number of leaves, stem height, SPAD greenness index, and above and belowground biomass – were measured to quantify the effect of microbial invaders on plants. Overall, we found the plant response to *Bacillus* introduction could not be generalized, and indeed varied across the different plant genotypes.

Propagule pressure is one of the most common explanatory factors that impacts success in micro- and macro-organism invasion. Therefore, we interrogated the introduction of *B. veleznesis* at four levels of propagule pressure: 0 (control group), 50% of the commercially recommended rate (3.4 x $10^9$ CFU/gallon), the recommended rate (6.9 x $10^9$ CFU/gallon), and double the recommended rate (13.8 x $10^9$ CFU/gallon) in a field experiment. We hypothesized that (1) *P. trichocarpa* will support *B. velezensis* introductions; and (2) the establishment and rate of systemic spread will increase with increased propagule pressure. Our on-going field study is a fully factorial cross of 4 tree genotypes x 4 propagule pressure levels x 2 inoculation timing levels x 3 destructive harvest time points (n = 5, 480 plants total). We have completed two destructive harvests and one non-destructive harvest to-date. Sample processing and analyses are in progress. Preliminary findings suggest that *B. velezensis* GB03 inoculations did not alter microbial community composition in bulk soil two weeks post-planting when plants were inoculated at the time of cutting propagation. Harvests will continue non-destructively to monitor persistence of *Bacillus velezensis* GB03 and legacy effects in the microbial community over time. This is the first temporal study of biofungicide systemic spread in a biofeedstock plantation and has applications for national energy security.

**Funding statement:** The Secure Ecosystem & Engineering Design Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research, under FWP ERKPA17. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

**Title:** Explainable-AI driven feature engineering of CRISPR-Cas9 sgRNA efficiency leads to quantum biological insights into genome editing

**Authors:** Jaclyn Noshay[1]* (noshayjm@ornl.gov), Ashley Cliff[1,2], Jonathan Romero[1,2], Stephan Irle[3], David Kainer[1], Daniel Jacobson[1], and **Paul Abraham**[1]

**Institutions:** [1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]The Bredesen Center, University of Tennessee, Knoxville, TN; [3]Computational Chemistry and Nanomaterial, Oak Ridge National Laboratory, Oak Ridge, TN

**Website URL:** https://seed-sfa.ornl.gov/

**Project Goals:** The Secure Ecosystem Engineering and Design (SEED) Science Focus Area (SFA), led by Oak Ridge National Laboratory, combines unique resources and expertise in the biochemistry, genetics, and ecology of plant-microbe interactions with new approaches for analysis and manipulation of complex biological systems. The long-term objective is to develop a foundational understanding of how non-native microorganisms establish, spread, and impact ecosystems critical to U.S. Department of Energy missions. This knowledge will guide biosystems design for ecosystem engineering while providing the baseline understanding needed for risk assessment and decision-making.

**Abstract Text:**

High-throughput genome-wide studies using CRISPR-Cas tools have transformed capabilities for genetic manipulation in the laboratory. However, the performance of these tools is prone to error and increased uncertainty beyond model organisms and laboratory-controlled conditions. This is because current models for sgRNA prediction have primarily been trained on a narrow range of model species including human, mouse, drosophila, and zebrafish. While general "rules" have been defined in these organisms, the mechanistic underpinnings are not well understood. Additionally, with extreme variation in the genetic architecture between species and kingdoms, mammalian data alone lacks the information to generate an accurate predictive model in another organism. Herein, we utilized a feature engineering machine learning model, iterative random forest (iRF), to better understand the features of importance when considering an effective sgRNA. Using recently published data assessing genome-wide activity of gRNA for *E. coli*, we identified traits important for sgRNA design in a bacterial species and observed immense variation when tested with the same feature set in human and fungal species. We show the influence of expanding positional encoding to larger k-mers to capture intricate interactions in local and neighboring nucleotides. Most interestingly, we have utilized a novel feature set, quantum chemical tensors, that can be applied across species to improve sgRNA efficiency prediction as well as greatly enhance our understanding of the intricate quantum biological processes involved in CRISPR-Cas9 machinery. These advancements will provide a means to improve the safety and reliability of biosystem design and ecosystem engineering using non-model organisms.

**Funding Statement:** The Secure Ecosystem & Engineering Design Science Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research, under FWP ERKPA17. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

**Title:** Understanding the role of the Plasminogen-Apple-Nematode domain in host-plant invasion by pathogens.

**Authors:** Debjani Pal,[1] (pald@ornl.gov), Kuntal De,[1] Timothy B. Yates, [1] Kai Feng, [1] Wellington Muchero, [1] and **Paul A. Abraham**

**Institutions:** [1]Oak Ridge National Laboratory, Oak Ridge, TN 37831

**Website URL:** https://seed-sfa.ornl.gov/

**Project Goals:** The Secure Ecosystem Engineering and Design (SEED) Science Focus Area (SFA), led by Oak Ridge National Laboratory, combines unique resources and expertise in the biochemistry, genetics, and ecology of plant-microbe interactions with new approaches for analysis and manipulation of complex biological systems. The long-term objective is to develop a foundational understanding of how non-native microorganisms establish, spread, and impact ecosystems critical to U.S. Department of Energy missions. This knowledge will guide biosystems design for ecosystem engineering while providing the baseline understanding needed for risk assessment and decision-making.

**Abstract Text:** Invasive pathogens, causing diseases such as septoria stem canker caused by the fungal pathogen *Sphaerulina musiva*, negatively affect ecosystems and economies around the world. The exact mechanism by which these pathogens become established remains largely unknown, and most studies assume a predominant role by pathogen-derived factors in pathogenesis. In this project, we are investigating the role of susceptibility host-genes in the pathogenesis of *Populus* by *S. musiva*. A G-type-Lectin receptor-like kinase (G-type LecRLK) encoded in *Populus* was suggested to function as a susceptibility factor that enhanced infection by *S. musiva*. Subsequently, we demonstrated that the G-type LecRLK gene falls in the same category as proteins carrying the Plasminogen-Apple-Nematode (PAN) domain, which is highly implicated in cell invasion and pathogenesis parasitism. Herein, we will present evidence showing that this protein domain mediates receptor ubiquitination, cellular internalization, and proteolysis. We propose a model by which the cellular internalization of ubiquitinated receptors functions as an invasion platform for the pathogen to enter host cells. Identification of these genetic factors advances our ability to assess or modulate success of this invasive pathogen in *Populus*.

**Funding Statement:** The Secure Ecosystem & Engineering Design Science Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research, under FWP ERKPA17. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

**Title:** Understanding microbial establishment in the rhizosphere using quantitative trait-locus mapping and CRISPR Cas editing

**Authors**: Oumar Sacko[1*] (sackoo@ornl.gov), Ilenne Del Valle Kessra[1*] (delvallekeid@ornl.gov), Joshua K. Michener,[1] Carrie A. Eckert,[1] and **Paul E. Abraham[1]**

**Institutions:** [1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

**Website URL:** https://seed-sfa.ornl.gov/

**Project Goals:** The Secure Ecosystem Engineering and Design (SEED) Science Focus Area (SFA), led by Oak Ridge National Laboratory, combines unique resources and expertise in the biochemistry, genetics, and ecology of plant-microbe interactions with new approaches for analysis and manipulation of complex biological systems. The long-term objective is to develop a foundational understanding of how non-native microorganisms establish, spread, and impact ecosystems critical to U.S. Department of Energy missions. This knowledge will guide biosystems design for ecosystem engineering while providing the baseline understanding needed for risk assessment and decision-making.

**Abstract Text:**

Plant-microbe interactions in the rhizosphere are essential to ensure plant health and productivity. Recently, commercial biofertilizers and biopesticides have become available to enhance plant growth and pest resistance. However, these products have inconsistent performance as the microbes present in these products fail to survive and proliferate when competing against native microbes present in the rhizosphere. Bacteria have multiple traits encoded in their genome that provide them with advantages or disadvantages under different field conditions. For this reason, we need to investigate the genetic determinants for microbial establishment and persistence in the rhizosphere at a genome-wide scale.

In this project, we are applying two high-throughput methods, bacterial genome shuffling and CRISPR-mediated genome editing, to interrogate bacterial genotype-phenotype relationships. We use as a model system the bacterium *Bacillus velezensis,* a sporulating Gram-positive bacterium that has been shown to promote plant growth by secreting beneficial secondary metabolites and acting as antagonist of pathogenic fungi.

First, we are using genome shuffling to recombine isolates of *B. velezensis* and construct a strain panel for bacterial quantitative trait-locus (QTL) mapping. We have demonstrated genome shuffling in *Bacillus velezensis* FZB42, an important and robust root colonizer, and are extending this approach to investigate recombination between strains of *B. velezensi*s. Based on greenhouse studies conducted in the SEED SFA, we are prioritizing parental *B. velezensis* strains with variable plant phenotypes and developing genetic tools for these strains to enable shuffling.

Our second approach uses CRISPR-Cas mediated genome editing to obtain genome-wide knockout/knockdown libraries. By knocking out/down different genes, we can study how the loss of a function impacts microbial interactions with plants and the environment. We are validating Cas9 and dead Cas9 (dCas9) library generation in the Gram-positive model organism *B.*

*subtilis* using a library targeting 10-15 gRNA per annotated coding region. These experiments will be used to train a computational model of CRISPR Cas efficiency to predict successful gRNA integration and cutting in other bacterial genomes. In parallel, we are expanding the genetic tools available for *B. velezensis* by improving transformation efficiency via natural competency and electroporation. We are also characterizing genetic parts (e.g., promoters, origins of replication, etc.) to engineer a robust Cas9, dCas9, and gRNA expression system in this non-model organism.

Genome-scale genetics in *B. velezensis* will improve our understanding of the genetic factors affecting microbial establishment, persistence, and functionality. This information will be helpful to predict the effect of biostimulants on plant health and productivity.

**Funding Statement**: The Secure Ecosystem & Engineering Design Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research, under FWP ERKPA17. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DEAC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123

**Title:** Development of CRISPR-Cas editing tools in *Sphaerulina musiva* towards controlling its establishment and pathogenicity in *Populus* ecosystems

**Authors:** Joanna Tannous[1]* (tannousj@ornl.gov), Cole Sawyer,[1,2] David Kainer,[1] Alyssa Carrel,[1] Mindy Clark,[1] Jesse Labbe,[1,3] Wellington Muchero,[1] Melissa A. Cregger,[1] Carrie Eckert,[1] and **Paul Abraham[1]**

**Institutions:** [1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN; and [3]Current address: Invaio Sciences, Cambridge, MA

**Website URL:** https://seed-sfa.ornl.gov/

**Project Goals:** The Secure Ecosystem Engineering and Design (SEED) Science Focus Area (SFA), led by Oak Ridge National Laboratory, combines unique resources and expertise in the biochemistry, genetics, and ecology of plant-microbe interactions with new approaches for analysis and manipulation of complex biological systems. The long-term objective is to develop a foundational understanding of how non-native microorganisms establish, spread, and impact ecosystems critical to U.S. Department of Energy missions. This knowledge will guide biosystems design for ecosystem engineering while providing the baseline understanding needed for risk assessment and decision-making.

**Abstract Text:**

The genus *Populus* are economically important biofuel crops cultivated worldwide, but mainly in the Northern hemisphere to fulfill the demands for bioenergy and fiber production. Poplars and their hybrids' widespread distribution and usage is limited by their vulnerability to various diseases, of which the leaf spot and canker disease caused by the invasive fungal pathogen *Sphaerulina musiva* is the most detrimental. Human-mediated transport and introduction has and will continue to result in the establishment and spread of this invasive species throughout the United States. For this reason, we need to understand the biotic and abiotic determinants for the establishment, spread, and virulence of *S. musiva -Populus* pathosystem.

In this project, we established the first CRISPR-Cas9 gene-editing protocol to successfully transform *S. musiva*. We are leveraging from this established tool to advance our understanding of genomic factors affecting above- and below-ground establishment of *S. musiva* and its virulence on *Populus* trees. Firstly, we generated and validated knockout strains of *S. musiva*, to examine the role of the effector gene *ecp2* in the pathogenicity behavior of this fungus on *Populus.* This has been done through pathogenicity experiments conducted on detached *Populus* leaves. Later, we established a closed system *in planta* to identify and characterize more genetic markers implicated in the establishment of *S. musiva* within native *Populus* soil microbial communities through RNA sequencing analysis. Ultimately, transcriptomic data from this experiment, will be complemented

with additional knockouts to confirm the function of the identified genetic markers in the establishment of this pathogen within the native soil microbiota.

**Funding Statement:** The Secure Ecosystem & Engineering Design Science Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research, under FWP ERKPA17. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

# Biofilm Distribution in a Porous Medium Reactor Simulating Shallow Subsurface Conditions

K. Massey[3], H.J. Smith[3], S. Altenburg[3] (sara.altenburg@montana.edu), H. Dreesbach[3], A.B. Cunningham[3], K. Walker[4], A. Putt[4], E. Dixon[4], D.C. Joyner[4], S. Gushgari-Doyle[1], K.G. Cagugao[1], B. Enalls[1], Y. Fan[6], Y. Fu[6], D. Ning[6], R. Chakraborty[1], M.W. Fields[3], T. C. Hazen[4,5], J. Zhou[6], A.P. Arkin[1,2] and **P.D. Adams[1,2]**

**[1]Lawrence Berkeley National Lab, Berkeley**; [2]University of California at Berkeley; [3]Montana State University, Bozeman; [4]University of Tennessee, Knoxville; [5]Oak Ridge National Lab, Oak Ridge; [6]University of Oklahoma, Norman.

*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

Microorganisms in the terrestrial subsurface play important roles in nutrient cycling and degradation of anthropogenic contaminants, functions essential to the maintenance of healthy aquifers. Microorganisms have the potential to change the geochemical properties of the shallow terrestrial subsurface, and previous studies have uncovered significant roles microorganisms can play in groundwater processes, such as biogeochemical cycling. Much of the attention given to the shallow terrestrial subsurface has been focused on the effects of contamination and how microorganisms function in these systems, with far less emphasis on understanding how hydraulic variables influence subsurface microbial ecology. To fully understand how environmental factors impact microbial community dynamics, interactions, succession, colonization, and dispersal in the shallow subsurface environment it is essential to understand the link between microbiology and hydrology. An up-flow packed bed reactor (PBR) was designed to simulate select field conditions (*i.e.*, flow rate and particle size) observed at the Oak Ridge National Laboratory-Field Research Center (ORNL-FRC) to observe how environmental factors, including hydraulic variables such as average pore velocity, influences metabolic activity, community establishment, and cell distribution in a micropore environment. The goals were to understand how environmental variables impact distribution and metabolic activity of microbial cells in a pore microenvironment using native sediment bug trap material under hydraulic properties based upon field conditions (flow rate and particle size). The PBR contained a porous medium of silica oxide particles (74-300 μm), and the size range was based upon particle size assessment of sediment material from the ORNL-FRC. The water phase of the system was a basal groundwater medium that contained low levels of sugars, amino acids, and

nucleosides/nucleotides as the C and N sources that were based upon metabolomic characterization of sediment extracts from the ORNL-FRC. The inocula for the PBRs consisted of sediment material in samplers that were incubated down-well and retrieved from three FRC wells each at distinct pH values (4, 6.3, or 7). The three PBRs were run in parallel with a steady-state flow rate that resulted in an average pore velocity of 0.313 cm/h. Following 4 months of incubation, biomass, cell concentrations, cell distribution, and microbial community analysis for each reactor were evaluated. The pH 4 reactor had the largest biomass and highest activity but had the lowest diversity amongst the pH conditions. The two circumneutral reactors (pH 7 and 6.3) had lower biomass concentrations and activity but had microbial communities that were more diverse than pH 4. The reactors showed different trends in how microbial biomass was distributed through the porous medium as well as distances to other cells and/or cell aggregates. The measured distances were also compared to substrate concentrations over distances predicted by a model based upon diffusion coefficients for molecular classes (*i.e.*, sugars, amino acids, nucleotides/nucleosides). Overall, the data and predictions demonstrate that under *ex situ* conditions meant to simulate porous media flow (*e.g.*, porosity, flow, particle size) at the ORNL-FRC, cells that are part of a diverse microbial community can be on average 20 to 80 μm apart with an average of 2 to 9 cells/particle. Based on diffusivity of potential substrates and measured distance ranges between cells, sugar levels could be approximately 5 to 20 μM whereas amino acids and nucleotides/nucleosides would likely be at sub-micromolar levels between nearest cell/aggregate neighbors. Furthermore, we developed methods to visualize the localization of active and non-active cells within the porous medium and the PBR was able to enrich predominant populations observed in the field. The results have implications for elucidating the impacts of environmental factors on metabolic activity and cell distributions in an impacted, subsurface environment.

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Dynamics of nitrate reduction through process partitioning in a synthetic community**

A.E. Otwell[4], A.V. Carr[3,4,*] (avcarr@uw.edu), K.A. Hunt[4], J. Wilson[3], J.J. Valenzuela[3], H.J. Smith[5], M. De Raad[1], C.J. Petzold[1], S. Turkarslan[4], R. Chakraborty[1], T.R. Northen[1], M.W. Fields[5], N.S. Baliga[3,4], D.A. Stahl[4], A.P. Arkin[1,2] and **P.D. Adams[1,2]**

[1]Lawrence Berkeley National Lab, Berkeley; [2]University of California at Berkeley; **[3]Institute for Systems Biology, Seattle; [4]University of Washington, Seattle;** [5]Montana State University, Bozeman.

*http://enigma.lbl.gov*

**Project Goals: ENIGMA - Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods. Denitrification is presumed to be a major anaerobic microbial process in the low-oxygen zones of the contaminated subsurface of the Oak Ridge Field Research Center (FRC) due to the high concentrations of nitrate. To simulate in the laboratory the exchange of nitrogen intermediates presumed to be occurring in this and other subsurface environments, we established a denitrifying co-culture composed of isolates from this field site. This co-culture provides a controlled system to help inform the ecological, energetic, and functional significance of denitrification pathway complementation. In this and future work, this model system is being utilized to address fundamental questions regarding the importance of nitrogen intermediate exchange in denitrifying populations. Key questions include the relevance of this type of metabolic specialization to population dynamics, community assembly, gene expression, resource partitioning, and system resiliency.**

**Abstract:**

Perturbations, whether from natural phenomena or anthropogenic pollutants, can have severe effects on community function and ecosystem health. In polluted environments, such as the Fields Research Center (FRC), changes in community state can disrupt the cycling of nutrients like nitrogen and lead to the production of the greenhouse gases such as nitrous oxide. Here we highlight efforts to characterize the physiology of a nitrate reducing synthetic community and the environmental implications of pathway partitioning in the context of community function and resilience. We show that partitioning of the denitrification pathway into smaller functional units appears to be a common strategy among organisms isolated from diverse environments. We also show how two particular isolates obtained from the FRC, *Rhodanobacter R12* and *Acidovorax 3H11*, function independently and as a community in nitrate reducing conditions. Independently,

neither organism can completely reduce nitrate to nitrogen gas but together they complement one another to perform complete denitrification. Using kinetic models, transcriptomics, and estimates of community relative abundance we show how *Rhodanobacter R12* and *Acidovorax 3H11* are able to cooperate to enhance community growth and perform complete denitrification. We highlight how each organism responds to variations in nitrate concentration and the role of nitrite in modulating species growth. We also highlight efforts to develop and refine metabolic models for each organism and the insights these models provide. This model community for bidirectional metabolite exchange is now being used as a basis for evaluating biotic and abiotic controls of denitrification in the contaminated subsurface of the Oak Ridge FRC.

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

# Linking Community Composition and Resource Utilization

Y. Wang[1]* (Ywang10@lbl.gov), M. de Raad[1]* (MdeRaad@lbl.gov), Y. Chen[1], S.M. Kosina[1], B.P. Bowen[1], R. Chakraborty[1], C.J. Petzold[1], T.R. Northen[1], A.P. Arkin[1,2] and **P.D. Adams[1,2]**
[1]Lawrence Berkeley National Lab, Berkeley; [2]University of California at Berkeley
*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

There is an urgent need to improve our understanding of the connections between microbial community composition to their *in situ* activities. Exometabolomic methods provide direct measures of metabolite use and production and these data are increasingly available for diverse collections of isolates. However, the degree that isolate metabolic activities are conserved in communities is unclear due to the difficulty in measuring growth rates and resource use in mixed communities. Mass spectrometry based proteomics is routinely used for bacterial protein profiling and can be used for the identification of bacteria. Here, we developed a defined medium, the Northen Lab Defined Medium (NLDM), to characterize bacterial isolates and use protein profiling to track both community structure and resource use partitioning among community members.

NLDM not only supports the growth of diverse soil bacteria but also is defined and therefore suited for exometabolomic profiling. Metabolites included in NLDM were selected based on their presence in R2A medium and soil collected from the Oak Ridge Field Research Center (ORFRC), elemental stoichiometry requirements, as well as knowledge of metabolite usage by different bacteria [1]. NLDM supported the growth of 108 of the 110 phylogenetically diverse soil bacterial isolates (spanning 36 different families) tested and all of its metabolites were trackable through liquid chromatography mass spectrometry analysis.

Next, we constructed a five-member bacterial Synthetic Community (SynCom) based on our previous analysis of the translationally-active microbes present in soil samples from the ORFRC [2]. Resource utilization of the 5 individual SynCom members as well as the SynCom itself was examined using NLDM and exometabolomic profiling. Simultaneously, quantification of the community structure on NLDM was determined using proteomics.

The combination of exometabolomic profiling, using the newly developed NLDM medium, and proteomics enables simultaneous monitoring of population dynamics and substrate use partitioning in mixed communities. Next steps will include integrating stable isotope probing (SIP) into the current setup to track substrate usage by each individual member within the community.

**References**

[1] Jenkins, S. et al. Construction of Viable Soil Defined Media Using Quantitative Metabolomics Analysis of Soil Metabolites. Front. Microbiol. 8, 2618 (2017).

[2] Couradeau, E. et al. Probing the active fraction of soil microbiomes using BONCAT-FACS. Nat. Commun. 10, 1–10 (2019).

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

# Investigating Microbial Ecophysiology of Aerobic and Anaerobic Microbes Through the ENIGMA Environmental Atlas

Romy Chakraborty[1], Brandon C. Enalls,[1] (bcenalls@lbl.gov) Sara Gushgari-Doyle,[1] Mon Oo Yee, Lauren M. Lui,[1] Torben N. Nielsen,[1] Xiaoqin Wu,[1] Farris L. Poole II,[2] Michael W.W. Adams,[2] T. C. Hazen,[3] A.P. Arkin[1,4] and **P.D. Adams[1,4]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA; [3] University of Tennessee, Knoxville TN
[4]University of California, Berkeley, CA

*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

Microbes have diverse and innovative metabolisms that allow for the transformation of carbon and other geochemically relevant elements in the environment. Understanding how these metabolisms intimately link microbes to their habitats is critical for understanding geochemical processes in the environment. As part of ENIGMA, we take an integrated approach to investigate microbes catalyzing key ecophysiological processes at the Oak Ridge Reservation Field Research Center (ORR-FRC) including diverse cultivation strategies to obtain new isolates and supplement our increasing strain repository, genomic characterization of newly obtained isolates to inform metabolic poise, and physiological characterization to confirm metabolic functions *in vivo*.

Here, we highlight two ENIGMA Environmental Atlas projects that exemplify our research goals and interdisciplinary approaches. First, we present an analysis of seven distinct strains of a single genus, *Arthrobacter*, isolated from varying depths of a single ORR FRC sampling location and adjacent groundwater well, linking niche ecotype to genotype and phenotype. *Arthrobacter* are a highly prevalent aerobic subsurface bacterial genus known for complex carbon metabolisms, and are abundantly present at ORR-FRC. Through analysis of circularized high-quality genomes and laboratory phenotyping, we were able to group these isolates into several ecotypes that demonstrate functional and genomic capacities related to carbon degradation that are specific to their biogeochemical conditions of origin, and that genes found in plasmids and predicted genomic islands support several of these functions. We then performed a pangenome analysis to compare the genomic diversity of our strains to other environmental *Arthrobacter* isolates from diverse environments. These data suggest that volatile accessory genomes of *Arthrobacter* sp. likely contribute to the genus' high potential for adaptability in the environment.

Second, we report on our efforts towards understanding the influence that minerals have on anaerobic microbial communities that impact carbon transformations in the subsurface. Methanogenic archaea in the

subsurface tend to be outcompeted for electron donors by other anaerobes such as sulfate-reducing bacteria. However, geochemical analysis of sediment cores collected from ORR FRC suggest the coexistence of methanogenesis and sulfate reduction within similar depth horizons, where semiconductive iron minerals were also abundant. We hypothesized that iron minerals may serve as an electron conduit that supports methanogenesis through extracellular electron transport despite the presence of competing sulfate-reducing bacteria. Amending ORR FRC sediments with iron minerals ferrihydrite and hematite using butyrate as the primary electron donor, we observed that the presence of hematite stimulated both methanogenesis and sulfate reduction whereas ferrihydrite only stimulated sulfate reduction. We posit that sulfide-derived electrons are being transferred to methanogens through hematite mineral conduits. Future work examining both the composition of the microbial communities *via* 16S analysis as well as the redox and coordination state of the iron minerals using synchrotron-based methods will help us better understand how conductive minerals aid in sustaining complex and competing microbial metabolisms in nature. These projects demonstrate two examples of the research performed with aerobic and anaerobic bacteria from ORR-FRC under the aims of the ENIGMA Environmental Atlas.

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

# Characterization of a Nitrate-Respiring, Multi-Metal-Resistant *Bacillus* Species Highly Abundant in Heavily Contaminated ORR FRC Subsurface

J.L. Goff[1*] (JLGoff@uga.edu), E.G. Szink[1], M.P. Thorgersen[1], L.M. Lui[2], T.N. Nielsen[2], F.L. Poole II[1], A.D. Putt[3], Y. Fan[4], J.P. Michael[4], Y. Wang[4], D. Ning[4], Y. Fu[4], J.D. Van Nostrand[4], E.R. Kelly[3], K.A. Lowe[5], M. Rodriguez Jr.[5] Y. Chen[2], C.J. Petzold[2], J.-M. Chandonia[2], T. C. Hazen[3,5], J. Zhou[4], M.W.W. Adams[1], A.P. Arkin[2,6], and **P.D. Adams[2,6]**

[1]University of Georgia, Athens; [2]Lawrence Berkeley National Lab, Berkeley; [3]University of Tennessee, Knoxville; [4]University of Oklahoma, Norman; [5]Oak Ridge National Lab, Oak Ridge [6]University of California at Berkeley

*http://enigma.lbl.gov*

**Project Goals: ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) is using a systems biology approach to understand the interactions between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field and computational methods.**

A significant goal of microbial ecology is the development of models that accurately predict how environmental perturbations will impact the structure and function of microbial communities. Identification of representative isolates is essential for the high-resolution physiological studies required for proper parameterization of these models. Frequently, pure-culture laboratory studies fail to select appropriate isolates that are relevant to an environment either in terms of taxonomic abundance or metabolic function. Here, we present the successful isolation and characterization of a *Bacillus cereus* strain (designated CPTF) that is extremely abundant in the subsurface of the Oak Ridge Reservation Field Research Center (ORR FRC) that is immediately adjacent to the highly-contaminated S-3 ponds (Area 3). The S-3 ponds contain high levels of legacy uranium, nitric acid, and other mixed metals waste that has leached into the surrounding environment. The CPTF 16S rRNA gene has 100% sequence identity to the most abundant ASV observed across an ORR FRC Area 3 subsurface metagenome survey. The CPTF-matching ASV is present in all 32 subsurface samples collected from Area 3. Its abundance across individual samples ranges from 0.02 – 40.5% with a median abundance of 3.8%.

The ORR FRC Area 3 subsurface poses a significant physiological challenge to microorganisms living at the site—with high concentrations of nitrate, mixed metals waste, and low pH. The porewaters of the soils contain nitrate at concentrations up to 19,600 ppm and pH values as low as 3.0. Notably, the CPTF-matching ASV is observed across the full span of measured porewater nitrate concentrations. Furthermore, as the CPTF-matching ASV is ubiquitous across the sampled Area 3 subsurface, it also exists concurrent with elevated levels of

uranium (up to 1130 ppm) and other metals measured in the same survey. On-going work seeks to understand the persistence of CPTF at the ORR FRC despite these significant environmental stressors. Experimental work has shown that CPTF grows robustly with high site-relevant levels of nitrate and high concentrations of nitrite. CPTF respires nitrate with nitrite as the major end-product. Phenotypic observations and genome analysis suggest that nitrite can support fermentative ammonification by this strain. CPTF also has high tolerance to multiple metals at site-relevant concentrations both individually and in combination. The metals tested include U, Al, Ni, Cd. Mn, Fe, Cu, and Co. Genome analysis of CPTF revealed metal efflux pumps, metal-responsive transcriptional regulators, and numerous other metal resistance determinants. High-throughput proteomic analysis is underway to understand the cellular response to ORR groundwater-relevant metals under nitrate-respiring conditions to explore how mixed waste contamination impacts the physiology of this abundant contaminated site isolate. The results of this work will facilitate our understanding of the persistence of dominant taxa at highly-contaminated DOE sites.

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**ENIGMA Long Read Sequencing and Assembly for Microbial Genomes:**
**Current Capabilities and Future for Metagenomics and KBase Integration for Assembly**

Lauren M. Lui*[1] (lmlui@lbl.gov), Torben N. Nielsen*[1] (torben@lbl.gov), John-Marc Chandonia[1], Adam P. Arkin[1,2], **Paul D. Adams[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of California, Berkeley, CA

*https://enigma.lbl.gov*

**Project Goals: ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) uses a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

Achieving a causal understanding of a microbial system requires mapping mechanisms by which organisms grow, cooperate, and compete in complex environments. These mechanisms include ecological phenomena and abiotic factors that influence behavior and survival. One of the critical requirements for reaching this level of understanding is fully resolving the genomes of the community, including plasmids and viruses, so that the functional roles specified by their genomes can be assayed and discovered. While the challenges of gene functional annotation and linking genotype and phenotype loom beyond simply obtaining genomes, the underlying challenge at the present remains to generate high-quality genomes for both isolates and metagenomes. The base genome along with its relative abundance constitute the most important foundational data needed to infer and parameterize models of microbial system dynamics.

Despite advances in sequencing technology throughput, until recently it has been very difficult or impossible to completely finish genomes from both isolates and microbial communities. Finishing genomes is difficult and laborious without the use of reads that are longer than any repetitive DNA element, which can be thousands of base pairs. In addition, for metagenome-assembled genomes, it is challenging to associate plasmids with their host chromosomes. Both these barriers can be overcome using new technologies in long read sequencing from Pacific Biosciences and Oxford Nanopore Technologies. These technologies can produce read lengths in the multiple thousands and can resolve DNA methylation modification. Error rates in calling nucleotide identity have been dropping as well, reducing the need for more accurate (Illumina) short reads for polishing. Long reads enable effective assembly of replicons and can even permit strain-level resolution variants. Methylation profiling can often tie multiple replicons such as genomes and plasmids together in the same organism even when assembled from a complex mixture of organisms. Full genome assemblies wherein all replicons are associated accelerates studies of phylogeny, adaptive evolution, and facilitates better assessment of base genetic capability. While this technology is progressing it still remains challenging to apply it effectively to more complex samples of natural communities and diverse enrichments.

The ENIGMA SFA has spent time developing pipelines for sequencing and assembly of long read data from microbial isolates and metagenomes to help achieve the goal of casual microbial ecology. We have developed the capability to isolate diverse organisms, to extract the high molecular weight (HMW) DNA needed for single molecule long read sequencing, and to perform the sequencing using Oxford Nanopore Technologies MinION sequencers. For example, we have successfully made long read libraries using DNA extracted from ENIGMA groundwater and sediment samples. From one sample we recovered more than 30 finished grade genomes, including two 6 Mb genomes, as well as what we believe to be a bacteriophage larger than 1 Mb, which would make it the largest bacteriophage genome known (currently the largest identified phage is ~735kb). This is from one sample; from nearly all short read metagenome studies there are only approximately 100-200 fully complete microbial genomes.

We developed computational pipelines to process long read sequencing data since ENIGMA is generating next-generation sequencing datasets at scale. To characterize the microbial diversity and activity at the Oak Ridge Reservation at Oak Ridge National Laboratory, ENIGMA anticipates isolating thousands of bacteria and archaea, as well as generating spatio-temporal series of fully resolved enrichments and metagenomes from the site. These sequencing projects assist the goals of linking genotype to phenotype and understanding the temporal, dynamic, and complex factors influencing microbial community structure and activity at our site. ENIGMA uses isolates to help link genotype to phenotype by analyzing genomes in conjunction with transposon mutant libraries, metabolomics, and growth condition data. High quality genomes and metagenomes are essential for these types of experiments and ENIGMA science.

We are currently adding new functionality to DOE Systems Biology KnowledgeBase (KBase) by implementing tools for using long read data for assembly of isolates, assembly of metagenomes, and methylation detection. We are developing workflows within KBase to make these tools more broadly available across the ENIGMA SFA and to other scientists, especially for scientists that do not specialize in computational methods. These apps and workflows will enable ENIGMA, as well as other DOE SFAs and microbiologists to (1) address scientific questions that would otherwise be infeasible with isolate and metagenome assemblies using only short reads, (2) track provenance of data and methods used for assembly, and (3) share assemblies across the SFA for collaborations. By providing this new functionality in KBase, we will also provide a foundation for further extensions in KBase to support developments in long read technology.

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Mobile Genetic elements in the ORFRC: Native plasmids and Viral Genomes**

A. Kothari[1], A. Mukhopadhyay[1,2] **(**amukhopadhyay@lbl.gov**)**, A. Codik[1], S Roux[1], Y-W Yu[1], T. K. Owens[1], L. M. Lui[1], T. N. Nielsen[1], H. Zhang[1], A. Prieto[1], D. Soneja[1], J-M. Chandonia[1], M. Charrier,[1] L. Rajeev,[1] A. M. Rocha,[3] S. J. Spencer[3], X. Wu[1], S. Altenburg[5], S. W. Singer[1], E.J. Alm[3], R. Chakraborty[1], A.M. Deutschbauer[1,2], M.W. Fields[4], T. C. Hazen[5,6]A. P. Arkin[1,2], and **P.D. Adams[1,2]**

**[1]Lawrence Berkeley National Lab**; [2]University of California at Berkeley; [3]Massachusetts Institute of Technology, Cambridge; [4]Montana State University, Bozeman; [5]University of Tennessee, Knoxville; [6]Oak Ridge National Lab, Oak Ridge

*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

Mobile genetic elements such as plasmids and viruses can shape the function and evolution of natural microbial communities. Both viruses and plasmids have, for example, been shown to harbor and transfer genes to microbial hosts that improve microbial fitness in particular environments or conditions. Here, we present a set of studies to identify and functionally characterize mobile genetic elements in ground water microbiomes at the Oak Ridge Reservation study site (ORFRC), which is contaminated with heavy metals and nitrate. First[1], we optimized a cultivation-free method targeting plasmid DNA across a range of sizes and abundance to examine the plasmidome of groundwater samples. Plasmidome sequencing data from just two ground-water wells helped identify > 600 new assembled circular plasmids, many encoding metal and antibiotic resistance genes. In a follow-up study[2], we examined in more detail the surprising presence of genes encoding mercury tolerance, despite mercury not being one of the major contaminants at this site. Many homologs of *mer* genes (e.g., *merA*) were found in several assembled plasmid scaffolds, including a plasmid found to be abundant (based on sequence coverage) in both groundwater samples. Since this abundant plasmid, p5343, was not isolated but rather inferred from assembly of plasmidome sequence, we synthesized this 8kb plasmid to test it experimentally. We developed a mercury tolerance assay in *E. coli* and established that the *mer* genes on p5343 were functional in conferring mercury tolerance. More importantly we found that p5343 was able to naturally transform into several isolates including two Gram-positive bacteria that have not been transformed before. In a third recent study[3] we re-analyzed this large sequence dataset for the presence of phage genomes. In collaboration with the DOE Joint Genome Institute (JGI) we discovered 200 phage genomes, infecting some of the dominant

microbes in the ecosystem. Several of these phage genomes encode predicted metal and antibiotic resistance genes, presumably providing a selective advantage to infected hosts. Taken together, these plasmidome and phageome studies provide a first overview of the diversity of mobile genetic elements in ORFRC groundwater, along with valuable future directions. The current database of > 600 plasmids and 200 viral genomes can be mined for parts such as origins for developing genetic tools for manipulating many model and ORFRC bacterial isolates. Functional studies of genes encoded on these plasmids and viral genomes will reveal key insight into metal and antibiotic resistance in this environment, highlighting functions encoded by mobile genetic elements. Study of additional wells, both ground water and sediment will provide rich information of mobile genes and their potential role in the microbial ecology of this site.

**References**


1. Kothari et al mBio 2019 https://doi.org/10.1128/mBio.02899-18
2. Kothari et al mSystems 2019 https://doi.org/10.1128/mSystems.00588-19
3. Kothari et al mSystems 2021 https://doi .org/10.1128/mSystems.00537-21



*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**The ENIGMA Subsurface Observatory: A high resolution approach to studying a shallow contaminated groundwater system**

A.D. Putt[1]*, (aputt@vols.utk.edu), E.R. Kelly[1], K.F. Walker[1], M. Newcomer[2], M.W. Fields[3], J.L. Goff[4], E.G. Szink[4], M.P. Thorgersen[4], F.L. Poole II[4], Y. Fan[5], J.P. Michael[5], P.J. Walian,[2] D. Ning[5], J.D. Van Nostrand[5], K.A. Lowe[6], M. Rodriguez Jr[6], D.C. Joyner[1], T. C. Hazen[1,6], M.W.W. Adams,[4] J. Zhou[5] A.P. Arkin[2,7] and **P.D. Adams[2,7]**

[1]University of Tennessee, Knoxville; **[2]Lawrence Berkeley National Lab, Berkeley**; [3]Montana State University, Bozeman; [4]University of Georgia, Athens; [5]University of Oklahoma, Norman; [6]Oak Ridge National Lab, Oak Ridge; **[7]University of California at Berkeley**.

*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies uses a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. The field to laboratory efforts of ENIGMA rely on high resolution field data to link genetic, ecological, and environmental factors to the structure and function of microbial communities. The ENIGMA SSO is a critical infrastructure under design to provide unprecedented spatiotemporal resolution for groundwater time series analyses. The SSO is a major component of ENIGMAs three-aim approach for integrating and developing laboratory, field, and computational methods.**

The ENIGMA (*http://enigma.lbl.gov*) subsurface observatory (SSO) is a shallow subsurface study site being designed to investigate the spatiotemporal effects of subsurface perturbations. Once established, the SSO will feature a total of 9 boreholes in a 3 x 3 borehole configuration. Each of the nine boreholes will contain four separate wells screened to different depths. The deepest of wells will be in the saturated zone, two wells will divide the shallow saturated and variably saturated zones, and a single well in the unsaturated vadose zone. The multi-level design enables observations of parallel, hydrologically coupled wells oriented along a flow path. Tracking dynamics across the coupled locations will provide a more reliable estimation of differential dispersal forces and improved detection of the persistent and active subpopulations acting under the different local conditions. Once the physical infrastructure is established, the subsurface observatory will provide a spatiotemporal time series of unprecedented resolution.

The SSO multi-level design provides a physical infrastructure for investigations of depth-dependent changes in the biogeochemistry with a focus on the zone of variable saturation. The SSO will be established in a zone with strong stratification of taxonomic and geochemical composition and a strong impact of hydrogeological forces on vertical mixing and recharge within the variably saturated zone.

A three-dimensional model of the shallow subsurface was constructed from a grid of over 100 cone penetrometer pushes which has been used to identify zones of low and high permeability. The identification of high permeability zones has been used to guide the placement of the SSO wells as a part of the development of a high resolution subsurface observatory. A network of continuous monitoring water levels and a multiparameter sonde has indicated precipitation events as a major source of natural subsurface perturbation. Rain events not only increase the water level, but also increase the specific conductivity, dissolved oxygen, and pH. While local infiltration occurs over the course of hours, continuous monitoring indicates that over half of the recharge may come from regional infiltration transported in the underlying bedrock and fractured transition zone. This regional recharge may be a major mechanism of material transfer into the site and is supported by increased concentrations of contaminants and low pH measured in sediment samples collected above the bedrock transition zone. Flow rates from pump tests, slug tests, soil tests, and permeability estimates from the volumetric three dimensional model all share a high level of agreement that the average flow rates in the unconsolidated material are likely between $10^{-5}$ and $10^{-7}$ m/s. In total, this program aims to further select an ideal site with high similarity in sediment type composition, where well to well connectivity is highly probable, and where differential dispersal forces can be modeled.

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Genome-scale metabolic and regulatory network reconstruction of *Pyroccous furiosus***

Dmitry A. Rodionov [1]*(rodionov@sbpdiscovery.org), Ying Zhang [2]*(yingzhang@uri.edu), Ke Zhang, [2] Jason Vailionis, [2] James R. Crosby,[3] Ryan G. Bing,[3] Hailey O'Quinn, [4] Gina L. Lipscomb, [4] Tania N. N. Tanwee,[4] Robert M. Kelly,[3] and **Michael W. W. Adams[4]**

[1]Sanford-Burnham-Prebys Med. Discovery Institute, La Jolla, CA; [2]University of Rhode Island, Kingston, RI; [3]North Carolina State University, Raleigh, NC; [4]University of Georgia, Athens, GA;

**Project Goals: The goal of this project is to establish the two non-model microorganisms *Caldicellusiruptor bescii* and *Pyroccous furiosus* as platforms for sustainable production of industrial chemicals using renewable plant biomass. We aim to engineer *C. bescii* and P*. furiosus* to reincorporate $CO_2$ and $H_2$ generated during fermentation of lignocellulose as an additional source of carbon and energy, potentially reaching net zero $CO_2$ emission. Furthermore, *P. furiosus* will be engineered with enzymes from *C. bescii* to allow it to degrade non-pretreated plant biomass. System-wide metabolic and regulatory modeling of both organisms will be used to optimize biomass degradation and desired product yield and selectivity.**

*Pyrococcus furiosus* is an extremely thermophilic, strictly anaerobic, sugar-utilizing archaeon microorganism that grows up to 103°C on starch, laminarin, maltose, trehalose, cellobiose and beta-glucan oligosaccharides, but not crystalline cellulose, xylan or monosaccharides In this study, we applied a subsystems-based approach combining comparative genomics, transcriptional regulon prediction, and genome-scale modeling to reconstruct an integrated view of the metabolic and regulatory network of *P. furiosus*. The complete genomes of over 30 *Pyrococcus and Thermococcus* species were used for ortholog mapping and comparative analysis. Functional gene assignments, genome context analysis, comparative analysis of orthologous genes and DNA upstream regions, gene co-occurrence analysis and protein similarity searches were performed in the SEED environment [1]. We also used genome annotations from Swiss-Prot, KEGG, TCDB, and RegPrecise databases and published experimental data. The previously generated RNASeq datasets for whole-genome gene expression and transcriptional start sites obtained for *P. furiosus* grown on different carbon sources (glucose, maltose, cellobiose) were used for validation of reconstructed transcriptional regulons and for refinement of transporter specificities.

The reconstruction and analyses of genome-scale models (GEMs) combines the stoichiometry of metabolic processes with the definition of condition-specific metabolic constraints into predicting microbial growth and biochemical production. The GEM curation was done with the support of PSAMM software [2] to incorporate the known and predicted metabolic functions of enzymes and transporters. We have obtained a first draft of the *P. furiosus* GEM that contains 409 metabolic genes, 860 metabolites (non-unique) and 538 metabolic reactions, thereby covering 81% of genes with EC number assignments. These include a native RuBisCo enzyme that allows *P. furiosus* to incorporate $CO_2$ into its central carbon metabolism, a diverse range of catabolic and anabolic pathways, and biomass production equations that incorporate the experimentally calibrated proportion of major cell components. Besides the genes encoding metabolic enzymes, the *P. furiosus* GEM contains 198 genes encoding components of metabolic transporters (89% of total number of putative transporter genes). Growth predictions made by the *P. furiosus* GEM were validated by matching metabolic simulations with growth measurements in batch and chemostat culture using defined media. This

model, alongside our preexisting model of *C. besci* [3], serves as a stepping stone for the engineering of future strains to enable and enhance the yields of bio-based fuels and chemicals.

Five industrial compounds have been selected as engineering targets with the potential to achieve net zero carbon emission. We have developed strategies for the production of these compounds by incorporating enzymes from other thermophiles, as well as engineering the *P. furiosus* RuBisCo into *C. bescii*. Simulations using the two GEMs demonstrated that both *P. furiosus* and *C. bescii* can theoretically achieve negative $CO_2$ production, while generating products when $CO_2$ and $H_2$ are supplied. Additionally, enhanced product formation has been predicted by simulation of multi-gene knockouts and introduction of hydrogenases from *P. furiosus* to *C. bescii*.

Previously, we used a comparative genomics approach to reconstruct the carbohydrate utilization regulatory network in *C. besci* and related bacteria [4]. Here, we used the same approach to identify DNA-binding motifs and reconstruct regulatory networks for 15 out of 65 TFs encoded in *P. furiosus* genome. Most of these TFs are local regulators, while two reconstructed regulons (SurR and TrmBL1) include genes from multiple metabolic pathways. TrmBL1 is a global regulator of the carbohydrate metabolism that co-regulates large sets of genes involved in the starch/maltose utilization, glycolysis and gluconeogenesis. The reconstructed metabolic and regulatory networks are used to generate metabolic models for *C. besci* and *P. furiosus*, and to guide engineering strategies to improve bioproduct formation. The goal is to increase carbon flux from substrates to engineered products, which we hypothesize is key to producing industrially relevant titers and yields.

**IMPORTANCE** In this study, we developed a predictive model for simulating the metabolism of the non-model organism, *P. furiosus*. The simulation predictions can provide potential directions for the more efficient metabolic engineering design for bio-based chemicals. We will combine modeling and regulatory mechanisms and use extensive experimental validation to enable model parameterization. The systems-wide integration of models at the metabolic and regulatory levels will enable discovery of “non-intuitive” designs that may not be apparent from knowledge-based optimization of targeted pathways.

**References**

1. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 42:D206-14. doi: 10.1093/nar/gkt1226.

2. Steffensen JL, Dufault-Thompson K, Zhang Y (2016) PSAMM: A Portable System for the Analysis of Metabolic Models. *PLoS Comput Biol* 12. doi:10.1371/journal.pcbi.1004732.

3. Zhang K, Zhao W, Rodionov DA, Rubinstein GM, Nguyen DN, Tanwee TNN, Crosby J, Bing RG, Kelly RM, Adams MWW, Zhang Y. (2021) Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for bio-based chemical production. *mSystems*. 6(3):e0135120. doi: 10.1128/mSystems.01351-20.

4. Rodionov DA, Rodionova IA, Rodionov VA, Arzamasov AA, Zhang K, Rubinstein GM, Tanwee TNN, Bing RG, Crosby JR, Nookaew I, Basen M, Brown SD, Wilson CM, Klingeman DM, Poole FL 2nd, Zhang Y, Kelly RM, Adams MWW. (2021) Transcriptional regulation of plant biomass degradation and carbohydrate utilization genes in the extreme thermophile *Caldicellulosiruptor bescii*. *mSystems.* 6(3):e0134520. doi: 10.1128/mSystems.01345-20.

*This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019391 and DE-SC0022192*

# From fast to furious phenotyping of microbes: advancing the understanding of microbial diversities and gene functions using droplet microfluidics and high throughput sequencing

F. Song[1,*] (fsong@lbl.gov), X. Lyu[2], M. Moore[2], A. Hung[2], J.V. Kuehl[1], H. Carlson[1], M. Price[1], V. Mutalik[1], A.M. Deutschbauer[1], A.P. Arkin[1,2] and **P.D. Adams[1,2]**

**[1]Lawrence Berkeley National Lab, Berkeley**; [2]University of California at Berkeley.

*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

Knowing the phenotypes of diverse microbial species and mutations is essential for understanding their ecological roles and genetic functions. Previous methods for microbial phenotyping use fluorescence or optical density ($OD_{600}$) as the readout and 96 or 384 well plates as the containers. These methods became the standard and have been very popular. Although throughput has been dramatically increased by using lab automation, it does not meet the current demands. Thousands of microorganisms are now routinely isolated by environmental microbiology programs and are in need of basic phenotyping. Large-scale libraries of genetic variations of these, made to characterize and engineer their function, similarly are in dire need of rapid characterization. In addition, the well-plate-based method is expensive- consuming high quantities of disposable plasticware and reagents- and plagued by cross contaminations amongst the wells on the same plate. The contaminants could not be identified by the optical based readout. It is imperative to accelerate the phenotyping speed, decrease the cost and plastic utility, and avoid the effects of cross contamination.

To address problems of scale, we developed a droplet-based high throughput platform for microbial phenotyping using genetic marks as the readout and water-in-oil droplets as containers. The new method consists of droplet generation, droplet cultivation, cell pooling, spike-in controls addition for abundance quantification, DNA extraction, high throughput sequencing, and associated bioinformatic pipelines for data analysis. Briefly, single cells from a mix of hundreds of microbial species (or mutants) are captured in individual water-in-oil droplets. They are allowed to grow within these droplets for some period. The emulsion is then broken, and droplet contents are pooled along with a spike-in standard (a microbe or gene which is not in the pool of microbes). Sequencing of the population before and after this operation enables comparison of the relative abundances of each microbe at the beginning and in the end of the cultivation. In essence, the fitness of hundreds of microbes under the tested condition is obtained by amplicon sequencing in one set-up. Since the system is miniaturized and scalable, we were able to screen many different conditions in parallel using a multichannel syringe/pressure pump. We have demonstrated the efficacy of this method by comparing results from classical plate-based assays to our droplet method for seventeen well characterized organisms under the conditions of R2A and Xylose. We are then applying this technology to physiologically

characterize ENIGMA field isolates under conditions predicted to be important for field processes.

In addition to characterizing microbial isolates, we also combined this method with Random Barcode Transposon-site Sequencing (RB-TnSeq)[1] to study the gene functions of microbes during phage infection. As a proof of principle, we assayed an *E. coli* RB-TnSeq library with two *E. coli* phages (T4 and N4) to discover genes which confer resistance or sensitivity. Our droplet results are consistent with the findings from bulk assays[2] and also uncovered few new hits that were not seen in bulk assays. The droplet assays are useful in identifying medium strength resistance mutants that may be lost in bulk assays because of stronger selection pressure.

The droplet-based high throughput phenotyping method will enable us to link microbial genotypes with phenotypes, and further achieve a mechanistic understanding of field data leading to pointed hypotheses of the role of particular taxa, proteins, co-factors and their interactions in creating systems behavior.

## References


1. Wetmore, Kelly M., et al. "Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons." MBio 6.3 (2015): e00306-15.
2. Mutalik, Vivek K., et al. "High-throughput mapping of the phage resistance landscape in E. coli." PLoS biology 18.10 (2020): e3000877.



*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Energy and Carbon Optimized Conversion of Lignocellulose to Biobased Chemicals by Extreme Thermophiles**

Tania N.N. Tanwee (tt57098@uga.edu),[1*] James Crosby,[2] Hailey O'Quinn,[1] Gina L. Lipscomb,[1] Ryan G. Bing,[2] Tunyaboon Laemthong,[2] Ke Zhang,[3] Jason Vailionis,[3] Ying Zhang,[3] Dmitry Rodionov,[4] Robert M. Kelly,[2] and **Michael W. W. Adams[1]**

[1]University of Georgia, Athens, GA; [2]North Carolina State University, Raleigh, NC; [3]University of Rhode Island, Kingston, RI; [4]Sanford-Burnham-Prebys Med. Discovery Institute, San Diego, CA

**Project Goals: The goal of this project is to establish the two non-model microorganisms *Caldicellulosiruptor bescii* and *Pyrococcus furiosus* as platforms for sustainable production of industrial chemicals using renewable plant biomass. We aim to engineer *C. bescii* and *P. furiosus* to reincorporate $CO_2$ and $H_2$ generated during fermentation of lignocellulose as additional sources of carbon and energy, to potentially reach net zero $CO_2$ emission. Furthermore, *P. furiosus* will be engineered with enzymes from *C. bescii* to allow it to degrade non-pretreated plant biomass. System-wide metabolic and regulatory modeling of both organisms will be used to optimize biomass degradation and desired product yield and selectivity.**

The extreme thermophiles *Caldicellulosiruptor bescii* ($T_{max}$ = 90°C) and *Pyrococcus furiosus* ($T_{max}$ = 103°C) share the distinction within the bacterial and archaeal domains, respectively, of being the most thermophilic member having a functional genetic system. Over the last decade, *C. bescii* has been metabolically engineered to produce ethanol, acetone, and various alcohols (1-3) and *P. furiosus* has been engineered to produce lactate, ethanol, and 3-hydrxypropionate (4-6). Fermentation at high temperatures reduces the risk of contamination, facilitates product recovery, minimizes cooling costs, and enhances biomass solubilization (7). These advantages, combined with the native ability of *C. bescii* to deconstruct non-pretreated lignocellulosic biomass (8-11), offer environmentally sustainable, economic platforms to engineer production of industrial products. Our labs have made recent advancements toward strengthening the knowledge-base of *C. bescii* by defining the regulatory pathways of carbohydrate utilization (8) and constructing a genome-scale metabolic model (12). We aim to improve our understanding of *P. furiosus* by applying similar approaches. One aspect of the experimental work is focused on engineering *P. furiosus* with (hemi)cellulases from *C. bescii* to enable its growth on lignocellulosic substrates, since *P. furiosus* can utilize cellobiose but not cellulose or xylan (13). Additional strains of *C. bescii* and *P. furiosus* are under construction to produce industrially-relevant compounds. A key aspect of the engineering strategy is to recycle $CO_2$ and $H_2$ generated from lignocellulose fermentation. Target chemicals have been selected strategically to use the $CO_2$ generated during one biomass-to-bioproduct formation for incorporation into a second bioproduct. This biomass-to-bioproduct conversion will be enhanced by regenerating additional redox cofactors from $H_2$. To accomplish this experimentally, we will leverage the carbon fixing and hydrogenase enzymes from *P. furiosus* in genetic engineering of *C. bescii*. The genome-scale metabolic models of both *C. bescii* and *P. furiosus* are being used to evaluate each

prospective engineered pathway to enhance production capacity of target products while optimizing carbon and electron fluxes.

References:


1. Williams-Rhaesa, A. M., Rubinstein, G. M., Scott, I. M., Lipscomb, G. L., Poole Ii, F. L., Kelly, R. M., and Adams, M. W. W. (2018) Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii*. *Metab Eng Commun* **7**, e00073
2. Straub, C. T., Bing, R. G., Otten, J. K., Keller, L. M., Zeldes, B. M., Adams, M. W. W., and Kelly, R. M. (2020) Metabolically engineered *Caldicellulosiruptor bescii* as a platform for producing acetone and hydrogen from lignocellulose. *Biotechnol Bioeng* **117**, 3799-3808
3. Rubinstein, G. M., Lipscomb, G. L., Williams-Rhaesa, A. M., Schut, G. J., Kelly, R. M., and Adams, M. W. W. (2020) Engineering the cellulolytic extreme thermophile *Caldicellulosiruptor bescii* to reduce carboxylic acids to alcohols using plant biomass as the energy source. *J Ind Microbiol Biotechnol* **47**, 585-597
4. Hawkins, A. B., Lian, H., Zeldes, B. M., Loder, A. J., Lipscomb, G. L., Schut, G. J., Keller, M. W., Adams, M. W. W., and Kelly, R. M. (2015) Bioprocessing analysis of *Pyrococcus furiosus* strains engineered for $CO_2$-based 3-hydroxypropionate production. *Biotechnol Bioeng* **112**, 1533-1543
5. Keller, M. W., Lipscomb, G. L., Loder, A. J., Schut, G. J., Kelly, R. M., and Adams, M. W. W. (2015) A hybrid synthetic pathway for butanol production by a hyperthermophilic microbe. *Metab Eng* **27**, 101-106
6. Keller, M. W., Lipscomb, G. L., Nguyen, D. M., Crowley, A. T., Schut, G. J., Scott, I., Kelly, R. M., and Adams, M. W. W. (2017) Ethanol production by the hyperthermophilic archaeon *Pyrococcus furiosus* by expression of bacterial bifunctional alcohol dehydrogenases. *Microb Biotechnol* **10**, 1535-1545
7. Straub, C. T., Counts, J. A., Nguyen, D. M. N., Wu, C. H., Zeldes, B. M., Crosby, J. R., Conway, J. M., Otten, J. K., Lipscomb, G. L., Schut, G. J., Adams, M. W. W., and Kelly, R. M. (2018) Biotechnology of extremely thermophilic archaea. *Fems Microbiol Rev* **42**, 543-578
8. Rodionov, D. A., Rodionova, I. A., Rodionov, V. A., Arzamasov, A. A., Zhang, K., Rubinstein, G. M., Tanwee, T. N. N., Bing, R. G., Crosby, J. R., Nookaew, I., Basen, M., Brown, S. D., Wilson, C. M., Klingeman, D. M., Poole, F. L., Zhang, Y., Kelly, R. M., Adams, M. W. W., and Summers, Z. M. (2021) Transcriptional regulation of plant biomass degradation and carbohydrate utilization genes in the extreme thermophile *Caldicellulosiruptor bescii*. *mSystems* **6**, e01345-01320
9. Lee, L. L., Crosby, J. R., Rubinstein, G. M., Laemthong, T., Bing, R. G., Straub, C. T., Adams, M. W. W., and Kelly, R. M. (2020) The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in 'Caldi World'. *Extremophiles* **24**, 1-15
10. Straub, C. T., Khatibi, P. A., Wang, J. P., Conway, J. M., Williams-Rhaesa, A. M., Peszlen, I. M., Chiang, V. L., Adams, M. W. W., and Kelly, R. M. (2019) Quantitative fermentation of unpretreated transgenic poplar by *Caldicellulosiruptor bescii*. *Nat Commun* **10**
11. Basen, M., Rhaesa, A. M., Kataeva, I., Prybol, C. J., Scott, I. M., Poole, F. L., and Adams, M. W. W. (2014) Degradation of high loads of crystalline cellulose and of unpretreated plant biomass by the thermophilic bacterium *Caldicellulosiruptor bescii*. *Bioresource Technol* **152**, 384-392
12. Zhang, K., Zhao, W., Rodionov, D. A., Rubinstein, G. M., Nguyen, D. N., Tanwee, T. N. N., Crosby, J., Bing, R. G., Kelly, R. M., Adams, M. W. W., and Zhang, Y. (2021) Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for bio-based chemical production. *mSystems* **6**, e0135120
13. Koning, S. M., Elferink, M. G. L., Konings, W. N., and Driessen, A. J. M. (2001) *C*ellobiose uptake in the hyperthermophilic archaeon *Pyrococcus furiosus* is mediated by an inducible, high-affinity ABC transporter. *J Bacteriol* **183**, 4979-4984



This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0022192.

**Investigating the abiotic control of denitrification processes using synthetic communities and laboratory simulations**

Jacob J. Valenzuela[1]* (jvalenzu@isbscience.org), J. Wilson[1], S. Turkarslan[1], H. Smith[2], A. Otwell[3], K. Hunt[3], F. Poole[4], Xiaoxuan Grace Ge[4], J.L. Goff[4], M. Thorgersen[4], M. Wells[5], P. Walian[6], A. M. Deutschbauer[6], T. R. Northen[6], M. W.W. Adams[4], R. Chakraborty[6], D. A. Elias[5], D. A. Stahl[3], M. W. Fields[2], N. S. Baliga[1,2], A. P. Arkin[6,7] and **P. D. Adams[6,7]**

**[1]Institute for Systems Biology, Seattle, WA**; [2]Montana State University, Bozeman, MT; [3]University of Washington, Seattle, WA; [4]University of Georgia, Athens, GA; [5]Oak Ridge National Lab, Oak Ridge, TN; [6]Lawrence Berkeley National Lab, Berkeley CA; [7]University of California at Berkeley, CA

*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods. Thus, ENIGMA has been organized into several campaigns involving multiple institutes with varying expertise.** *Here we describe an overarching goal of the Environmental Simulations and Modeling Campaign to simulate, model, predict, and characterize mechanistic underpinnings of $N_2O$ emissions in varying ecological contexts (pH, metal availability, oxygen, etc.) using field isolates assembled into synthetic communities.*

**Abstract: (Please limit to 2 pages.)**

The Field Research Center (FRC) at Oak Ridge, TN has some of the highest subsurface nitrate concentrations [>10g/L] ever recorded. This concerningly large pool of subsurface nitrate, which is a remnant of legacy activities, can end up as the greenhouse gas nitrous oxide ($N_2O$) via incomplete denitrification or as nitrogen gas ($N_2$) when completely denitrified. It is critical to understand the environmental drivers that favor either complete denitrification ($N_2$ emission) or incomplete denitrification (e.g., $N_2O$ emission) in the subsurface so that models can predict the fate of excess nitrate. For instance, at the FRC we have observed that wells with a pH below neutral and high nitrate levels emit large amounts of $N_2O$. In addition, monitoring wells after rainfall events revealed a sudden decline in pH up to 1.5 units over a matter of hours. We therefore hypothesized that if the process of complete denitrification is partitioned among multiple organisms (i.e., incomplete denitrifiers), this coupled reaction is subject to disruption by abiotic factors that may lead to increased $N_2O$ off-gassing. To test this hypothesis, we have established a synthetic community (SynCom) of two field isolates --*Rhodanobacter sp*. R12 and *Acidovorax sp*. 3H11-- which together can perform complete denitrification but cannot independently. Therefore, we have generated a cross-campaign initiative to elucidate different mechanisms of abiotic control including pH shifts, microaerobic environments, and metal availability. Using time course experiments we determined that a shift in pH from neutral pH 7 to pH 6 is enough to decouple the complete denitrification process of the SynCom resulting in

significant increases in $N_2O$ emissions. Transcriptome analysis of the SynCom at differing pH conditions, suggest dynamic changes in community composition and physiological states. Transmission electron microscope images suggest very different morphologies between the two field isolates that may play an important role in carbon, nitrogen, and phosphorus fluxes between the organisms. Current experiments are focused on shifts in pH at differing C/N ratios, oxygen, and metal availability (e.g., Ni) that can shift complete denitrification to incomplete. Insights from these studies can be utilized to define reaction terms for predictive modeling at the field site.

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Diurnal and Seasonal Fluctuations with the Subsurface: A 17-Week Survey of Groundwater and Sediment in 27 Contaminated Wells**

K. F. Walker[1,2]*, E. R. Dixon[1,3], D. C. Joyner[1,3], K. A. Lowe[1], F. L. Poole[4], X. Ge[4], M. P. Thorgersen[4], D. Ning[5], Y. Fan[5], J. P. Michael[5], J. D. Van Nostrand[5], L. M. Lui[6], X. Wu[6], J. L. Goff[4], M. W. W. Adams[4], R. Chakraborty[6], D. A. Elias[1], R. L. Wilpiszeski[1], J. Zhou[5,6], M. W. Fields[7], T. C. Hazen[1,2,3,6], A.P. Arkin[6], and **P. D. Adams[6]**

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]Bredesen Center of Interdisciplinary Research and Education, Knoxville, TN; [3]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN; [4]University of Georgia, Athens, GA; [5]University of Oklahoma, Norman, OK; [6]Lawrence Berkeley National Laboratory, Berkeley, CA; and [7]Montana State University, Bozeman, MO.

*http://enigma.lbl.gov*

Project Goals: **ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) uses a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

Spatiotemporal variability of groundwater within contaminated wells greatly effects the geochemistry and microbial communities present. This survey of 27 wells at the DOE Oak Ridge Reservation (ORR) Y-12 Complex in Oak Ridge, Tennessee, set out to obtain diurnal and seasonal fluctuations within three levels (low, moderate, and high) of nitrate and heavy metal contamination. With this data, we created 3D geochemical models using RockWorks of these areas to study changes within the attached and unattached microbial communities in relation to groundwater geochemistry.

Measurements were gathered in 27 previously established groundwater wells, four days a week, for the span of 17 weeks (70 days total) spanning from July to December 2019 to build both diurnal and seasonal time series of geochemistry and microbial communities. In-field geochemical measurements were obtained using In-Situ Aqua TROLL 600s, including dissolved oxygen (DO), pH, conductivity, oxidation-reduction potential (ORP), and nitrate concentration. Samples were then taken from each well to measure 52 metals, anions, organic acids, and total organic and inorganic carbon/nitrogen. Throughout the study, results show diurnal and seasonal changes in geochemistry with wide variations between each well and levels of contamination. The subsurface geochemistry was also greatly affected by rainfall events, which was evident after two months of regional drought conditions. Additionally, one well in each level of contamination (3 wells total) was selected to complete a microbial analysis by sampling for

groundwater (unattached) and sediment (attached) microbial communities. Groundwater was filtered through 8µm and 0.2µm filters for 16S rRNA and metagenomic analysis (420 filters). In each of the three wells, 18 unamended sediment traps were placed to complete a time series analysis of the attached microbial community's sediment geochemistry.

Results of all analyses are linked to groundwater flow vectors (using the Geotech Colloidal Borescope) on-site weather data (using a HOBO RX300 Weather Station) to produce a correlation analysis between source water and flow paths, groundwater geochemistry, weather events, and levels of contamination on a spatiotemporal scale. With this data, we aim to establish a predictive model between groundwater geochemistry and microbial communities to inform future ENIGMA groundwater and sediment sampling in a planned subsurface observatory.

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

# High-Throughput Genetic Analysis of Rhodanobacter Reveals Genes Important for Metal Tolerance

Li-Wen Wang (Audrey_LiWenWang@lbl.gov)[1]*, Trenton K. Owens[1], Hans K. Carlson[1], Alexey E. Kazakov[1], Morgan N. Price[1], A.M. Deutschbauer[1,2], A.P. Arkin[1,2], and **P.D. Adams[1,2]**

**[1]Lawrence Berkeley National Lab, Berkeley**; [2] University of California at Berkeley

*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

**Abstract**:
*Rhodanobacter* species are highly abundant at the Oak Ridge Reservation site, which is contaminated with heavy metals, high nitrate, and low pH [1]. To uncover the molecular mechanisms underlying their survival in this extreme environment, we constructed a whole-genome barcoded transposon mutant library in *Rhodanobacter* sp. FW104-10B01 (10B01), through which the fitness of nearly all non-essential genes can be evaluated in parallel through competitive growth assays and DNA barcode sequencing. Our early attempts to mutagenize 10B01 were inefficient, which we hypothesized was due to the native restriction modification systems in this strain. To overcome this inefficiency, we used PacBio sequencing of the 10B01 genome to identify methylated motifs and then constructed a custom transposon delivery vector minus these motifs. This new vector was much more efficient (100-fold) in making mutants in 10B01, and we subsequently used this vector to construct a final library of over 460K uniquely barcoded 10B01 mutants. Using this library, we conducted over 100 mutant fitness assays in the presence of elevated metal concentrations, including for many of the major selective stressors at the site, such as U, Mn, Al, Cd, Zn, Co, and Ni [1]. From these data, we identified heavy-metal efflux pumps that were important for fitness in the presence of a number of metals, and in many instances these efflux pumps were specific for only one or a handful of metals. To enable single-gene follow-up investigation, we archived a collection of 192 plates of individual mutants in monoculture (>18K single mutants). By tracking transposon insertions using barcode sequencing, we are able to accurately locate mutants in the archived library and thereby connect specific Rhodanobacter genotypes to phenotypes through analysis of mutants [2, 3, 4]. Overall, our new genetic tools and resources for *Rhodanobacter* will facilitate research on the mechanisms by which these bacteria survive in heavily contaminated environments.

## References


1. Carlson, H. K., Price, M. N., Callaghan, M., Aaring, A., Chakraborty, R., Liu, H., ... & Deutschbauer, A. M. (2019). The selective pressures on the microbial community in a metal-contaminated aquifer. The ISME journal, 13(4), 937-949.
2. Shiver, A. L., Culver, R., Deutschbauer, A. M., & Huang, K. C. (2021). Rapid ordering of barcoded transposon insertion libraries of anaerobic bacteria. Nature Protocols, 16(6), 3049-3071.
3. Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., ... & Deutschbauer, A. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio, 6(3), e00306-15.
4. Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., ... & Deutschbauer, A. M. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. Nature, 557(7706), 503-509.



*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

# A Field-Scale Omics-Enabled Groundwater Reactive Transport Model

Y. Wang[1*] (yajiao.wang-1@ou.edu), D. Ning[1], G. Wang[1], N. Xiao[1], Z. Shi[1], P. Zhang[1], Z. He[1], J.D. Van Nostrand[1], L. Wu[1], T.C. Hazen[2,3], J-Z. Zhou[1], A.P. Arkin[4,5] and **P.D. Adams**[4,5]

[1]**University of Oklahoma, Norman**; [2]University of Tennessee, Knoxville; [3]Oak Ridge National Lab, Oak Ridge; [4]Lawrence Berkeley National Lab, Berkeley; [5]University of California, Berkeley.

*http://enigma.lbl.gov*

**Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.**

Microbial activity substantially impacts the temporal and spatial distribution of contaminants through diverse reactions and transport in groundwater. The reactive transport model (RTM) in groundwater is aimed to quantitatively describe and predict the distribution of chemicals, which integrates various active hydrological, geochemical, and microbial processes. However, current groundwater reactive transport models are limited in the ability to predict microbial activity, especially when the structure and functions of the microbial community are dynamic due to variations of environmental conditions. The increasing availability of the meta-omics data makes it feasible to quantitatively simulate and predict the dynamics of microbial community structure and functions in groundwater. Our idea is to refine the biotic reactions in the reactive transport processes by incorporating the microbial functional group information through omics measurements to explicitly represent microbially-mediated biogeochemical reactions.

The primary flow field and solute transport model were built using the commonly applied MODFLOW and MT3D-USGS models. In addition, we modified MT3D-USGS to incorporate multiple (i.e., 22) microbial functional groups, which mediate the critical metabolic processes in anaerobic fermentation of organic matter, methanogenesis, methane oxidation, oxic/anoxic degradation of organic matter, uranium (VI) reduction, denitrification, nitrification, nitrogen fixation, dissimilatory sulfate reduction and oxidation, dissimilatory iron (III) reduction and oxidation. In addition, the new model development will include the assimilatory processes of methane oxidation, nitrate reduction, sulfate reduction, and iron (III) reduction, microbial mortality, abiotic iron (II) oxidation and iron (III) reduction, immobilization of uranium (IV) and iron (II, III), and sorption of various substances.

The measurements of chemical concentrations and functional gene abundances will be used to constrain and validate the new model. However, since the groundwater reactive transport process

combines various complex processes, it is challenging to estimate the key parameters. Therefore, we developed an parameter optimization procedure for field-scale groundwater RTM based on the Shuffled Complex Evolution (SCE) algorithm. To enhance the computation efficiency, we developed the parallel computation program with the OpenMPI interface on the supercomputer. We considered multiple objectives in the calibration of model parameters. Each objective evaluates the goodness-of-fit of a specific observed variable, e.g., water table, concentrations of ammonium/nitrate/nitrite, sulfate/sulfide, iron (II/III), uranium (VI), gases ($N_2O$, NO, $CH_4$, $CO_2$, $H_2$), microbial biomass, functional gene/enzyme abundances. The overall objective function is the weighted average of multiple objective functions.

This new model was applied to the Emulsified Vegetable Oil (EVO) injection project in 2009 at the ORIFRC site. Here we present the updates of the modeling results on the model development and parameterization. The total area of the modeling zone was around 3700 $m^2$, which contains the main flow path of injected chemicals. The modeling area was discretized into grids with 0.5 m × 0.5 m × 2 m for the finite-difference calculation. At first, we used the MODFLOW and PEST pilot point method to estimate the spatial distribution of flow-related parameters, i.e., hydraulic conductivity and recharge rate, and simulated the transient flow field after EVO injection. The mean residual of the water heads in the optimization was <0.01 m. Then we estimated the porosity and dispersivity with tracer concentration due to its weak reactivity. To build the baseline RTM, we estimated the reactive parameters and simulated the chemical transport process with biomass-based microbial reaction rate. Next, we will integrate the omics data into the reaction rates and use functional gene abundances to constrain the model. By comparing with the baseline RTM, we will evaluate the omics-informed reactive transport model to see if we could improve modeling performance and reduce uncertainty by introducing necessary functional enzymes/genes.

This new model could help better understand how different microbial functional groups contribute to the transformations of the contaminants in groundwater after EVO injection in the complex interactions between microbial, hydrological, and geochemical processes. It can also be extended to quantitatively and systematically predict the fate of contaminants in groundwater under different scenarios (e.g., extreme flood, low temperature, the dosage of different carbon sources).

*This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Title:** Optimal Experimental Design (OED) of Biological Systems

**Authors:** Puhua Niu (niupuhua.123@tamu.edu),[3*] Maria J. Soto[2], Byung-Jun Yoon[1,3], Edward R. Dougherty[3,] **Francis J. Alexander**[1,] Ian Blaby[2], and Xiaoning Qian[1,3],

**Institutions:** [1]Brookhaven National Laboratory, Upton, NY; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Texas A&M University, College Station, TX.

**Project Goals:** The project's overall goal is to develop optimal experimental campaigns to achieve a particular objective, namely metabolite yield alteration. The optimal experiments will be designed by quantifying the cost of uncertainty in the current predictive model—a transcriptional regulatory network (TRN) model that regulates metabolism—and selecting the experiments that are expected to maximally reduce the model uncertainty that affects the attainment of the aforementioned objective. This approach will serve as a proof of principle, demonstrating the significant potential of computationally guided biology in areas directly relevant to BER's missions.

**Abstract Text:** There has been extensive research on *in silico* modeling and prediction of genome-scale metabolic behavior, mostly focusing on mutant strain design with metabolic reaction network modeling [1]. To further model transcriptional regulations, these metabolic network models are also integrated with genetic regulatory relationships involving transcription factors (TFs) that may regulate metabolic reactions. Transcription regulation is often integrated via "transcriptional regulatory constraints" with various heuristics for flux-balance analysis (FBA) of metabolic networks. However, many of these computational tools were often only validated for selected model organisms with curated data and network models, despite access to high-throughput technologies allowing genome-scale engineering.

A major achievement in our project during the past year is the development of TRIMER (Transcription Regulation Integrated with MEtabolic Regulation), a genome-scale computational pipeline for integrative modeling and analysis of TF-regulated metabolism. Specifically, a Bayesian network (BN) is employed in TRIMER, instead of local TF-gene conditional probabilities or transcriptional regulatory constraints, thereby aiming at effectively capturing the global transcriptional regulatory relationships that may affect metabolism. Through this BN, the influence of transcription regulation (and its changes) on metabolic behavior under different conditions can be predicted more accurately via more flexible conditional probability inference, by linking transcription regulation to metabolism based on prior knowledge on TF-gene-reaction interactions. Details of our TRIMER package and comprehensive results can be found in [2, 3].

To experimentally validate the predictions generated by TRIMER, we assayed 163 TF deletants vs. wildtype for biomass and metabolite levels, which served to confirm the performance of the tool.

While *E. coli* is a well-understood model organism, with extensive prior knowledge on both gene regulation and metabolic reaction pathways as well as significant amount of accumulated expression data, one critical challenge to overcome when applying TRIMER to less-studied organisms—especially, in the context of optimal experimental design (OED)—is to understand the model uncertainty or sensitivity due to incomplete knowledge and/or noisy data. Building upon TRIMER, we further investigated how model uncertainty in the TRN may affect the metabolic flux prediction and the TF KO experiment design. With a quantified uncertainty, we aim to further optimize the outcome for OED under uncertainty in an efficient manner (i.e., fewer experiments and less guesswork), whereby optimization is achieved by optimally (most favorably) improving the model or the microbial system represented by the model relevant to the objective (i.e., maximizing the metabolite yield). We here present the recent research progress on model uncertainty analyses as well as the planned research directions based on our analyses.

**References/Publications**


[1] Amit Varma and Bernhard Ø Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use.Bio/technology, 12(10):994–998,1994.

[2] Puhua Niu, Maria J. Soto, Byung-Jun Yoon, Edward R. Dougherty, Francis J. Alexander, Ian Blaby, Xiaoning Qian. TRIMER: Transcription Regulation Integrated with MEtabolic Regulation, *iScience*, 24(11):103218, 2021.

[3] Puhua Niu, Maria J. Soto, Byung-Jun Yoon, Edward R. Dougherty, Francis J. Alexander, Ian Blaby, Xiaoning Qian. Predicting condition-dependent metabolite yield with TRIMER, *STAR Protocols*, under review (minor revision).



**Funding Statement:** The materials presented in this paper are based upon the work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under contract number DE-SC0012704

**Title:** Genome-wide mapping of cis-regulatory elements and regulation of nitrate assimilation in *Phaeodactylum tricornutum*

**Authors:** Sarah R. Smith,[1,2] Hong Zheng,[2] Pratap Venepally,[2] Leo Baumgart,[3] Ronan O'Malley,[3] and **Andrew E. Allen[2,4]*** (aallen@ucsd.edu)

**Institutions:** [1]Moss Landing Marine Laboratories, San José State University, Moss Landing, CA; [2]J. Craig Venter Institute, La Jolla, CA; [3]Joint Genome Institute, [4]Scripps Institution of Oceanography, UC San Diego, La Jolla, CA; and [4]University of State, City

**Website URL:**

**Project Goals:** Short statement of goals. (Limit to 1,000 characters)

**Abstract Text:** Despite the importance of diatoms in the marine environment, and their relevance for biofuels development, little is known about how diatoms sense and respond to shifts in environmental conditions. Transcription factors (TFs) regulate gene expression programs by binding DNA and promoting (activating) or blocking (repressing) recruitment of RNA polymerase to initiate transcription, and undoubtedly have a role in the response of diatoms to their environment. We used DNA affinity purification sequencing (DAP-Seq), a high-throughput *in vitro* method, to characterize transcription factor binding sites genome-wide. Using this method, we have successfully mapped the transcription factor binding sites for 58 TFs from the *P. tricornutum* genome. Several TF classes failed *in vitro* to generate significant signals, which could be due to incorrect gene models, the lack of true DNA binding domains, or insufficient maturation (lack of heteromers or posttranslational modifications). The most successful classes of TFs in the DAP-seq pipeline were the bZIPs, heat shock factors (HSFs), Myb and MybSHAQKY TFs. Our results corroborate findings from previous investigators for the few transcription factors that have been functional characterized, and greatly expand the catalog of these key molecular components of signal transduction cascades in diatoms. Genome-wide, several bZIPs (Aureochrome1a, bZIP10, bZIP11, bZIP13, bZIP15) were associated with patterns of gene expression driven by diel signatures, pointing to a role for these TFs in regulating the shift between illuminated and dark cell physiology. Fewer TFs were associated with overall macronutrient status. Careful dissection of the architecture of transcription factor binding sites (TFBS) in the promoters of a tightly coordinated set of nitrate assimilation genes that are highly sensitive to nitrate availability (highly nitrate sensitive, or HNS regulon) revealed the existence of TFBS hotspots within different promoters. Analysis of these hotspots suggests that HNS genes are likely regulated by a combination of activation and repression from various TFs, including light and cAMP-sensitive bZIPs, low N induced HSFs, and a homolog of a fungal regulator of the nitrate regulon. This mode of regulation shares little similarity with what is known from cyanobacteria, plants, other algae (including *Chlamydomonas*), and filamentous fungi, and gives insight into the functional significance of the HSFs in diatoms, which have undergone massive evolutionary radiation relative to HSFs in other eukaryotes. Precise

knowledge of mechanisms regulation transcription from the HNS promoters will further our understanding of how diatoms integrate the assimilation of nitrate with other cellular energetic demands during growth. This detailed knowledge is also essential in synthetic biology to most effectively design transgenic lines that use the nitrate reductase (NR) promoter as an inducible promoter.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC00006719 and DE-SC0008593.

**Unravelling xanthophyll pigment biosynthesis in algae**

Yu Bai,[1*] (yu.bai@colostate.edu), Maxwell A. Ware,[1] Martin Lohr,[2] Graham Peers[1] and **Andrew Allen[3]**

[1]Colorado State University, Fort Collins; [2]Johannes Gutenberg-Universität, Mainz; [3]J. Craig Venter Institute, La Jolla, CA

**Project Goals**: *Reprogram metabolic networks using in vivo synthetic modules to increase the flux of energy and carbon into biofuel precursors. Goal 1) Profiling the transcriptome, proteome and metabolome to investigate cell responses to physiologically relevant conditions. Goal 2) Identify and manipulate key factors involved in the control of inorganic C assimilation, photosynthetic efficiency and regulation of lipid accumulation. Goal 3) Development of Phaeodactylum genome reconstruction/modeling and our development of novel synthetic genomic tools to achieve our overall goal of increasing productivity.*

Diatoms are the dominant eukaryotic primary producers of the world's oceans. Their versatile peripheral photosynthetic light harvesting antenna, Fucoxanthin Chlorophyll a/c binding Proteins (FCPs), enable diatoms to survive under extreme light conditions through enhanced light absorption and efficient energy dissipation compared to plants. The main accessory xanthophyll pigment, fucoxanthin, gives cells their distinct brown color and an extended spectral range for photosynthesis. Meanwhile, the novel diadinoxanthin (Dd)-diatoxanthin (Dt) cycle xanthophylls contribute to the FCP's switch between a light harvesting and energy dissipation state. Notably, the biosynthetic pathway of these xanthophylls was previously unknown. By applying reverse genetic techniques, we isolated three green *Phaeodactylum tricornutum* mutants, *zep1*, *vdl2* and *crtiso5* which cannot produce fucoxanthin and exhibit a reduced functional absorption cross-section of PSII. A fucoxanthin biosynthesis pathway is proposed based on the structures of the carotenoids accumulated in the mutants. Another mutant, *zep3,* is unable to convert Dt into Dd after the shift from high light to low light, resulting in irreversible non-photochemical quenching (NPQ). The constitutively accumulated diatoxanthin in *zep3* lines leads to a lower photochemical efficiency of photosystem II (≈60% of WT). Phylogenies of the VDL, ZEP and CRTISO families from plants and algae, suggest that *zep1, zep3, vdl2 and crtiso5* evolved via multiple duplications and neofunctionalization of their ancestor genes during the process of secondary endosymbiosis that led to Stramenopile algae. These discoveries open the potential for the manufacture of important nutraceutical pigments and can be applied to engineering the efficiency of photosynthesis.

***Funding:*** *This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science – grant # DE-SC0008595*

**CRISPR interference (CRISPRi) for Dynamic Regulation of Gene Expression in Diatoms**

Tessema Kassaw[1*](tessema.kassaw@colostate.edu), Graham Peers[1], **Andrew Allen**[2]

[1]Colorado State University, Fort Collins, CO; [2] J. Craig Venter Institute, San Diego, CA.

***Project Goals****: Reprogram metabolic networks using in vivo synthetic modules to increase the flux of energy and carbon into biofuel precursors. Goal 1) Profiling the transcriptome, proteome and metabolome to investigate cell responses to physiologically relevant conditions. Goal 2) Identify and manipulate key factors involved in the control of inorganic C assimilation, photosynthetic efficiency and regulation of lipid accumulation. Goal 3) Development of Phaeodactylum genome reconstruction /modeling and our development of novel synthetic genomic tools to achieve our overall goal of increasing productivity.*

CRISPRi-based gene inactivation is an attractive tool to probe genome-wide gene functions and to reprogram metabolic networks due to its high target specificity, reversibility, and multiplexing capacity. However, the current state of the art for gene expression knockdown in most algae is RNAi. While RNAi can occasionally be successful in producing strains with lowered cellular enzyme activities, it is less efficient, and the underlined mechanism is still unknown in diatoms. We recently developed two episome-based small molecule responsive transcriptional control systems in the model diatom alga, *Phaeodactylum tricornutum*. These systems are activated by the addition of exogenous β-estradiol and digoxin. We demonstrated that our transcriptional control systems are tunable and reversible maintaining a dynamic range of up to ~180-fold. We employed our inducible system to regulate the expression of *P. tricornutum* codon optimized dCas9 to establish a CRISPRi system for at will control of target protein production from an episome. We first assessed the feasibility of our CRISPRi approach using eYFP reporter system. We found that up to 80% repression can be achieved by targeting the reporter protein close to its transcription start site using quantitative fluorescence assay. We then selected and designed sgRNA for multiple endogenous loci (nitrate reductase, urease) to demonstrate the β-estradiol inducible CRISPRi application. Based on our preliminary result, we observed ~20% and ~70% growth reduction by targeting nitrate reductase and urease genes respectively. To our knowledge this is the first report on exploiting inducible CRISPRi system for targeted gene inactivation in any algal system. The procedure described here will expand the synthetic biology toolkits in diatoms and can be adapted for high-throughput interrogation of genome-wide gene functions and redirection of metabolic fluxes towards target chemical production.

***Reference:***

*Kassaw, T.K.; Paton, A.J.; Peers G. Episome-Based Gene Expression Modulation Platform in the Model Diatom Phaeodactylum tricornutum. ACS Synthetic Biology, 2022. DOI 10.1021/acssynbio.1c00367.*

***Funding:*** *This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science – grant # DE-SC0008595*

**Modeling carbon metabolism of the diatom *Phaeodactylum tricornutum* during nitrogen starvation and during high light and low light conditions**

Amy Zheng[1*](amy.o.zheng@vanderbilt.edu) Bo Wang[1*] (bo.wang.2@vanderbilt.edu), Miles Crockett[2], Michael T. Guarnieri[3], James K. McCarthy[4], Graham Peers[5], Jamey D. Young[1,2], **Andrew Allen[4,6]**

[1] Department of Chemical and Biomolecular Engineering, [2] Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN. [3] National Renewable Energy Laboratory, Golden, CO. [4] Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA. [5] Department of Biology, Colorado State University, Fort Collins, CO. [4] Microbial and Environmental Genomics Group, J. Craig Venter Institute. La Jolla, CA. [6] Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA.

https://www.jcvi.org/diatom-systems-biology

**The diatom *Phaeodactylum tricornutum* (Pt), a model photosynthetic eukaryotic microbe, has the ability to accumulate up to 45% of dry cell weight as triacylglycerol (TAG), a neutral lipid and precursor to biodiesel[1]. To take advantage of this innate ability, we need to understand how metabolic pathways adjust to changing environmental conditions. The long-term goal of this project is to promote efficient production of high-value and fuel-related compounds through optimization of metabolic fluxes in Pt. Building upon our expertise in $^{13}$C metabolic flux analysis (MFA),[2] our current goal is to develop novel experimental protocols and data analysis workflows to enable $^{13}$C flux analysis of Pt. We are currently investigating the metabolic adjustments of Pt to three variables, *i.e.*, light, nitrogen availability, and genetic knockout of TAG degradation enzymes, which strongly impact cell growth and lipid accumulation.**

In our first study, we varied the intensity of light supplied to the Pt culture. We compared metabolic fluxes inside wild-type (WT) Pt cells grown under low-light (60 μE $m^2$ $s^{-1}$) or high-light (250 μE $m^2$ $s^{-1}$) conditions. We compared the metabolic measurements and net fluxes in these two conditions using $^{13}$C metabolic flux analysis. We demonstrated that carbon was fixed at a faster rate under the high light conditions compared to low light conditions, and the cell mass composition and TAG profiles were different between the two conditions. We aim to compare the flux maps to understand how photosynthesis activity affects central carbon metabolism and TAG accumulation in Pt.

In a second study, we investigated metabolic fluxes inside wild-type Pt and a nitrate reductase (NR) knock-out strain in response to changing nitrogen availability in the culture medium. We studied three cultures, *i.e.*, Pt-WT with nitrate (WT/N+), Pt-WT without nitrate (WT/N-), and Pt-NR with nitrate (NR/N+). We found WT/N− accumulated much higher levels of TAG and total lipids than that of WT/N+, which is consistent with previous studies. Meanwhile, the chlorophyll, protein content and free amino acid levels inside WT/N− cells dropped substantially relative to that in WT/N+. In contrast, the carbohydrate content, urea and metabolites in the TCA cycle of WT/N− cells increased dramatically compared to that in WT/N+. Our results are consistent with previous findings that genes associated with urea cycle are upregulated while

expression of urea-degrading urease is downregulated in WT Pt cells under nitrogen starvation conditions. Interestingly, although NR/N+ showed a biomass composition similar to that of WT/N−, its carbohydrate content was about 50% higher than that of WT/N−. $^{13}$C-labeling and targeted metabolomics revealed that NR/N+ cultures maintained smaller metabolite pool sizes in the TCA cycle and nitrogen assimilation pathways but exhibited higher labeling rates compared to WT/N−. Our $^{13}$C-MFA results have revealed remarkable differences in the metabolic fluxes between WT/N+, WT/N- and NR/N+.

In a third study, we aimed to characterize metabolic changes in an acyl-CoA dehydrogenase knockout (ACAD-KO) Pt strain. When Pt cultures are switched from nitrogen-depleted to nitrogen-replete media, WT cells rapidly degrade the accumulated TAG while ACAD-KO cells retain their TAG stores. Comparing the ACAD-KO strain to WT after nitrogen repletion, we observed increases in TCA cycle labeling in the ACAD mutant. We hypothesize that the TCA cycle in the WT strain is being fed by the breakdown of the TAG, resulting in lower labeling.

Our findings based on $^{13}$C MFA will help us to understand how Pt metabolism adapts to various environmental conditions and genetic modifications, which will guide strain engineering efforts to maximize TAG biosynthesis in Pt. We will conduct future MFA studies of Malic Enzyme overexpression will further our understanding of TAG accumulation and the facilitative role of cross compartment redox shuttling through modification of a central metabolic enzyme, in diatoms.[4]

(Supported by grant *DE-SC0018344: Design, Synthesis, and Validation: Genome Scale Optimization of Energy Flux through Compartmentalized Metabolic Networks in a Model Photosynthetic Eukaryotic Microbe* from the Department of Energy.)

1. Hu, Q. et al. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant Journal* **54**, 621-639, doi:10.1111/j.1365-313X.2008.03492.x (2008).
2. Jazmin, L.J. et al. Isotopically nonstationary $^{13}$C flux analysis of cyanobacterial isoburyaldehyde production. *Metabolic Engineering* **42**, 9-18, https://doi.org/10.1016/j.ymben.2017.05.001 (2017)
3. Levitan, O. et al. Remodeling of intermediate metabolism in the diatom Phaeodactylum tricornutum under nitrogen stress. *PNAS* **112**, 412-417, https://doi.org/10.1073/pnas.1419818112 (2015)

**Title: Bacterial Community Response to Environmental Change Varies With Depth in the Surface Soil.**

**Authors:** Kristin M. Barbour[1]* (kbarbou1@uci.edu), Claudia Weihe[1], Steven D. Allison[1,2], Eoin L. Brodie[3], Michael L. Goulden[1], Adam C. Martiny[1,2], Kathleen K. Treseder[1], and **Jennifer B. H. Martiny[1]**

**Institutions:** [1]Department of Ecology and Evolutionary Biology, University of California, Irvine, California; [2]Department of Earth System Science, University of California, Irvine, California; [3]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, California

**Project Goals:** We sought to assess whether bacterial response to above-ground environmental changes is uniform throughout three layers of the surface soil: the leaf litter layer, 0-2 cm of bulk soil, and 0-10 cm of bulk soil. We aimed to characterize how bacterial community composition responded to drought, wildfire, and temporal variation using the ongoing Loma Ridge Global Change Experiment. We did this in two adjacent ecosystems, a grassland and coastal sage scrubland, to further investigate how ecosystem type may influence bacterial response to environmental change throughout the surface soil.

**Abstract Text:** Bacterial communities in the organic leaf litter layer and bulk (mineral) soil are sensitive to environmental change. However, despite close interactions between these communities, the leaf litter layer has historically been studied in isolation from the bulk soil. Whether bacterial response to environmental changes is uniform throughout the surface soil remains unclear. Here, we simultaneously characterized how bacterial community composition in three surface soil layers (the leaf litter layer, 0-2 cm of bulk soil, and 0-10 cm of bulk soil) responded to a wildfire burning through a decade-long drought simulation in two adjacent ecosystems, a grassland and coastal sage scrubland. We found that bacterial communities in all three surface soil layers were distinct in composition and responded to drought, ecosystem type, and temporal variation. Moreover, the impact of these environmental changes on bacterial community composition decreased with depth in the surface soil. For instance, bacterial response to drought simulation was three-fold higher in the leaf litter layer than in the top 10 cm of bulk soil, with drought treatment explaining 5% and less than 2% of the compositional variation respectively (PERMANOVA: $P < 0.01$ in all cases). Wildfire altered bacterial composition in the leaf litter layer (PERMANOVA: $P < 0.05$) while there was no significant change in composition with 0-10 cm of the bulk soil (PERMANOVA: $P > 0.05$). Further, previous exposure to drought did not influence bacterial response to the wildfire. These data suggest that considering soil depth when assessing the impact of environmental conditions on the surface soil microbiome may improve predictions about the degree to which microbial communities, and therefore C flux, will respond to future environmental change.

**Funding Statement:** This research was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research grants DE-SC0016410 and DE-SC0020382 and the Ridge to Reef NSF Research Traineeship, award DGE-1735040.

**Title:** Litter Microbial Trait-Based Strategies in Response to Drought

**Authors:** Brianna K. Finley[1]* (bkfinley@uci.edu),[1] Bahareh Sorouri,[1] Kathleen Treseder,[1] Jennifer Martiny,[1] Adam Martiny,[1] Alejandra Rodriguez Verdugo,[1] Michael L. Goulden, [2] Shi Wang,[2,3] Eoin Brodie,[1] and **Steven Allison**

**Institutions:** [1] University of California, Irvine, CA; [2] University of California, Berkeley, CA; [3]Ecology Department, Climate and Ecosystem Sciences, Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA

**Project Goals:**

The overarching goal of this project was to better understand how microbial trait-based responses under environmental change – such as drought - influence biogeochemical cycling. Specifically, we combine different genomic methods with field-based manipulations of microbiomes to understand microbial functional responses to drought over two years. This project consists of multiple activities. First, we conducted field-based litter microbiome experiments and laboratory assays of whole microbial communities to measure microbiome community shifts and dispersal contributions to drought responses. Secondly, we will assess microbial demographic changes as well as trait tradeoffs of functional genes under different life history strategies (high growth yield, resource acquisition, and stress tolerance) via metagenomic shotgun sequencing. Lastly, we use knockout approaches to validate the genetic basis of drought tolerance phenotypes observed in our experiments.

**Abstract Text:**

Microbial community shifts under environmental change are shown to influence carbon and nutrient cycling. However, the extent to which evolutionary adaptation, demographic changes, and dispersal affect microbial trait-based responses to global change such as drought and subsequent consequences for biogeochemical cycling are poorly understood. Here, we implemented a two-year field plant litter decomposition experiment within the Loma Ridge Global Change Experiment (LRGCE) to assess effects of drought on microbiome functioning with and without the effects of dispersal. The LRGCE has undergone drought manipulations (roughly 50% rainfall reduction) since 2007 on two adjacent vegetation communities: grass and coastal sage shrub. To remove legacy effects of drought on vegetation and microbiomes, we sterilized leaf litter from both vegetation types and inoculated each litter with microbiomes from unsterilized litter from the ambient precipitation plots. We used 0.2 μm pore nylon mesh "microbial cages" to limit dispersal as well as 2 mm pore microbial cages to allow for dispersal within the ambient and drought plots. After two destructive sampling time points 5 and 11 months after microbial cage deployment, we did not observe a difference in litter mass loss between drought and ambient conditions but found about 25% greater litter mass loss from the coastal sage shrub litter compared to the grass across treatments, as well as an interactive effect between vegetation type and dispersal. From the microbiomes isolated from the field experiment,

we will conduct shotgun sequencing to assess changes in microbiome demographics under drought responses with and without the effect of dispersal. Additionally, we will identify and quantify key functional genes associated with different microbial life history strategies under the YAS framework (high growth yield, resource acquisition, and stress tolerance). This will be conducted both at the metagenomic scale of entire microbiomes, as well as quantify different growth rates of genomes using quantitative stable isotope probing within metagenomes. In order to explore the genomic basis for microbial fitness traits, we are using random-barcoded transposon mutagenesis and sequencing (RB-TnSeq). We have developed pools of tagged transposon mutants in a highly-active litter-degrading bacterium (Erwinia LR017) originating from our field evolution experiment. Assays were designed to explore the costs of gene disruptions to the fitness of this bacterium in media containing substrates relevant to litter decomposition, including under osmotic stress representative of drought conditions. This information is then used to provide direct genotype-to-phenotype relationships for improved model prediction of how traits and trade-offs interact to control microbial litter decomposition.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-FOA-0002059.

# Identification of Pennycress (*Thlaspi arvense* L.) Proteins that Influence Lipid Droplet Formation and Modulate Neutral Lipid Accumulation

Julius Ver Sagun[1*] (juliusver.sagun@unt.edu), Athanas Guzha[1], Mariel Cruz Gómez[2], Allison M. Barbaglia[2], Erich Grotewold[2], Kent D. Chapman[1], and **Ana Paula Alonso**[1]

[1]BioDiscovery Institute, Department of Biological Sciences, University of North Texas, Denton, TX USA; [2]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI USA

**Project Goals: To investigate the function of pennycress candidate genes involved in lipid storage and stability using transient expression in *Nicotiana benthamiana*, and to generate stable transgenic pennycress lines for improved seed oil content and composition.**

The finite nature of crude oil-derived fuels coupled with their adverse effects on the environment requires the search for alternative, renewable sources of energy that are more environmentally friendly. Pennycress (*Thlaspi arvense* L.) has been identified as a promising alternative crop for aviation fuel production. It is an annual winter *Brassicaceae* growing in most temperate regions including North America, and produces seeds with high oil content (26-39%). The average yield of pennycress seeds is 1,500 kg $ha^{-1}$, corresponding to 600–1200 L $ha^{-1}$ of oil, which is higher than that of common oil crops such as soybean and camelina. While pennycress benefits from the fully sequenced genome and research tools of the closely related model plant *Arabidopsis thaliana*, there are still significant challenges associated with establishing gene function that would make pennycress much more valuable as a dedicated bioenergy oilseed crop. Transcriptional analysis of 22 pennycress accessions resulted in the identification of 35 potential gene candidates whose expression levels were correlated with seed oil yield (DE-SC0019233). After transient over-expression of the coding sequences of these 35 pennycress genes, 15 enhanced the number of cytoplasmic lipid droplets, increased the total neutral lipid content, and altered the fatty acid composition in the neutral lipid fraction of *Nicotiana benthamiana* leaves. Confocal microscopy also showed that protein products of two of these ORFs - a lipid transfer protein (LTP6) and a lipid droplet associated protein (LDAP3) - mainly localized to lipid droplets. These two ORFs were cloned into expression constructs behind the *CaMV 35S* constitutive promoter or the pNAP seed specific promoter and transformed into pennycress plants through the floral-dip method. Future work is aimed at characterizing seeds of these transgenic lines to examine the impact of LTP6 and LDAP3 overexpression on seed lipid content and composition. These results may lead to new targets to effectively improve oil content and composition in pennycress seeds, and may ultimately contribute to the production of sustainable aviation fuel.

*This research was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0020325*

# Using $^{13}$C-labeling to Unravel the Temporal and Spatial Production of Seed Oil in Developing Embryos of Pennycress, a Promising Source for Sustainable Aviation Fuel

Umesh Prasad Yadav[1]* (umeshprasad.yadav@unt.edu), Trevor B. Romsdahl[1,2], Kent D Chapman[1], **Ana Paula Alonso[1]**

[1]Biodiscovery Institute, University of North Texas, Denton, TX; [2]University of Texas Medical Branch, Galveston, TX

**Project Goal: Assessing temporal and spatial regulation of storage lipid accumulation in developing Pennycress seeds using $^{13}$C-labeling and mass spectrometry imaging.**

The US military and commercial aviation industry consume nearly 20 billion gallons of jet fuel per year. Development of sustainable alternatives to fossil-derived jet fuels is critical due to unpredictable prices, finite fossil fuel sources, and concerns over environmental impact. To mitigate the impact of the growing demand for jet fuel, the US aviation industry is targeting an increase in the production of sustainable aviation fuel by 3 billion gallons per year by 2030. In the last few years, pennycress (*Thlaspi arvense* L.) emerged as a promising oilseed crop, especially suited for aviation fuel production, due to its oil content and fatty acid composition. In addition to its excellent biofuel potential, pennycress requires low agricultural inputs and can serve as a cover crop when grown in a summer/winter rotation cycle with other conventional commodity crops, such as corn and soybean. Improvements to the understanding of regulatory factors that limit oil yield in pennycress seeds will be instrumental to advance the goal of developing bio-based aviation fuels for the future. Culture conditions were optimized to provide $^{13}$C-labeled glucose to pennycress siliques to follow the incorporation of labeled carbon in intracellular metabolites and analyze lipid synthesis and storage in pennycress seeds. This approach allows tracing metabolism from carbohydrate sources to storage oils *in vivo* conditions. For this purpose, 100% [U-$^{13}$C]-glucose was supplied for 120 h to siliques 16 days after pollination (DAP) to track the incorporation of $^{13}$C-labeled acetyl fragments into *de novo* synthesized fatty acids in seed plastids, and in elongated fatty acids in the cytoplasm. The percentage of labeling in plastidic and cytosolic acetyl fragments was found to be significantly lower in the axis of the embryos in comparison to the cotyledons. Moreover, mass spectrometry imaging (MSI) was performed in $^{13}$C-labeled pennycress embryos to analyze the isotopologues of phosphatidylcholine (PC), an important membrane lipid and intermediate in storage oil biosynthesis. MSI displayed a greater $^{13}$C-labeling of PC molecular species in the cotyledons than the embryonic axis[1]. Monitoring the temporal incorporation of $^{13}$C-acetyl units into fatty acids of developing embryos showed that the $^{13}$C-labeling in acetyl fragments slowly increased to 12%, whereas $^{13}$C incorporation in sugars, amino acids, and organic acids occurred more quickly and reached a plateau. These results provide insights on the temporal and spatial production of oil in pennycress seeds, and will guide metabolic engineering efforts to produce higher oil yield for use as aviation fuel.

Reference:

1. Romsdahl, T.B.; Kambhampati, S.; Koley, S.; Yadav, U.P.; Alonso, A.P.; Allen, D.K.; Chapman, K.D. Analyzing Mass Spectrometry Imaging Data of 13C-Labeled Phospholipids in Camelina sativa and Thlaspi arvense (Pennycress) Embryos. Metabolites 2021, 11, 148.

*This research was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through the Genomic Science Program grant no. DE-SC0020325.*

**Title:** Application of a Novel $^{2}H$ Isotope Tracer Approach to Characterizing Metabolic Thermodynamics in *C. thermocellum*

**Authors:** Melanie M Callaghan[1,2]* (callaghan2@wisc.edu), Tyler B Jacobson[1,2], Bishal D Sharma[2,3], David M Stevenson[1,2], Daniel G Olson[2,3], Lee R lynd[2,3], **Daniel Amador-Noguez[1,2]**.

**Institutions:** [1]University of Wisconsin-Madison, Madison, WI; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN; [3]Dartmouth College, Hanover, NH.

**Project Goals:** Thermodynamic analysis can help us understand how energy is transferred and transformed within metabolic networks and has emerged as a powerful tool for pathway design and metabolic engineering. This project integrates thermodynamic analysis with advanced mass spectrometry, computational modeling, and metabolic engineering to develop an approach for *in vivo* determination of Gibbs free energies ($\Delta G$) in metabolic networks. This project also investigates how the thermodynamics of biosynthetic pathways in microbial biofuel producers change dynamically as substrates are depleted or products accumulate.

**Abstract:** Consolidated bioprocessing (CBP) relies upon efficient cellulolytic microbial activity to break down plant cellulose, making hard-to-access substrates into metabolizable starting material. *Clostridium thermocellum* is an excellent candidate for CBP, as it is an anaerobic bacterium capable of efficient cellulose catabolism. Previous work has begun to characterize the metabolic networks of *C. thermocellum* that facilitate the conversion of cellulose feedstocks into $C_2$ and $C_4$ alcohol products. These studies found that the thermodynamics of central metabolism in *C. thermocellum* are vastly different than those of other model fermenters, such as anaerobically-grown *Escherichia coli* and *Thermoanaerobacterium saccharolyticum*, having a lesser overall drop in Gibbs free energy and thus more limited thermodynamic driving force [1,2]. The thermodynamics of *C. thermocellum* metabolism is likely a key limitation to achieving high product titers, and consistent with this hypothesis we observe non-optimal ethanol titers during *C. thermocellum* fermentations with high substrate loading.

Established methods for identifying thermodynamic bottlenecks, which are helpful for identifying engineering targets for increasing product yields, have relied upon feeding microbial cultures isotopically labeled ($^{2}H$, $^{13}C$) substrates. Unfortunately, this option is not viable for high concentrations of cellulose substrates. Here, we present an alternative approach utilizing deuterated water ($^{2}H_2O$, or "heavy water") as the source of isotope label during microbial fermentations. We have successfully used HPLC-MS to measure the incorporation of $^{2}H$ from heavy water into central metabolism, including intermediates of glycolysis, the TCA cycle, and the pentose phosphate pathway, as well as amino acids. Labeling patterns are consistent with known metabolic signatures in this organism, indicating that our $^{2}H_2O$ isotope tracer approach is a cost-effective method that is not limited by labeled substrate availability. Application of this technique to high substrate fermentation conditions, as well as various *C. thermocellum* mutants, will provide key information about metabolic flux during these fermentations. These data will

contribute to our holistic understanding of feedstock-to-bioalcohol metabolism, allowing us to identify barriers and optimize pathway engineering.

**References/Publications**


1. Jacobson TB, Korosh TK, Stevenson DM, Foster C, Maranas C, Olson DG, Lynd LR, Amador-Noguez D. 2020. *In vivo* thermodynamic analysis of glycolysis in *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum* using $^{13}C$ and $^{2}H$ tracers. mSystems 5:e00736-19. doi: 10.1128/mSystems.00736-19
2. Cui J, Stevenson D, Korosh T, Amador-Noguez D, Olson DG, Lynd LR (2020). Developing a cell-free extract reaction (CFER) system in *Clostridium thermocellum* to identify metabolic limitations to ethanol production. Front. Energy Res. 8:72. doi: 10.3389/fenrg.2020.00072 .


**Funding Statement:**


This material is based upon work supported by the Department of Energy (DOE) Early Career Research Program under award no. DE-SC0018998 and the Center for Bioenergy Innovation (CBI) under award no. DE-AC05-00OR22725, subcontract no. 4000158665.

**Title:** Optimizing Biological Nitrogen Fixation on Sorghum Using Microbial Communities

**Authors:** April MacIntyre,[1,2] Claire Palmer[1*] (claire.palmer@wisc.edu), Quinlan Kiefer,[1] Jaron Thompson,[1] Vânia C. S. Pankievicz,[1,3] **Sushmita Roy,[1] Wilfred Vermerris,[4] Ophelia Venturelli,[1]** and **Jean-Michel Ané[1]**

**Institutions:** [1]University of Wisconsin-Madison, Madison, WI, USA; [2]Now at Valent BioSciences Inc., Libertyville, IL, USA; [3]Now at GoGenetic, Curitiba, PR, Brazil; and [4]University of Florida, Gainesville, FL, USA

**Website URL:** https://sonar.bact.wisc.edu/

**Project Goals:** We aim to promote biological nitrogen fixation on sweet sorghum, a target crop for biofuel production, using a systems biology-based approach involving both the plant and associated microbes. On the microbial side, we plan to do this by 1) *isolating and characterizing microbes associated with sorghum aerial roots*, 2) *forming representative communities from these isolates and investigating community dynamics and functions by building computational models to predict their behaviors*, and 3) *testing the efficacy of these synthetic communities to promote biological nitrogen fixation, and, by extension, sorghum sustainability.*

Cereal biofuel crops like corn and sorghum cannot form classical symbiotic relationships with nitrogen-fixing bacteria like legumes and rely on synthetic fertilizer and less efficient associative nitrogen fixation[1]. Improving biological nitrogen fixation (BNF) on cereals to improve crop production sustainability and reduce environmental degradation is, therefore, a primary agronomical goal. Some indigenous corn landraces from Central America can obtain 29%-82% of their nitrogen from the air through BNF[2]. Diazotrophic microbes live in the low oxygen, sugar-rich root mucilage of these landraces. Several poorly characterized sorghum accessions also display this trait. Breeding this trait into biofuel-suitable sorghum lines and characterizing and optimizing the BNF potential through microbes are target areas for research and development.

To improve BNF on the microbial side, we have isolated a collection of ~90 individual strains from aerial root mucilage. Preliminary mucilage microbiome data confirmed we covered a wide diversity of strains in the mucilage. We have initiated a much larger comparative meta-amplicon study between field-derived corn and sorghum mucilage 16S, ITS, and *nifH* sequences to elucidate how location, species, and genotype influence mucilage microbial communities.

All environmental isolates have been screened for nitrogen fixation and plant growth-promoting phenotypes, such as auxin production, siderophore production, phosphate solubilization, and ethylene degradation. While <10% of isolates have functional nitrogen-fixing capabilities, many of the isolates display one or more plant-growth-promoting traits.

Carbohydrate analysis has confirmed the monosaccharide composition of the complex polysaccharide making up the mucilage[3]. We currently screen the mucilage for other nitrogen and carbon sources using untargeted metabolomics and biochemical assays. We have found that

ammonium and nitrate are both present in 0-1 mM and 0-3 mM concentrations in sorghum mucilage, respectively, reflecting possible BNF activity. Sorghum mucilage supports nitrogen fixation, as confirmed through acetylene reduction assays.

In parallel to the investigation of environmental isolates, we have begun work with a simplified seven-member community previously identified on corn roots and several additional diazotrophic strains of interest[4]. We will use this synthetic community to investigate the molecular mechanisms and metabolic interactions influencing community-level nitrogen fixation. To do this, we will use high-throughput, combinatorial, *in vitro* community assembly experiments in artificial mucilage media designed from our chemical analyses of the mucilage. We will use our comparative meta-amplicon study's *in vitro* and field data to build computational models at different resolutions to elucidate inter-species interactions promoting community growth, survival, and BNF. These experiments will further guide the selection of defined communities that we can study in the context of sorghum accessions that support the best diazotrophic communities. With these defined communities, we will confirm survival, nitrogen fixation, and promotion of sorghum nutrition, growth, and health in the field.

**References**


1. Pankievicz, V. C. S., Irving, T. B., Maia, L. G. S. & Ané, J.-M. Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops. *BMC Biol.* **17**, 99 (2019).
2. Van Deynze, A. *et al.* Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota. *PLoS Biol.* **16**, e2006352 (2018).
3. Amicucci, M. J. *et al.* Strategy for Structural Elucidation of Polysaccharides: Elucidation of a Maize Mucilage that Harbors Diazotrophic Bacteria. *Anal. Chem.* **91**, 7254–7265 (2019).
4. Niu, B., Paulson, J. N., Zheng, X. & Kolter, R. Simplified and representative bacterial community of maize roots. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2450–E2459 (2017).



**Funding Statement:** The authors gratefully acknowledge funding from the U.S. Department of Energy (BER) grant no. DE-SC0021052.

# A Systems Understanding of Nitrogen-Fixation on the Aerial Roots of Sorghum

Saddie Vela[1*](saddievela@ufl.edu), Emily Wolf [1*] (reed.emily@ufl.edu), Jennifer Wilker,[2] Saptarshi Pyne,[2] April MacIntyre,[2] Sushmita Roy,[2] Jean-Michel Ané,[2] **Wilfred Vermerris[1]**

[1]University of Florida, Gainesville, FL
[2]University of Wisconsin – Madison, Madison, WI

https://sonar.bact.wisc.edu/

**Project Goals: The overall goal of this project is to understand better the molecular and cellular networks controlling biological nitrogen fixation in sorghum aerial roots using a combination of genetics, synthetic bacterial communities, and systems biology.**

Since the Green Revolution, the intensive use of synthetic fertilizers has been the most common strategy to alleviate the limitation of nitrogen (N) availability on crop yields. With the desire to produce bioenergy crops on low-productivity lands to avoid competition with food production, the heavy use of fertilizer necessary to ensure reasonable yields raises the cost of production. It increases the environmental footprint of producing bioenergy crops. Sorghum is an attractive bioenergy crop due to its ability to produce high biomass yields with limited inputs and withstand biotic and abiotic stresses. To further enhance sorghum's bioenergy potential, we are investigating the mechanisms that enable specific sorghum accessions to support symbiotic interactions with N-fixing microbes in the mucilage of their aerial roots, similar to what has been observed for tropical maize in Mexico[1]. We are employing a genome-wide association study of two panels of genetically diverse genotypes, the sorghum minicore[2], a collection of 242 landraces, and the sorghum association panel (SAP)[3], a collection of 388 sorghum genotypes representing all major cultivated races and important U.S. breeding lines and their progenitors. Traits of interest include the number of nodes producing aerial roots, the total number of aerial roots, aerial root length, and aerial root diameter. We are also investigating the effect of the nitrogen fertilization level and location (Florida vs. Wisconsin) on these traits. Based on the initial screening in the summer and fall of 2021, aerial roots were observed in 114 accessions of the minicore (47%), but only 18 accessions of the SAP (5%), suggesting aerial root formation is a trait that has been under negative selection in modern breeding programs. In parallel, we have initiated seven breeding populations from which biomass sorghum to support BNF will be selected. In addition, we are using single and bulk transcriptome profiling to investigate the sorghum gene regulatory networks that enable mucilage production and N-fixation and interactions of sorghum with diverse microbial species. From this analysis, we expect to infer cell type-specific gene regulatory networks and critical regulatory genes that drive the dynamics of these processes. Candidate genes that are important to enable biological nitrogen fixation (BNF) will be tested through reverse genetics approaches. This will be based on near-isogenic lines of RTx430 that are under development, in which alleles for loci enabling the formation of aerial roots have been introgressed. As a result of these activities, we will enhance our understanding of BNF in sorghum and have genetic tools at our disposal to enable the commercial development of N-fixing bioenergy sorghums to display high biomass productivity on marginal lands, with a reduction in both inputs.

1. Van Deynze, A. *et al.* Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota. *PLoS Biol.* **16**, e2006352 (2018).

2. Upadhyaya, H. D. *et al*. Developing a mini core collection of sorghum for diversified utilization of germplasm. Crop Sci. 49, 1769–1780 (2009).
3. Casa, A. M. *et al.* Community resources and strategies for association mapping in sorghum. Crop Sci. 48, 30–40 (2008).

*The authors gratefully acknowledge funding from the U.S. Department of Energy (BER) grant no. DE-SC0021052.*

**Rapid Design and Engineering of Smart and Secure Microbiological Systems**

Gyorgy Babnigg,[1] Jeffrey Dewey,[1] Fatima Foflonker,[2] Michael Fonstein,[1] Sara Forrester,[1] Stephanie Greenwald, [1] Christopher S. Henry,[2] Michael Irvin, [2] Jessica L. Johnson,[1] Peter E. Larsen,[1] Filipe Liu,[2] Carla M. Mann,[2] Sarah Owens, [1] Arvind Ramanathan,[2] Rebecca Weinberg,[1] Marie-Francoise Gros,[1,3] Philippe Noirot,[1,3] Tomoya Honda,[4,5] Zhiying Zhao,[4,5] Yasuo Yoshikuni, [4,5] Steven R. Fleming,[6] Ashty S. Karim,[6] Brenda Wang,[6] Michael C. Jewett,[6] Jaehyun Lee,[1,7] Mark Mimee,[7] and **Dionysios A. Antonopoulos**[1*] (dion@anl.gov)

[1]Biosciences Division, Argonne National Laboratory, Lemont, IL; [2]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL; [3]National Research Institute for Agriculture, Food and Environment (INRAE), France; [4]Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Joint Genome Institute, Berkeley, CA; [6]Northwestern University, Evanston, IL; [7]University of Chicago, Chicago, IL

**Project Goals: The long-term goal for this Project is to realize secure biodesign strategies for microbial systems that operate in the dynamic abiotic and biotic conditions of natural environments, thus enabling systems-level and rational biological design for field use. There are several key challenges to incorporating safeguard systems at the design stage including: (1) lack of knowledge for how well safeguards operate across the broad set of environmental and physiological conditions that an organism experiences; (2) a need to integrate the safeguard with other cellular components so that it can sense and recognize specific signals from the intracellular or extracellular environment, and mediate a response; and (3) a need for rapid and reliable methods to engineer and optimize the biological components for safeguard construction and functional integration. To address these challenges, we propose to utilize recent advances in the fields of synthetic biology, artificial intelligence (AI), and automation, which together pave the way for a paradigm shift in our understanding of the ways that cellular function can be designed at the level of bacterial communities.**

The design and application of successfully engineered biosystems requires an understanding of how engineered microbes will interact with other organisms – either as one-on-one competitors, for example, or in the context of microbial consortia. Engineering microorganisms from first principles for non-laboratory, environmental applications is inherently challenging because: (1) engineered systems tend to quickly revert back to their wild-type behaviors; and (2) these systems typically pay a price in reduced fitness making them uncompetitive against invasive contaminating species (i.e., metabolic burden). A key question is how do sensing, signaling, and metabolism contribute to the stabilization and destabilization of these interactions? Here, the organization, control, stabilization, and destabilization of natural and engineered microbes will be investigated through a synthetic biology approach. The approach will enable development of (1) single-strain systems capable of detecting and responding to target organisms in the environment; (2) a pipeline for refining and engineering biological constructs in new non-model host organisms; and (3) improved systems for the rapid designing, engineering, and assaying of new biological modules. This coupled approach to design and build safeguard systems for intrinsic biocontainment that are predictable and portable across bacterial species, would focus on microbes that are part of the beneficial plant microbiome. A long-term goal

beyond the proposed research is to enable the engineering of communities of microorganisms based on first principles of biological design that mimic the smart performance of microorganisms observed in natural systems. This will enable a new vision of biosecurity and biocontainment that harnesses the underlying mechanisms of resource management occurring within and between organisms.

*This Project is funded by the Biological Systems Science Division's Genomic Science Program, within the U.S Department of Energy, Office of Science, Biological and Environmental Research.*

**Developing "Smart" Single-strain Systems Capable of Detecting and Responding to Target Organisms in the Environment**

Gyorgy Babnigg[1*] (gbabnigg@anl.gov), Jeffrey Dewey,[1] Fatima Foflonker,[2] Michael Fonstein,[1] Christopher S. Henry,[2] Jessica L. Johnson,[1] Arvind Ramanathan,[2] Tomoya Honda,[3,4] Zhiying Zhao, [3,4] Yasuo Yoshikuni, [3,4] Steven R. Fleming,[5] Ashty S. Karim,[5] Brenda Wang,[5] Michael C. Jewett,[5] Jaehyun Lee,[1,6] Mark Mimee,[6] and **Dionysios A. Antonopoulos**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, IL; [2]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Joint Genome Institute, Berkeley, CA; [5]Northwestern University, Evanston, IL; [6]University of Chicago, Chicago, IL

**Project Goals: The long-term goal for this Project is to realize secure biodesign strategies for microbial systems that operate in the dynamic abiotic and biotic conditions of natural environments, thus enabling systems-level and rational biological design for field use. There are several key challenges to incorporating safeguard systems at the design stage including: (1) lack of knowledge for how well safeguards operate across the broad set of environmental and physiological conditions that an organism experiences; (2) a need to integrate the safeguard with other cellular components so that it can sense and recognize specific signals from the intracellular or extracellular environment, and mediate a response; and (3) a need for rapid and reliable methods to engineer and optimize the biological components for safeguard construction and functional integration. To address these challenges, we propose to utilize recent advances in the fields of synthetic biology, artificial intelligence (AI), and automation, which together pave the way for a paradigm shift in our understanding of the ways that cellular function can be designed at the level of bacterial communities.**

The design and application of successfully engineered biosystems requires an understanding of how engineered microbes will interact with other organisms – either as one-on-one competitors, for example, or in the context of microbial consortia. Engineering microorganisms from first principles for non-laboratory, environmental applications is inherently challenging because: (1) engineered systems tend to quickly revert back to their wild-type behaviors; and (2) these systems typically pay a price in reduced fitness making them uncompetitive against invasive contaminating species (i.e., metabolic burden). A key question is how do sensing, signaling, and metabolism contribute to the stabilization and destabilization of these interactions?

We are applying a model-driven design strategy to develop a "smart" platform strain that is: (1) capable of detecting and responding to the presence of an invasive strain in the community; (2) capable of eliminating another strain; and (3) retains its own internal safeguards so the platform strain can also be eliminated on demand. The development of a platform strain that is capable of targeted elimination of an invasive strain, followed by triggered self-destruction will have many practical applications of relevance to DOE, from plant growth health to biomanufacturing, as well as providing the base toolkit for ultimately engineering more complex microbial community interactions.

The research activities in this portion of our project are being conducted in the context of the following three scenarios: (1) building a strain which can be eliminated in a controlled manner

with an external chemical inducer (i.e., testing its safeguard system); (2) building a biosensor containing strain capable of sensing the presence of other organisms and then responding; and (3) building a strain that differentiates its population in a probabilistic fashion to mitigate metabolic burden in response to a signal. Accordingly, work has been divided into subtasks that increase the complexity of the proposed platform strain stepwise, by integrating new modules optimized *in vitro* and *in vivo* and regulatory responses. These circuits are then integrated into target microbes via the chassis-independent recombinase-assisted genome engineering (CRAGE) method [1]. The final set of subtasks are focused on developing the modeling approaches for predicting the overall system behavior arising from each tested platform design. These modeling methods are to be applied to design optimal combinations of biosensor properties, response system activation, and the response itself to achieve a desired overall effect. More importantly, this system establishes many foundation stones for the development of more complex synthetic biology-based systems.

## References

1. Wang, G., et al., *CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria*. Nature Microbiology, 2019. 4: p. 2498-2510.

*This Project is funded by the Biological Systems Science Division's Genomic Science Program, within the U.S Department of Energy, Office of Science, Biological and Environmental Research.*

**Guiding Data-Driven Integrative Design of Secure Biological Systems with Artificial Intelligence Techniques**

Carla M. Mann[1,3]* (cmann@anl.gov), Michael Irvin,[1,3] Rebecca Weinberg,[2] Gyorgy Babnigg,[2] Christopher Henry,[1] Dionysios A. Antonopoulos,[2] and **Arvind Ramanathan**[1]

[1]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL; [2]Biosciences Division, Argonne National Laboratory, Lemont, IL; [3]These authors contributed equally

**Project Goals: The long-term goal for this Project is to validate strategies for secure microbial systems that operate in dynamic environmental conditions, thus enabling systems-level and rational biological design for a range of applications. There are several key challenges to incorporating safeguard systems at the design stage including: (1) lack of knowledge for how well safeguards operate across the broad set of environmental and physiological conditions that an organism experiences; (2) a need to integrate the safeguard with other cellular components so that it can sense and recognize specific signals from the intracellular or extracellular environment, and mediate a response; and (3) a need for rapid and reliable methods to engineer and optimize the biological components for safeguard construction and functional integration. To address these challenges, we propose to use recent advances in the fields of synthetic biology, artificial intelligence (AI), and automation, which together pave the way for a paradigm shift in our understanding of the ways that cellular function can be designed at the level of bacterial communities.**

Developing organismal systems to apply rational biological design to areas of need necessitates development of mechanisms to safely contain those organisms to protect the environment and public. Rational design of biological modules (e.g., biosensors, novel enzymes for biosynthesis and/or degradation, control circuits) to enable stability in engineered microbial systems also presents challenges due to the size of potential design-search spaces. For example, optimizing a 10 amino acid span of a 300 residue-length protein has $10^{20}$ potential combinations embedded within a highly non-linear (and potentially sparsely sampled) space. Scaling this type of optimization problem to the level of pathways and cellular systems produces an effectively infinite search space. This requires developing AI frameworks to intelligently explore that space to reduce time and effort needed to conduct Design, Build, Test, and Learn (DBTL) cycles.

As part of this optimization process, we are developing a bacterial "self-destruct" mechanism that activates under specific environmental conditions. This mechanism relies on a gRNA in the bacterium that self-targets a genomic locus such that the gRNA has low activity in the lab, but under changing cellular conditions in response to a specific environmental condition, becomes highly active. As gRNA activity can widely vary between target locations within the same gene [1, 2], and the same location targeted under different physiological conditions can lead to gRNA activity changes orders of magnitude in size, machine learning (ML) and AI strategies are needed to elucidate rules governing CRISPR/Cas gRNA activity under changing conditions. In order to optimally identify gRNAs capable of producing desired behavior, we developed a deep-learning model, CRISPRAct, that combines a natural language model with a neural network (NN) model to predict the fold change in cell population as a proxy for gRNA activity. This model utilizes Google AI's ALBERT [3] architecture by treating the 21 bp upstream and 21 bp downstream of the PAM site as "sentences" composed of seven "words" which themselves

comprise three bases. Our model is pre-trained on representative *Escherichia coli* genomes, then fine-tuned while a regression model is built on top of the pre-trained model. The NN model uses 444 features comprising physicochemical and positional features [1] along with environmental and cellular conditions (e.g., media type and growth phase). Predictions from the models are combined through polynomial regression to predict the fold change in cell population for a given gRNA. CRISPRAct produces a Mean Absolute Error (MAE) of 0.39 and Spearman Correlation Coefficient of 66.4% at different timepoints in variable physiological conditions.

Environment-responsive systems are not only useful in biocontainment, but also lay foundations for further development of self-contained biological modules that respond to cellular signals. We are thus developing biological modules that interpret and respond to complex dynamics of intracellular and extracellular signaling interactions. To accomplish this, we combine AI, ML, and dynamical modeling approaches to capture the dynamics of our biological modules and their impact on phenotypes. Our dynamical models (DMs) encode, as first-principles, complex networks of interactions between proteins and their regulatory mechanisms that underlie functions targeted in rational biological design objectives. We leveraged an ML approach that uses feature selection to identify a testable, low-dimensional representation of intracellular signaling to guide more efficient exploration of the design search space [4]. We also developed convolutional autoencoders (CAEs) that enable efficient approximation of computationally-expensive DM simulations. The AI model treats multivariate results of DM simulations as images wherein each row of pixels represents the simulated time course of a cellular component. Using a DM with 14 observable proteins and 100 simulation timepoints, we generated 10,000 1400-pixel images (each representing distinct configurations of the initial amount and activity of proteins) which were compressed via a CAE to a 196-dimensional latent space. Hyperparameters (e.g. the dimensionality of the latent space) were optimized using Optuna [5] which achieved a minimum running Mean Squared Error (MSE) loss of 0.02. This approach provides compressed latent representations of intracellular signaling dynamics, which enables them to readily support rational biological design by guiding more efficient exploration of an otherwise sparse nonlinear design-search space. The dynamical modeling and AI approach were developed in tandem and are part of a larger modeling and optimization loop that offers a synergistic impact on our model-based design objectives.

## References


1. Guo, J., et al., *Improved sgRNA design in bacteria via genome-wide activity profiling*. Nucleic Acids Res, 2018. 46(14): p. 7052-7069
2. Gutierrez, B., et al., *Genome-wide CRISPR-Cas9 screen in E. coli identifies design rules for efficient targeting*. bioRxiv, 2018: p. 308148
3. Lan, Z. et al., *ALBERT: A Lite BERT for self-supervised learning of language representations*. ICLR, 2020.
4. Irvin, M., et al. *The misleading certainty of uncertain data in biological network processes*. bioRxiv 2021.
5. Akiba, T., et al. *Optuna: A next-generation hyperparameter optimization framework*. Proc 25th ACM SIGKDD 2019.



*This Project is funded by the Biological Systems Science Division's Genomic Science Program, within the U.S Department of Energy, Office of Science, Biological and Environmental Research.*

# CRISPR-Act: AI-guided Prediction of a CRISPR Kill-switch Across Physiological Contexts

Rebecca Weinberg[1,4*] (rweinberg@anl.gov), Carla M. Mann,[2,4] Gyorgy Babnigg,[1] Sara Forrester,[1] Stephanie Greenwald, [1] Peter E. Larsen,[1] Sarah Owens, [1] Marie-Francoise Gros,[1,3] Philippe Noirot,[1,3] Arvind Ramanathan,[2] and **Dionysios A. Antonopoulos**[1]

[1]Biosciences Division, Argonne National Laboratory, Lemont, IL; [2]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL; [3]National Research Institute for Agriculture, Food and Environment (INRAE), France; [4]These authors contributed equally

**Project Goals: The long-term goal for this Project is to realize secure biodesign strategies for microbial systems that operate in the dynamic abiotic and biotic conditions of natural environments, thus enabling systems-level and rational biological design for field use. There are several key challenges to incorporating safeguard systems at the design stage including: (1) lack of knowledge for how well safeguards operate across the broad set of environmental and physiological conditions that an organism experiences; (2) a need to integrate the safeguard with other cellular components so that it can sense and recognize specific signals from the intracellular or extracellular environment, and mediate a response; and (3) a need for rapid and reliable methods to engineer and optimize the biological components for safeguard construction and functional integration. To address these challenges, we propose to utilize recent advances in the fields of synthetic biology, artificial intelligence (AI), and automation, which together pave the way for a paradigm shift in our understanding of the ways that cellular function can be designed at the level of bacterial communities.**

The development of genetically engineered organisms necessitates the creation of secure and efficient biocontainment systems to safely contain such organisms and thus protect the environment, public health, and public perception of scientific research. Microbial safeguards based on controlled activation of a self-targeting CRISPR/Cas9 "self-destruct" mechanism are transferable between different organisms, cost-effective, and relatively easy to implement. However, the variability of CRISPR/Cas cleavage efficiency within a genome (and even within the same gene) [1, 2] represents a challenge in choosing efficient self-targeting guide RNAs (gRNAs) for self-destruction, particularly as gRNA efficiency also varies under different environmental conditions. We have developed a machine-learning prediction method, CRISPRAct, that predicts gRNA efficiency across different environmental conditions to assist in identifying candidate gRNAs for use in secure biosystems.

We hypothesized that dynamic gene expression responses to varying physiological conditions would influence the cell-killing activity of the CRISPR/Cas9 system. To test this hypothesis, we screened the activity of a library of 180,000 gRNAs spanning the *E. coli* MG1655 genome, and compared cell-killing activity to control sequences with no genomic matches (~20,000). We used the log2 fold change in cell population between time points as a proxy for gRNA cutting activity. To assess the influence of physiological conditions on gRNA cutting, screens were conducted as time courses in three growth conditions: rich media in exponential growth (LB-E); defined media in exponential growth (M9-E); and rich media in stationary phase (LB-S).

In our initial library screens, we identified ~6,000 guides that were statistically overrepresented (via ANOVA testing) for physiology-specific functions. A small subset of

gRNAs (174) were "outlier switches" that were statistical outliers in their outstanding killing activity in one or more physiological conditions, but were also statistical outliers in their lack of activity under the other physiological condition(s). Additionally, there is a marked difference in GC content between guides that efficiently kill in the M9 media versus the LB media: guides that have low activity in rich media (LB) but high activity in minimal media (M9) are GC rich, while guides that have high activity in rich media low activity in minimal media are very AT rich. Further, these "switch" guides tend to localize to specific areas in the genome – there are six regions that active rich media/inactive minimal media guides localize to, while inactive rich/active minimal media guides localize to three narrow regions and two broad swathes within the genome. We also identified a small but statistically significant correlation between the number of sites a gRNA can target in the genome, and the gRNA's cutting efficiency, and correlations of varying strengths between the proximity of a gRNA target site to nucleoid-associated protein binding motifs in the primary genomic sequence. Collectively, the initial round of 200k library screens generated an extensive dataset of more than 530,000 data points used to develop CRISPRAct.

The CRISPRAct model is a two-part model that combines a natural language processing (NLP) model with a neural network (NN) model. The NLP model treats genomic sequences as a machine-interpretable "language", while the NN model features now incorporate gRNA positional and physicochemical properties (including nucleoid-associated protein binding motif proximity) with environmental conditions. The outputs of these models are combined through polynomial regression to predict the percent fold change of guide prevalence after Cas9 induction, as a proxy for the gRNA activity. CRISPRAct achieves a Mean Absolute Error of 0.39 and Spearman Correlation Coefficient of 66.4%, beating the correlation from the previous state of the art (in a single environmental condition) by about 12%. Importantly, CRISPRAct exhibited a comparable Mean Absolute Error across physiological conditions- 0.51 in LB-E, 0.52 in LB-S and 0.54 in M9-E. CRISPRAct is thus, to our knowledge, the first gRNA activity predictor capable of predicting behavior under different environmental conditions. We are currently assessing reproducibility of our screens. Empirically assessed correlations between screens varied by physiological condition, which informs our strategy to develop transfer learning. Ultimately, we anticipate that leveraging these physiological variations while training CRISPRAct will improve the robustness of our models and reduce costly retraining time as we move into novel genomic contexts.

## References

1. Guo, J., et al., *Improved sgRNA design in bacteria via genome-wide activity profiling*. Nucleic Acids Res, 2018. 46(14): p. 7052-7069.
2. Gutierrez, B., et al., *Genome-wide CRISPR-Cas9 screen in E. coli identifies design rules for efficient targeting*. bioRxiv, 2018: p. 308148.

*This Project is funded by the Biological Systems Science Division's Genomic Science Program, within the U.S Department of Energy, Office of Science, Biological and Environmental Research.*

**Title:** Probing Lignin Deconstruction and Catabolism in Soil *Pseudomonas* species

**Authors:** Rebecca A. Wilkes,[1] Allison Werner,[2] Neha P. Kamat,[1] Gregg Beckham,[2] and **Ludmilla Aristilde[1]** (ludmilla.aristilde@northwestern.edu),

**Institutions:** [1]Northwestern University, Evanston, IL; [2]National Renewable Energy Laboratory, Golden, Co

**Project Goals:** The overarching research goals are to elucidate the reaction network responsible for the metabolic functionalities of the outer membrane vesicles (OMVs) in soil *Pseudomonas* species with demonstrated accelerated lignin catabolism and subsequently evaluate the metabolic relationship of OMVs in fueling lignin-derived carbon fluxes towards intracellular biosynthetic pathways.

**Abstract Text:** The first objective of this project is to elucidate lignin deconstruction and carbon assimilation flux using whole-cell cultures. The working hypothesis is that the species optimized for coupled OMV-mediated lignin deconstruction and intracellular assimilation of lignin derivative will incorporate lignin carbons into cellular metabolites, whereas the species with only OMV-mediated lignin deconstruction will accumulate lignin derivatives in the spent media. In preliminary experiments, cells of different *Pseudomonas* species (*P. putida* KT2440, *P. putida* S12, and *P. protegens* Pf-5) are grown in nutrient solutions containing different lignin structures [lignosulfonate (LS) and lignin alkali (LA)] supplemented without or with glucose (a sugar) or acetate (a short-chain carboxylic acid). In addition to monitoring cell growth and carbon consumption from the lignin compounds, we are monitoring changes in the lignin structure. Thus far, we have used Fourier-transform infrared spectroscopy (FTIR) to track the vibrational peak (1585 $cm^{-1}$) associated with the lignin aromatic skeleton, which disappeared in the LS structure after two-week exposure to all three species. These results imply the breakdown of the LS aromatic structure. We performed total organic carbon analysis to determine the extent to which LS-derived carbon was consumed. We plan to apply nuclear magnetic resonance spectroscopy and high-resolution liquid chromatography-mass spectrometry analysis to identify modification in the LS structure and specific LS-derived breakdown products, respectively. We are also developing mixed isotope tracer experiments to capture the intracellular assimilation of lignin breakdown products.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022181.

**KBase: A case study illustrating the derivation and testing of mechanistic connections between geochemistry and the microbiome**

**Adam P. Arkin**[1], **Robert Cottingham**[3], **Chris Henry**[2], Paramvir S. Dehal (psdehal@lbl.gov)[1*], Benjamin Allen[3], Jason Baumohl[1], Kathleen Beilsmith[2], David Dakota Blair[4], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Ellen Dow[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Jason Fillman[1], Tianhao Gu[2], AJ Ireland[1], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Filipe Lui[2], Dan Murphy-Olson[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyneth Terry[1], Charles Trenholm[1], Sumin Wang[1], Doreen Ware[5], Pamela Weisenhorn[2], Elisha Wood-Charlson[1], Ziming Yang[4], Shinjae Yoo[4], Qizhi Zhang[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Argonne National Laboratory, Argonne, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4]Brookhaven National Laboratory, Upton, NY; [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
http://kbase.us

**Project Goals: The Department of Energy Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer-extensible platform that enables scientists to analyze their own data within the context of public data and share their findings across the system.**

DOE is investing in multiple projects that attempt to understand the mechanistic processes of environmental ecology. These large-scale projects typically have multiple field sites that are gathering diverse forms of data including amplicons, metagenomes, extracted isolates, and geochemistry across time and space. For example, the ENIGMA project has field sites at Oak Ridge National Laboratory aimed at mechanistically understanding the microbial biogeochemical processes within contaminated subsurface sediment and has collected data across hundreds of wells and sediment cores resulting in over 4000 samples. ENIGMA and similar projects cross correlate taxa and geochemistry to infer the taxa and genomic functions that are constrained in their locational abundance and activity by environmental conditions and which in turn transform those environments. This type of research generally involves teams performing different measurements across the same samples collaboratively integrating and analyzing the resultant data using a plethora of complex computational approaches.

Here, we illustrate how innovations within the KBase platform can help support projects like ENIGMA. This analysis leverages new functionality delivered over the past year, including a new representation of Samples (See poster ***KBase: Significant Improvements***) with the ability to upload complex organized sets of Samples with detailed sample attributes describing the environment and sampling process. These attributes are validated and ontologically described to facilitate data sharing, reuse and downstream analysis. For each sample, data representing measurements of the Sample, like geochemistry, 16S amplicon reads, shotgun metagenome reads

and isolates are directly linked, in system, to their corresponding sample. This allows us to cross-correlate geochemistry, taxa, and gene function, as well as to evaluate relationships using isolates prioritized in our analyses. Over the last year we have also added analytical functionality for many of these data types that enable rich exploration of microbial ecological questions. With the ENIGMA data we demonstrate how to: 1) capture and organize multiple forms of measurement data linked to samples; 2) explore this data to identify ubiquitous taxa; 3) correlate across amplicons and geochemistry to identify key taxa linked to biogeochemical processes of interest; 4) leverage metagenomic data to understand the gene functions, genome organization and composition of key taxa, and the identification of isolates of interest for follow-up in laboratory experiments; and 5) build metabolic models of these isolates and MAGs that allow us to test these relationships. Together, these operations drive towards a generalized ability to mechanistically link environmental attributes to genes and genomes using a formal modeling framework to generate and evaluate hypotheses. Furthermore, the use of KBase enables the ENIGMA team to to share and collaborate on analysis, ensuring good stewardship of their products. They can also publish their data, tools, analysis and final results and evaluation of these results in a larger context of work done by the environmental microbiology community that is shared on the KBase system and thus can be included in their analysis.

Samples collected from the ENIGMA project were organized by field campaign and uploaded into KBase using the newly developed Samples framework. This capability allows KBase to capture critical metadata associated with samples that can be used to make important inferences in downstream analysis. These samples have a variety of extracted measurement data that have also been uploaded into KBase. Metagenomic reads and assemblies, metagenome assembled genomes, geochemistry, and 16S amplicon data are linked in KBase to corresponding samples. This allows users to easily see all of the measured data of a sample. Related Samples are grouped together in Sample Sets.

We introduced new analysis tools in KBase to assign taxonomy to marker gene sequences, filter and transform taxonomic abundance matrices, and correlate amplicon abundances with chemical abundance measurements in environmental samples. An amplicon from the prevalent Rhodanobacter genus was highly correlated with nitrous oxide concentration in ENIGMA samples, prompting us to examine denitrification in Rhodanobacter and co-occurring microbes. To further our understanding of these communities, we have assembled shotgun metagenomic reads, extracted metagenome assembled genomes, and scanned their functional repertoires to identify key lineages involved in nitrate reduction.

We constructed draft metabolic models of the identified key isolates and merged these models into a compartmentalized community model to mechanistically explore the microbial community dynamics. Model predictions are able to capture individual species' contributions to dynamics in nitrate reduction relating to energy biosynthesis and co-factor recycling with changes in the abundance of nitrate in the environment.

This work is supported as part of the BER Genomic Science Program. The DOE Systems Biology Knowledgebase (KBase) is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

**How KBase Supports Education**

**Adam P. Arkin**[1], **Robert Cottingham**[3], **Chris Henry**[2], Ellen Dow[1*] (egdow@lbl.gov), Benjamin Allen[3], Jason Baumohl[1], Kathleen Beilsmith[2], David Dakota Blair[4], Jay Bolton[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Paramvir Dehal[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Jason Fillman[1], Tianhao Gu[2], AJ Ireland[1], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Filipe Lui[2], Dan Murphy-Olson[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyneth Terry[1], Charles Trenholm[1], Sumin Wang[1], Doreen Ware[5], Pamela Weisenhorn[2], Elisha Wood-Charlson[1], Ziming Yang[4], Shinjae Yoo[4], Qizhi Zhang[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Argonne National Laboratory, Argonne, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4] Brookhaven National Laboratory, Upton, NY; [5] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

http://kbase.us

**The Department of Energy Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer extensible platform that enables scientists to analyze their own data within the context of public data and share their findings across the system.**

The KBase user interface (UI) enables instructors to work with students to conduct hands-on data science research and analysis without the need for programming skills or computational resources. The KBase team works with instructors and researchers of varying skill and career levels to ensure the transfer of domain knowledge is accompanied by an understanding of bioinformatic tools and techniques. KBase has been supporting education in three main ways: establishing the KBase Educators program, participating in the Environmental Molecular Sciences Laboratory Summer School, and hosting a variety of topic-based workshops and webinars. Each program supports a different cross-section of the BER research community, with the overall goal of improving and expanding the next-generation of data analysis using KBase.

**KBase Educators**

The KBase Educators program (https://www.kbase.us/kbase-educators/) consists of biological and data science instructors, ranging from High School to Graduate level, that have adapted the KBase platform to their curriculum needs by developing modular, adaptable, and customizable instructional units using KBase Narratives. These instructional modules contain teaching resources, data, analysis tools, and mark-down utility to tailor instructions and learning goals. Each module can be adapted for independent class concepts that involve Genomics, Metagenomics, Phylogenetics, Pangenomics, and/or Metabolic Modeling. The KBase Educators Organization provides access to resources in KBase, and a KBase Users Slack group provides access to a community network of peers, supported by community-driven guidelines,

instructional templates, and KBase staff. Working groups are short-term commitments around topics that the community identifies as important or relevant to their teaching goals. The aim behind community working groups (https://www.kbase.us/educator-working-groups/) is to enable educators using KBase to collaboratively build resources to add to the Educators Org, while dividing the workload and providing accountability to deliver resources. The KBase team helps facilitate working group formation and setting up the initial kick-off meeting to ensure everyone is familiar with the scope and goal, providing support during development and release of materials.

**EMSL Summer School**

KBase collaborates with and supports instruction for the Environmental Molecular Sciences Laboratory (EMSL) Summer School for early career researchers. National laboratory and academic researchers provide lectures on multidisciplinary research over the course of a week, which are free and open to the public. In addition, post-doctoral researchers and Ph.D. students attended tutorial classes, which provided hands-on instruction and learning. KBase contributes to both lectures and hands-on training on the complex analysis of biological data. During the 2020 (https://www.kbase.us/multiscale-microbial-dynamics-modeling/) and 2021 (https://www.kbase.us/multi-omics-modeling-of-biochemical-pathways/) programs, KBase was used for analysis and sharing data on diverse topics using bacterial and fungal data.

**KBase Workshops and Webinars**

For special topics and collaborations, KBase hosts workshops that address needs of research groups, educators, and classes. Workshops target a smaller class size to better facilitate hands-on tutorials with attendees and students. For the broader audience of all KBase Users, KBase holds live webinars throughout the year with a live question & answer section, which introduce new features and topics with speakers including community developers and subject matter experts. Webinars are posted on the KBase YouTube channel (https://www.youtube.com/DOEKBase) for anyone to revisit and view after the event and often include public Narratives for users to test out new tools and workflows.

Through each of these approaches, KBase empowers skilled researchers and inspires the next generation of biologists and data scientists by providing a platform that seamlessly enables users to integrate conceptual knowledge with sophisticated systems biology investigative tools.

1. Dow EG, Wood-Charlson EM, Biller SJ, Paustian T, Schirmer A, Sheik CS, Whitham JM, Krebs R, Goller CC, Allen B, Crockett Z and Arkin AP. Bioinformatic Teaching Resources – For Educators, by Educators – Using KBase, a Free, User-Friendly, Open Source Platform. Front. Educ. 2021; 6:711535. doi: 10.3389/feduc.2021.711535

This work is supported as part of the BER Genomic Science Program. The DOE Systems Biology Knowledgebase (KBase) is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

# KBase: Significant Improvements to the DOE Systems Biology Knowledgebase in 2021

**Adam P. Arkin**[1], **Robert Cottingham**[3], **Chris Henry**[2*], Benjamin Allen[3], Jason Baumohl[1], Kathleen Beilsmith[2], David Dakota Blair[4], Jay Bolton[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Paramvir Dehal[1], Ellen Dow[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Jason Fillman[1], Tianhao Gu[2], AJ Ireland[1], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Filipe Lui[2], Dan Murphy-Olson[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyneth Terry[1], Charles Trenholm[1], Sumin Wang[1], Doreen Ware[5], Pamela Weisenhorn[2], Elisha Wood-Charlson[1], Ziming Yang[4], Shinjae Yoo[4], Qizhi Zhang[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Argonne National Laboratory, Argonne, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4]Brookhaven National Laboratory, Upton, NY; [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

http://kbase.us

**Project Goals: The Department of Energy Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer extensible platform that enables scientists to analyze their own data within the context of public data and share their findings across the system.**

Over the last year, substantial new improvements were integrated into the KBase platform, with the goal of empowering upload, integration, and analysis of large-scale multi-omics datasets progressing toward a mechanistic understanding of environmental microbiomes. Given their complexity and diversity, the study of environmental microbiomes may require the analysis of hundreds of samples, often with multi-omics data (e.g., metagenome, amplicon, metabolomics, etc.) collected for each sample. This creates multiple significant challenges in supporting the analysis of such data on a platform like KBase.

First, hundreds of disparate data files must be loaded and integrated into the KBase data-model, a task that was previously quite tedious as most upload tools in KBase operated on one file at a time. A new bulk-upload pipeline has been established, enabling users to load hundreds of objects at once. These tools are now being applied by the Genome Resolved Open Watersheds (GROW) CSP and Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) SFA to load hundreds of metagenomes and thousands of MAGs.

A second challenge involves the association of distinct data types collected from common experimental samples to: (1) allow for the organization of all data related to a complex experiment and (2) organize the data in such a way as to enable sophisticated cross-analysis of various data types. Toward this end, a Samples management system was integrated into KBase, enabling users to load sample metadata into KBase and associate sample IDs with data objects

on system (e.g., linking together the metagenome, metabolome, and MAGs derived from a common sample to facilitate integrated analysis). Tools are now being developed that exploit the relationship amongst data created by these sample associations to permit complex statistical analysis of data. We developed our Samples metadata, templates, and ID management systems in close collaborations with our partners at EMSL, JGI, NMDC, and ESS-DIVE.

A third challenge is handling all the datatypes required for multi-omics microbiome analysis, and to meet this challenge, this year KBase added deeper support and analysis pipelines for metabolomic and amplicon data. Now users can upload these new datatypes and integrate them with their broader microbiome data as well as reference data generally available on KBase (e.g., MGnify). Metabolomics and expression data can be visualized on rich metabolic maps, including new genome-wide maps for plants and fungi. Amplicon data can be visualized, compared, used to predict microbiome functional potential, and analyzed in the context of environmental data related to samples.

A fourth challenge is the difficulty associated with annotating protein functions within microbiomes. To improve support for this activity, and in collaboration with our User Working Groups, new tools have been added to KBase in the area of genome annotation. Users can now upload and compare many alternative annotations using tools developed by the LLNL Biofuels SFA; they can build models from those annotations and compare performance of competing annotations on predicting phenotypes like Biolog growth profiles; and they can apply new annotation algorithms like DRAM (integrated into KBase by the Wrighton lab) (1) and the Exascale Analysis tool (integrated into KBase by the Jacobson lab). Tools for metabolic reconstruction are improving dramatically for plants, fungi, and microbes, including new support for visualizing and painting data on metabolic reconstructions. All of this culminates in KBase offering a sophisticated set of tools to load, store, create, compare, test, and visualize annotations.

While much still remains to be done, this new functionality has significantly lowered the barrier for KBase users to develop a deeper, integrated, mechanistic understanding of complex environmental microbiome data based on rich and diverse experimental datasets.

**References**


1. Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Research. 48:16. 8883–8900. https://doi.org/10.1093/nar/gkaa621.



This work is supported as part of the BER Genomic Science Program. The DOE Systems Biology Knowledgebase (KBase) is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

**Science Focus Areas use KBase to share new tools and data**

**Adam P. Arkin**[1], **Robert Cottingham**[3], **Chris Henry**[2], Elisha Wood-Charlson[1*], Benjamin Allen[3], Jason Baumohl[1], Kathleen Beilsmith[2], David Dakota Blair[4], Jay Bolton[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Paramvir Dehal[1], Ellen Dow[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Jason Fillman[1], Tianhao Gu[2], AJ Ireland[1], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Filipe Lui[2], Dan Murphy-Olson[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyneth Terry[1], Charles Trenholm[1], Sumin Wang[1], Doreen Ware[5], Pamela Weisenhorn[2], Ziming Yang[4], Shinjae Yoo[4], Qizhi Zhang[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Argonne National Laboratory, Argonne, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4]Brookhaven National Laboratory, Upton, NY; [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

http://kbase.us

**Project Goals: The Department of Energy Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery environment designed for both biologists and bioinformaticians. KBase integrates a large variety of data and analysis tools, from DOE and other public services, into an easy-to-use platform that leverages scalable computing infrastructure to perform sophisticated systems biology analyses. KBase is a publicly available and developer extensible platform that enables scientists to analyze their own data within the context of public data and share their findings across the system.**

Science Focus Area (SFA) research occurs at the cutting edge of the Office of Science's Biological and Environmental Research (BER) strategic goals. However, dissemination of knowledge and the tools capable of extending that knowledge to other biological systems beyond traditional scientific publishing is still challenging. KBase, as a free, open source platform designed to support broad biological systems-level analyses and sharing, provides a unique opportunity for BER programs to expand their impact to the 20,000+ KBase users world-wide.

Here, we describe several SFA-KBase collaborative projects, ranging from efforts to explore strain variation in microbes, protein-ligand binding, or quantitative Stable Isotope Probing (qSIP); improving functional annotation or viral ecogenomics analyses; and engineering microbial communities or community function(s) to support resilience in a time of change. Completed projects will highlight contributions and impact, and newer collaborations will showcase what is to come. SFA projects are also included as KBase Collaborations under the KBase User Working Group website: https://www.kbase.us/research/user-working-groups/.

*SFA-KBase Collaborative projects (alphabetical):*

**Bacterial-Fungal Interactions (LANL)**: Add novel taxonomic annotation tools for bacterial and fungal sequences in KBase (released in KBase: GOTTCHA2, Centrifuge). More information: https://www.kbase.us/research/chain-sfa/.

**Biofuels (LLNL)**: Improve on existing functional annotation tools by allowing for multiple annotations to be uploaded, compared, and merged, in support of creating more robust metabolic models (released in KBase: import, compare, and merge metabolic annotations). Coming soon: probabilistic annotation and ensemble metabolic modeling. More information: https://www.kbase.us/research/stuart-sfa/.

**ENIGMA (LBNL)**: Establish reference-based metagenome workflow to enable detection and annotation of strain variants (released in KBase: Meta-Decoder Call Variants, Strain Finder v1, Fama Profiling). Coming soon: Long read assembly tools for microbial isolates and metagenomes. More information: https://www.kbase.us/research/adams-sfa/.

**Microbes Persist (LLNL)**: Integrate the tools that enable work on viruses (released in KBase: vConTACT2, VirSorter, VirMatcher). Create functionality to detect and fix errors in assembled contigs (in development). Coming soon: Workflow to analyze quantitative stable isotope probe (qSIP) data and integration of *microTrait* to enable user-defined trait-matrix generation from a set of target genomes present in the same environment for functional comparison. More information: https://www.kbase.us/research/pett-ridge-sfa/.

**Persistence Control (PNNL)**: Coming soon: Improved protein annotation using machine learning, multi-omics data integration, and structural prediction. More information: https://www.kbase.us/research/egbert-sfa/.

**Plant Microbe Interfaces (ORNL)**: Coming soon: Design and omics exploration of synthetic microbial communities. More information: https://www.kbase.us/research/doktycz-sfa/.

**Protein Structure (ORNL)**: Establish the first protein structure importer and ligand-binding tools in KBase (released in KBase: CCMPredPy Contact Mapping, AutoDock Vina Protein-Ligand Docking, Import Protein Structure from PDB). More information: https://www.kbase.us/research/mitchell-sfa/. To extend this SFA effort, KBase is collaborating with PDB to further improve our protein structure workflows and datatypes in KBase. More information: https://www.kbase.us/research/protein-data-bank/.

**Soil Microbiome (PNNL)**: Coming soon: Omics-enabled global gapfilling (OMEGGA) for phenotype-consistent metabolic network reconstruction of microorganisms and communities. More information: https://www.kbase.us/research/hofmockel-sfa/.

KBase encourages new projects and collaborations! We provide information on the benefits of collaborating with KBase, and outline how to get started.

This work is supported as part of the BER Genomic Science Program. The DOE Systems Biology Knowledgebase (KBase) is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

**Title:** Monochromatic control of bacteria/yeast consortia for fuel and chemical production

**Authors:** Makoto A Lalwani,[1] Hinako Kawabe,[1] Rebecca L Mays,[1] Shannon M Hoffman,[1] **José L Avalos**[1,2,3]* (javalos@princeton.edu)

**Institutions:** [1]Department of Chemical and Biological Engineering, [2]Andlinger Center for Energy and the Environment, [3]High Meadows Environmental Institute, Princeton University, Princeton, New Jersey.

**Project Goals:** The overall goal of this project is to develop optogenetic controls over the population dynamics and composition of microbial consortia. This includes unidirectional monochromatic controls of one strain in the consortia; bidirectional monochromatic controls of two strains; and multidirectional polychromatic controls of two or more strains. We will apply these new capabilities to optogenetically control various consortia for biofuel and chemical production, using systems that showcase different advantages of co-culture fermentations. We will show how varying light duty cycles can optimize microbial populations throughout co-culture fermentations to maximize chemical production. Thus, this work addresses the long-standing challenge of stabilizing and optimizing the population composition of microbial consortia for chemical production.

**Abstract Text:** Engineered microorganisms have enormous potential to produce biofuels, chemicals, and other valuable products from renewable substrates and waste. However, the genetic and metabolic burden that exogenous biosynthetic pathways impose on engineered strains negatively impact their robustness and productivity. In addition, containing a full biosynthetic pathway in a single strain (i.e., in a monoculture) prevents metabolic engineers from separately optimizing each of its components (or modules), forcing them instead to optimize the entire pathway in a single cell by making compromises that often limit productivity. This shortcoming is magnified when different modules compete with or inhibit each other. These factors greatly contribute to the insufficient productivities of most bioprocesses, and prevent them from being competitive with fossil fuels and non-sustainable manufacturing processes.

Using co-cultures of more than one organism, instead of monocultures, has been proposed as a powerful solution to address these challenges. This approach draws inspiration from natural microbial processes (e.g., biomass decay, or food digestion), which are almost always carried out by microbial communities. Splitting biosynthetic pathways across multiple microorganisms would reduce the burden on each individual member and prevent negative interactions between different metabolic modules. Furthermore, the efficiency of each module would be significantly improved by optimizing them separately in specialized strains (akin to the way different parts of a chemical process are optimized in separate reactors in a chemical plant). However, stabilizing the population composition of engineered microbial consortia has proven to be a formidable challenge, with the fastest growing members usually dominating the consortia and turning them into futile monocultures. Controlling inoculation ratios and engineering clever biological

orthogonalities or synthetic symbioses between members have done much to improve co-culture operations, but these strategies have still been insufficient to advance engineered consortia to commercial processes.

In this proof-of-principle study, we demonstrate that optogenetics is an effective strategy to dynamically control populations in microbial co-cultures. Using a new optogenetic circuit we call OptoTA, we regulate an endogenous toxin-antitoxin system, enabling tunability of *Escherichia coli* growth using only blue light. With this system we can control the population composition of co-cultures of *E. coli* and *Saccharomyces cerevisiae*. When introducing in each strain different metabolic modules of biosynthetic pathways for isobutyl acetate or naringenin, we found that the productivity of co-cultures increases by adjusting the population ratios with specific light duty cycles. This study shows the feasibility of using optogenetics to control microbial consortia populations and the advantages of using light to control their chemical production.

**References/Publications**

1. Lalwani MA, Kawabe H, Mays RL, Hoffman SM, **Avalos JL**. Optogenetic control of microbial consortia populations for chemical production. *ACS Synthetic Biology* **10** (8), 2015-2029 (2021).

**Funding Statement:** This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022155.

**Synthetic Genetic Circuits to Engineer Water Use Efficiency, Photosynthetic Efficiency, and Biocontainment**

Jennifer A. N. Brophy[1], Degao Liu[2], Matt Zinselmeier[2], José R. Dinneny[1], Daniel F. Voytas[2*] (voytas@umn.edu) and **Ivan Baxter**[3]

[1]Stanford University, Palo Alto, CA; [2]University of Minnesota, St. Paul, MN; [3]The Donald Danforth Plant Science Center, St. Louis, MO

www.foxmillet.org

**Project Goals:** This project aims to leverage *Setaria viridis* as a model system to develop novel technologies and methodologies to redesign the bioenergy feedstock *Sorghum bicolor* to enhance water use and photosynthetic efficiencies.

Improving *Sorghum bicolor* as a biofuel crop requires methods to manipulate gene expression in ways that alter important agronomic traits. To achieve precise control over gene expression, we are building synthetic genetic circuits. These circuits enable precise spatial and magnitudinal control over gene expression and offer an exciting means to reprogram plant development and control growth. To build circuits, we've generated libraries of transcriptional regulators composed of bacterial DNA binding proteins or catalytically inactive dCas9 fused to transcriptional activation or repression domains. We tested thirty-six plant transcription factors in Arabidopsis and Setaria protoplasts for their ability to activate both reporters and native plant genes. A subset of the best performing transcriptional activators were introduced as transgenes into Arabidopsis and shown to activate expression of native genes. The best synthetic transcription factors were also used to construct circuits that perform Boolean logic operations. These circuits can regulate spatial patterns of gene expression across root tissues and, as a proof of concept, we used them to selectively modify branching in Arabidopsis roots. In parallel, synthetic transcription factors are being deployed in an innovative strategy for biocontainment of transgenes. We are identifying genes (target genes) that compromise the viability of *Sorghum bicolor* when overexpressed. We plan to introduce mutations into the target gene so that it is no longer recognized by a programmable transcription factor, then we will transform the mutant *Sorghum bicolor* with a synthetic circuit that leads to overexpression of the lethal gene, such that when the plant is outcrossed, its progeny are inviable. Taken together, these data show how synthetic genetic circuits can be built and used to successfully modify important plant traits.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018277.

**Title: Engineering enhanced photosynthesis and water use efficiency in Sorghum**

Varsha Pathare[1], Robert DiMario[1], Asaph B. Cousins[1]
Jennifer Brophy[2], Willian Viana[2], José R. Dinneny[2]
Niteen Kadam[3], James Fischer[3], Andrew D.B. Leakey[3]

[1]Washington State University, Pullman, WA 99163
[2]Stanford University, Stanford, CA 94305
[3]University of Illinois at Urbana-Champaign, Urbana, IL 61810

url: www.foxmillet.org

**Project Goals:** This project aims to leverage *Setaria viridis* as a model system to develop novel technologies and methodologies to redesign the bioenergy feedstock *Sorghum bicolor* to enhance water use and photosynthetic efficiencies.

**Abstract:** The yields of $C_4$ bioenergy crops are increasing through breeding and improved agronomy but the amount of biomass produced for a given amount of water use (water use efficiency) remains unchanged. Therefore, our research focuses on three major control points of water-limited production, with a systems-approach to biodesign: (1) greater photosynthetic carbon assimilation; (2) reduced water use through greater stomatal resistance to water loss; and (3) enhanced acquisition of available water by roots.

Plant cell wall structural polymers (e.g., cellulose, hemicelluloses, and lignins) influence biofuel digestibility as well as $CO_2$ and water movement within the leaf. For example, both $CO_2$ and water conductance are influenced by cell wall thickness and porosity. We will present data testing the hypothesis that changes in the chemical properties of mesophyll and bundle sheath cell walls will influence $CO_2$ and water conductance within the leaf, and therefore photosynthetic and whole plant water use efficiency. We have conducted experiments with grasses genetically modified for cell walls with differences in glucuronoarabinoxylan composition and mixed-linkage glucans. This research will improve our understanding of how cell wall modifications needed for cellulosic biofuel production influence photosynthetic water use efficiency in important $C_4$ crops.

Recent increases in atmospheric [$CO_2$] means that $C_4$ crops increasingly have greater $CO_2$ supply than is needed to saturated photosynthesis. Therefore, reducing stomatal conductance by reducing the number or size of stomata can increase intrinsic water use efficiency without necessarily suffering a trade-off of reduced photosynthetic $CO_2$ fixation. We are testing a series of orthologs of Arabidopsis stomatal developmental genes to determine the best solution for reducing stomatal conductance without unwanted pleiotropic effects that can alter stem or reproductive development. In addition, we have performed transcriptomic profiling of developing leaves to identify the network of genes controlling differentiation of epidermal cells.

Roots represent the supply side of plant-water relations. Root architecture, which is the branching pattern of the root system in soil, determines the efficiency that water and nutrients are accessed, but also represent a cost to the plant in terms of carbon. Previous work by our group has shown that grasses exhibit a suppression of crown root development under drought and a dramatic induction of their growth upon rewatering. Our current work is focused on identifying the genes necessary for these responses and the design of a synthetic biology approach to finely tune root branching. Work that will be presented includes the cloning of a novel locus controlling the initiation of Setaria root development under well-watered conditions and the establishment of a synthetic biology toolkit that enables two-input logic gates to be constructed in plants that create predictable changes in gene expression and modify root architecture.

**Funding statement:** This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC SC0018277).

**Title:** Advanced Phenomic Approaches to Improving Water-Use Efficiency in Bioenergy Grasses

**Authors:** Therese LaRue[1], Heike Lindner[1], **Ivan Baxter**[2], José R. Dinneny[1] (dinneny@stanford.edu), Andrew D.B. Leakey[3], Todd Mockler[1] Darshi Bannan, Rachel Paul, Parthiban Prakash, Charles Pignon, Collin Luebbert[2], Allen Hubbard[2], Jennifer Barrett[2]

**Institutions:** [1]Stanford University, Stanford, CA; [2]Donald Danforth Plant Science Center, St. Louis, MO; [3]University of Illinois at Urbana-Champaign, Urbana, Illinois

**Website URL:** www.foxmillet.org

**Project Goals:** This project aims to leverage *Setaria viridis* as a model system to develop novel technologies and methodologies to redesign the bioenergy feedstock *Sorghum bicolor* to enhance water use and photosynthetic efficiencies. Through the advancement of phenotyping platforms we are developing a quantitative and integrative understanding of how multiple shoot and root traits vary genetically and in response to water-deficit.

**Abstract Text:** Crop yields are frequently limited by the availability of water during the growing season; however, efforts to improve the acquisition and efficient use of water by crops have been challenged by the difficulty of rapidly phenotyping traits that control these processes.We have established a comprehensive portfolio of tools, methodologies and phenotyping platforms that are enabling new levels of insight into root, photosynthetic and stomatal traits. The resulting knowledge of associations between genotype, phenotype and environment are informing our efforts to engineer greater water use efficiency.

We have leveraged automated imaging, weighing and watering to monitor biomass, water use and other growth parameters over time. We have combined these approaches with biochemical profiling (transcriptomic, metabolomic, elements and isotopic ratios) and pan-genome variation analyses to get an in depth understanding of how plants respond to water deficit.

Advances in phenotyping leaf morphological and physiological traits important to water use efficiency and drought stress of field-grown sorghum and setaria have been achieved by leveraging a suite of imaging technologies. Stomatal patterning on the leaf epidermis has been phenotyped by optical tomography combined with an AI-enabled automated analysis pipeline. Leaf specific leaf area and nitrogen content have been phenotyped by leaf hyperspectral reflectance. Leaf rolling in response to drought and plant productivity has been phenotyped by hemispherical photography below the canopy. Canopy water use and the rate of stomatal closure has been phenotyped by thermal imaging at two different scales. In each case, (1) the fidelity of the rapid phenotyping method has been demonstrated against traditional gold-standard methods and (2) the resulting trait data has been applied to understanding the genetic architecture of the traits in setaria and /or sorghum. GWAS coupled with analysis of pan-genomic sequence

variation in putative candidate genes has revealed new sets of candidate genes and loci as targets for engineering enhanced water use efficiency and drought response in C4 bioenergy crops. It is also revealing how interactions among traits create opportunities and constraints for crop improvement.

The characterization of root system architecture has been an obsession for plant physiologists and developmental biologists. Recent work across several labs have established a wide range of different methods that bring with them strengths and weaknesses related to throughput, physiological relevance, resolution, and amenable species. The GLO-Roots method has been a successful entree to this portfolio due to the non-sterile soil-like media that is used to grow the root systems, the span of the plant life cycle that can be explored, and the ease of applying this approach to the model genetic systems such as Setaria, Brachypodium and Arabidopsis. Bottlenecks in the GLO-Roots imaging pipeline have been addressed recently, and as a consequence, we have been able to generate data of a magnitude that is several fold greater than was envisioned before. In LaRue and Lindner et al. we present the GLO-Bot I robotics platform, which enabled the first characterization of species-level differences in root system architecture in a soil-like growth environment beyond the seedling stage. We established a platform for automated plant root growth, imaging and image analysis that involved innovations in plant growth systems, robotics and computer vision approaches. This integrated solution to root phenotyping provides a stunning model for future studies and will influence the field of plant phenomics. We identify the major contributors to variation in the spatiotemporal control of root system architecture. New traits we define, such as average lateral root angle per day, show significant correlations to climate variables and suggest potential adaptive advantages for plants having shallow root systems when temperature variation is high in a location. Through continued funding from DOE-BER we have established the GLO-Bot II system, which is constructed with industry-standard robotics components and promises a new level of reliability and throughput compared to its predecessor.

**References/Publications**

1. LaRue T, Lindner H, Srinivas A, Exposito-Alonso M, Lobet G, Dinneny JR (2021) Uncovering natural variation in root system architecture and growth dynamics using a robotics-assisted phenomics platform. bioRxiv preprint doi: https://doi.org/10.1101/2021.11.13.468476

**Funding Statement:** This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC SC0018277).

**Improvement of Sorghum Transformation and Genome Editing for the Development of Stable Lines for the Analysis of Photosynthetic and Water Use Efficiencies**

Albert Kausch[1*](apkausch@uri.edu), James Chamness[2], Kimberly Nelson-Vasilchik[1], Joel Hague[1], Asaph Cousins[3], Robert DiMario[3], Dan Voytas[2], Matthew Zinselmeier[2], Colby Starker[2], Yang Liu[2], Andrew Leakey[4], and Ivan Baxter[5].

[1]University of Rhode Island, West Kingston; [2]University of Minnesota, Saint Paul; [3]Washington State University, Pullman; [4]University of Illinois, Urbana; [5]Donald Danforth Plant Science Center, Saint Louis, MO

www.foxmillet.org

**Project Goals:** This project aims to leverage *Setaria viridis* as a model system to develop novel technologies and methodologies to redesign the bioenergy feedstock *Sorghum bicolor* to enhance water use and photosynthetic efficiencies.

The development of a genome-level knowledge base linking genes to phenotypes in sorghum for bioenergy goals through the use of genome editing and stable plant transformation technology is critical to understanding fundamental physiological functions and important to crop improvement. We contribute the central hub capability to create, test and cultivate transgenic and genome edited plants with the various laboratories involved with this project. We have established reliable protocols for the *Agrobacterium*-mediated introduction of experimental genetic constructs into Sorghum cv BTx430, and collaborate to generate the viable transgenics required for the ongoing investigations on this project. These experiments include:; (1) sorghum RNAi constructs for knockdowns such as for voltage-gated chloride channel proteins, alpha carbonic anhydrase 7 (CA) and nine-cis-epoxycarotenoid dioxygenase 4, and myb domain protein 60; (2) constructs to test the fidelity of phosphoenol pyruvate carboxylase (PEPC) promoter expression, CA overexpression and PEPC with altered kinetics; (3) additional versions of CA overexpression aimed to test a range of increased mesophyll CA activity; (4) Ta Cas 9, dTa Cas9, and, dCas9 transcriptional activator for improved editing, and; (5) constructs to evaluate improvements to the transgenic process with the intent to increase transformation frequencies and shorten the time to T1 seed. These lines are currently in various stages of the transgenic process. The recent developments using morphogenic regulator-mediated transformation (MRMT) is a breakthrough toward enabling rapid transformation and genome editing. We report the development of an improved rapid transformation method using MMRT technology with the potential to increase through-put and decrease time for our projects. Working with the Voytas lab, we have evaluated a public version of the MRMT vectors. The Voytas lab is also testing novel methods for delivering genome editing reagents, specifically the use of RNA viral vectors to deliver gRNAs through infection. Heritable gene editing through infection has been achieved in several dicot species, and we are working to implement the technology in *Setaria* and sorghum.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018277.

**Title:** Leveraging multi-omics data and pan-genomic sequence variation in *Setaria viridis* and *Sorghum bicolor* for candidate gene discovery and nomination for crop improvement.

**Authors:** Phil Ozersky[1], Allen Hubbard[1], Todd Mockler[1], Asaph Cousins[2], Andrew D.B. Leakey[3], Seung Yon (Sue) Rhee[4], **Ivan Baxter[1]**

**Institutions:** [1]Donald Danforth Plant Science Center, St. Louis, MO; [2]Washington State University; [3]University of Illinois at Urbana-Champaign, Urbana, Illinois; [4]Carnegie Institution for Science.

**Project Goals:** This project aims to leverage *Setaria viridis* as a model system to develop novel technologies and methodologies to redesign the bioenergy feedstock *Sorghum bicolor* to enhance water use and photosynthetic efficiencies. Through the advancement of candidate gene discovery platforms and leveraging multi-omics datasets including analyses of pan-genomic sequence variation, we are identifying candidate genes for crop improvement efforts. These candidate genes are implicated in key traits of interest including water use and photosynthetic efficiencies and responses to drought stress.

**Abstract Text:** Bioenergy crop yields are likely to be limited by the availability of water during the growing season. Historically, efforts to improve water and photosynthetic use efficiencies and resilience to abiotic stresses such as drought have been limited by the challenge of efficiently identifying and harnessing genes that underly these traits. We have established a suite of genomic and transcriptomic datasets, tools, and approaches that enable elucidation of new insights into pathways and genes involved in water use and photosynthetic efficiencies and responses to drought stress. Integration of these datasets, tools, and approaches provides insights into associations between genotype, phenotype and environment that inform our efforts to improve bioenergy sorghum through breeding and engineering. We have leveraged whole-genome re-sequencing datasets, *de novo* reference grade genome assemblies, multi-omics profiling (e.g., transcriptomic, metabolomic, elemental, and isotopic ratios), pan-genomic sequence variation analyses, GWAS, and multi-model phenotyping datasets. We have combined these datasets and approaches to understand the genetic architecture of how plants respond to water deficit and to reveal and nominate new sets of candidate genes and loci as targets for engineering enhanced water use efficiency and drought response in C4 bioenergy crops.

**Funding Statement:** This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC SC0018277).

**Title:** Systems analysis of the beneficial associations of sorghum with arbuscular mycorrhizal fungi studied with genetics, genomics, imaging and microbiomics

**Authors:** Jeff Bennetzen[1] (maize@uga.edu), Jonathan Arnold[1], Anny Chung[1], Katrien Devos[1], and Nancy Johnson[2]

**Institutions:** [1]University of Georgia, Athens, GA; [2]Northern Arizona University, Flagstaff, AZ;

**Website URL:** https://mycorrhrsee.com

**Project Goals:** Arbuscular mycorrhizal fungi (AMF) are obligate symbionts/biotrophs that can enhance the disease resistance, nutrient access and/or drought tolerance of their plant host through their interactions with roots, but may also serve as negative contributors to plant yield depending on the agricultural environment. The primary goals of this project are to identify the plant genes that attract specific AMF interactions, to determine the possible benefits or costs of this interaction to the biofuel crop sorghum under varied field conditions, and to generate a systems analysis to assist plant breeders and agronomists to use endogenous AMF to improve sorghum productivity.

**Abstract Text:** In the first year of this project to determine the nature of AMF interactions and contributions to sorghum biomass production, we have developed the field and laboratory assays necessary for the full completion of this five-year project. Identical fields were planted in Georgia and Arizona that contained a small subset of the Genome Wide Association Study (GWAS) sorghum biomass panel that will be comprehensively analyzed in subsequent years. The root system microbiomes were analyzed over a time course to find the best (*i.e.*, least variable) time(s) to sample for robust results in subsequent seasons. These plants were also assayed for mineral content and aboveground biomass yield, and scored for disease incidence and plant developmental stage. Dramatic differences between the fields in Georgia and Arizona were noted, with very high levels of insects, weeds and pathogenic fungi in Georgia. Initial investigation of some of the observed diseases by our collaborator, Dr. Bochra Bahri at the University of Georgia, uncovered the organisms responsible for these diseases, including some that had not been reported previously as causing disease on sorghum in Georgia or, in some cases, anywhere in the world. DNA analyses of the root microbiome are ongoing, with precise values now established for the time, cost and personnel required for this science at this scale. High throughput imaging of AMF in these roots has also been advanced tremendously and is described in another poster from our project in this conference. Additional projects are underway to, first, define the best methods for RNA purification and analysis and, second, to test the effects of seed and foliage fungicide treatments on the root microbiome that is investigated in this project. Efforts to isolate and genome sequence specific AMF ecotypes are underway as well, in order to provide richer AMF sequence datasets to scan for homology to the RNA and DNA sequences we will find in the studied microbiomes. A linear model has been developed and fitted successfully to a subset of 79 BAP lines in the initial GWAS study as a first step in developing a structural equation model for the entire GWAS study.

.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021386.

•

**Title: Adaptive GWAS using MINE for discovery of structural equations in the AMF-Sorghum Project**

**Authors:** Shufang Zhang,[1] (shufan.zhang@uga.edu), Michael Skaro,[1] A. Bouffier-Landrum,[1] Isaac Torrres,[1] Yue Wu,[1] Yinping Guo,[1] Lauren Stupp,[1] Camryn Felt,[1] Brooke Lincoln,[1] Sedona Spann ,[2] Nancy Johnson,[2] and **Jonathan Arnold**

**Institutions:** [1]University of Georgia, Athens; [2]Northern Arizona University, Flagstaff, AZ;

**Website URL:** https://Mycorrhisee.com

**Project Goals:** Arbuscular Mycorrhizal Fungi or AMF are obligate root symbionts, associating with at least 80% of terrestrial plant families. The overarching project goal is to understand the AMF symbiosis in sorghum with a systems biology approach. Our particular components of the project are to identify a systems diagram for the network of AMF-sorghum interactions and imaging of AMF structures in the roots.

**Abstract Text:** Arbuscular Mycorrhizal Fungi (AMF) are obligate root symbionts, associating with at least 80% of terrestrial plant families and about which little is known. In return for plant assimilated carbon, these fungi provide nutrients, such as nitrogen and phosphorus, to their host plants. A systems biology approach is being used to understand this symbiosis in sorghum. To identify sorghum genotypes and genes that influence AMF abundance and plant biomass performance, a multi-year Genome Wide Association Study (GWAS) of AMF in sorghum on Wellbrook Farm in Watkinsville, GA is underway. Year 1 of the GWAS involves 79 accessions from the Bioenergy Accession Panel (BAP). The field was planted in early June, 2021 in 3 blocks with 79 rows and 9 seedlings of the same accession per row for a total of 2200 seedlings. 3 plants were harvested per row before and after flowering at three time points and measured for dry canopy weight and fresh and dry root weight. A high-throughput protocol for imaging symbiotic structures was developed using ink and vinegar staining. A total of 576 images of root intersections are being captured per plant for a total of 800 plants (88+288+424). Machine learning methods are being used to segment and classify AM fungal structures from the images for GWAS mapping to the sorghum genetic map. The results from GWAS, host biomass, microbiome, morphological, microscopic and transcriptome studies will feed into systems models, including structural equation models to predict sorghum genotypes that will maximize production (*i.e.*, plant health) knowing indigenous AMF in the field under different field conditions.

.

**Funding Statement** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. **DE-SC0021386**.

•

# Metabolic flux analysis of an engineered sucrose-secreting strain of the cyanobacterium *Synechococcus elongatus*

Bo Wang (bo.wang.2@vanderbilt.edu) [1*], Cristal Zuniga (crzuniga@ucsd.edu) [2*], Michael T. Guarnieri [3], **Michael J. Betenbaugh** [4], **Karsten Zengler** [2,5-6], **Jamey D. Young** [1, 7]

[1] Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN. [2] Department of Pediatrics, University of California San Diego, La Jolla, CA. [3] National Renewable Energy Laboratory, Golden, CO. [4] Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD.[5] Department of Bioengineering, University of California San Diego, La Jolla, CA. [6] Center for Microbiome Innovation, University of California San Diego, La Jolla, CA. [7] Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN.

**The goal of this project is to combine autotrophs and heterotrophs as a novel synergistic and symbiotic platform for the production of sustainable biofuel precursors. Photosynthetic microorganisms fix sunlight and $CO_2$ and provide organic carbon source and oxygen to the heterotrophs that are prolific producers of complex metabolites. Synthetic microbial communities of cyanobacterium-fungus have been studied through genome-scale metabolic modeling and $^{13}C$ metabolic flux analysis. Our current focus is to evaluate the metabolic response of the cyanobacterium *Synechococcus elongatus* PCC 7942 to osmotic stress and sucrose secretion.**

The capability of cyanobacteria to produce sucrose from $CO_2$ and light has significant societal and biotechnological impact since sucrose can serve as a carbon and energy source for a variety of heterotrophic organisms. Efforts on strain development and process optimization have taken place since a decade ago, and technology has been advanced significantly. However, most efforts have focused on understanding local pathway alterations that drive sucrose biosynthesis and secretion rather than global flux re-routing that occurs following induction of sucrose production by NaCl treatment. Therefore, we investigated global metabolic flux alterations in a sucrose-secreting ($cscB^+$) strain versus the wild-type *Synechococcus elongatus* 7942. We used $^{13}C$ metabolic flux analysis (MFA) and Genome-Scale Modeling (GSM) as complementary approaches to elucidate differences in intracellular resource allocation by quantifying metabolic fluxes between these strains. We performed $^{13}C$-MFA and GSM for three cyanobacterial cultures – wild-type *S. elongatus* 7942 grown in BG11 minimal culture medium (WT), wild-type *S. elongatus* 7942 grown in BG11 supplemented with additional 100 mM NaCl (WT/NaCl), and *S. elongatus* $cscB^+$ strain grown in BG11 medium supplemented with additional 100 mM NaCl ($cscB^+$/NaCl) – all under photoautotrophic conditions. We quantitatively described the dramatic rewiring of metabolic fluxes in WT/NaCl and $cscB^+$/NaCl relative to that of the WT culture, and identified a potential metabolic bottleneck limiting carbon fixation and sucrose biosynthesis in the engineered $cscB^+$ strain. Our study also demonstrates that combining two complementary approaches, namely $^{13}C$-MFA and GSM, is a useful strategy to both extend the coverage of MFA beyond central metabolism and to improve the accuracy of flux predictions using GSM.

*This project is supported by Biological and Environmental Research, Department of Energy, through grant DE-SC0019388: Creating Multifunctional Synthetic Lichen Platforms for Sustainable Biosynthesis of Biofuel Precursors.*

**Plant-Mycorrhizal-Decomposer Interactions and Their Impacts on Terrestrial Biogeochemistry**

Nahuel Policelli[1]* (npol@bu.edu), Colin Averill[2], Edward R Brzöstek[3], Hui-Ling Liao[4], Haihua Wang[4], Ryan Tappero[5], Joseph E Carrara[3], Corinne Vietorisz[1], Jake Nash[6], Rytas Vilgalys[6], and **Jennifer M Bhatnagar**[1]

[1]Boston University, Boston, MA; [2]ETH Zürich, Zürich, Switzerland; [3]West Virginia University, Morgantown, WV; [4]University of Florida, Quincy, FL; [5]Brookhaven National Laboratory, Upton, NY; and [6]Duke University, Durham, NC.

**Project Goals: We aim to determine the role of plant and soil resources in shaping interactions between coniferous plants, symbiotic ectomycorrhizal fungi, and free living saprotrophs, and the impact on plant and soil biogeochemistry.**

Ectomycorrhizal fungi (EMF) are the main tree roots symbionts in temperate and boreal forests and can mine soil for nitrogen (N), increase trees carbon (C) allocation belowground, and interact with free-living saprotrophic microbes (SAPs) that decompose soil C. We hypothesized that plant-EMF-SAP interactions and their controls over ecosystem biogeochemistry are context-dependent but predictable, based on resource (C and N) availability to both SAPs and EMF. In a greenhouse-based synthetic ecosystem experiment, we grew *Pinus taeda* seedlings with and without a highly specific EMF symbiont (*Suillus cothurnatus*), under high and low levels of soil C, soil N, and plant C (ambient vs. elevated carbon dioxide -$CO_2$). To assess soil microbial activity, we measured soil $CO_2$ release under the different conditions. We also measured plant biomass and EMF root colonization. We found that when more plant C was available, EMF suppressed soil $CO_2$ release under low soil C but had no effect under high soil C. By contrast, when less plant C was available, EMF promoted $CO_2$ release under high soil C conditions. Elevated soil N had a tendency to suppress the EMF effect on soil C-derived $CO_2$ release, although the result was not significant. Biomass was higher for those plants growing with EMF, regardless of soil C. Plants inoculated with EMF doubled the biomass of non-inoculated plants under high $CO_2$ conditions. EMF root colonization of inoculated plants was similar between treatments. Together, our results show that the direction of EMF-SAP interactions depends on soil C availability, potentially reversing according to plant C allocation belowground.

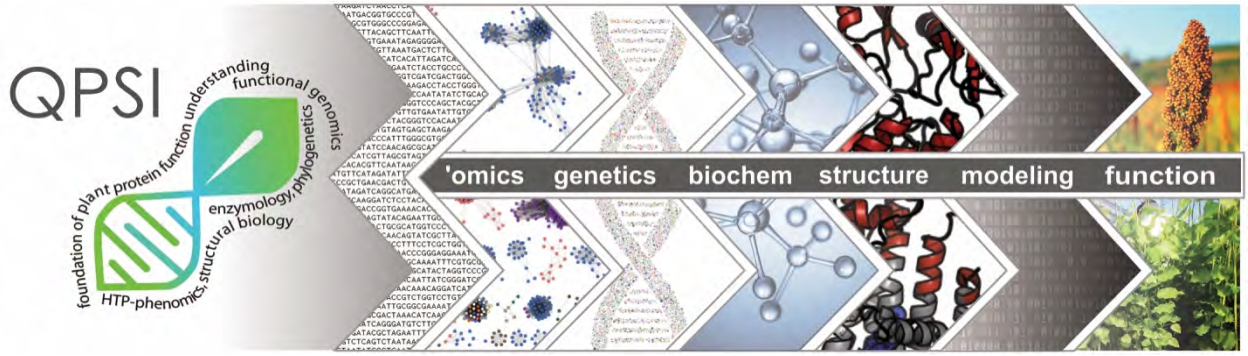
*Funding statement: The authors acknowledge DOE BER awards DE-SC0020403, AN_360581, and JGI/EMSL FICUS 51536.*

**Understanding poplar and sorghum micronutrient stress by integrating functional genomics with molecular-level experimentation**

**Crysten E. Blaby-Haas[1*] (cblaby@bnl.gov)**, Meng Xie[1], Timothy Paape[1] Qun Liu[1], Doreen Ware[4,5], Daifeng Wang[3], Sam Seaver[2], Lifang Zhang[4], Dimiru Tadesse[1], Michael Regulski[4], Miriam Pasquini[1], Changxu Pang[1], Sunita Kumari[4], Desigan Kumaran[1], Chiarg Gupta[3], Nicolas Grosjean[1], Nicolas Gladman[4], Aditi Bhat[1], Jeremiah Anderson[1]

[1]Brookhaven National Laboratory, Upton, NY; [2]Argonne National Laboratory, Argonne, IL; [3]University of Wisconsin, Madison, WI; [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [5]US Department of Agriculture, Agricultural Research Service, Ithaca, NY, USA

**Project Goals: The Quantitative Plant Science Initiative (QPSI) is a capability that aims to bridge the knowledge gap between genes and their functions. A central aspect of our strategy is combining genome-wide experimentation and comparative genomics with molecular-level experimentation. In this way, we leverage the scalability of 'omics data and bioinformatic approaches to capture system-level information, while generating sequence-specific understanding of gene and protein function. By incorporating molecular-level experimentation in our workflow, we are addressing the question of how a protein functions and establishing mechanistic insight into how sequence variation impacts phenotype. This knowledge serves as a touchstone for accurate genome-based computational propagation across sequenced genomes and forms the foundation for robust predictive modeling of plant productivity in diverse environments.**




Trace metal nutrients are of growing importance in maximizing bioenergy/bioproduction crop yield in marginal soil. Bioavailability in the soil is dynamic and variable, and yield-impacting deficiency can suddenly appear. Because these elements are essential for the proper assimilation and metabolism of macronutrients such as nitrogen, poor macronutrient availability can be exacerbated by metal deficiency. To support the development of bioenergy crops with improved micronutrient stress resilience, our goal is to develop a genome-based, molecular-level and system-level understanding for the two most abundant trace metal nutrients in plants: zinc and iron. Focusing on the bioenergy crops poplar and sorghum, we have completed a large-scale, hydroponics-based, transcriptomics timecourse to understand how these plants respond to and acclimate to different concentrations of metals in their environment. This data will be used to

construct gene-regulatory networks and a computational simulation of cofactor availability in the chloroplast, the major metal sink and site of carbon fixation and energy generation. We are also employing an interdisciplinary approach to provide a layer of experimentally grounded sequence-specific understanding of molecular-level functions for major players involved in metal homeostasis. Comparative genomics provides an in-silico platform to generate protein function hypotheses. Hypotheses are tested with reverse genetics in model organisms and biochemical assays of protein family members. Structure-function studies supply mechanistic insight into how sequence space translates into molecular function. These studies will improve our model of micronutrient dynamics in bioenergy crops and provide a sequence-based foundation for establishing how metal scarcity and excess affect plant health and biomass.

*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER).*

**A New Structural Paradigm in Heme Binding – a Novel Family of Plant Heme Oxidases.**

Nicolas Grosjean*†[1] (ngrosjean@bnl.gov), Desigan Kumaran†[1], Meng Xie[1], Ian Blaby[1,2] and **Crysten E. Blaby-Haas**[1]

[1]Biology Department, Brookhaven National Laboratory, Upton, NY
[2]Current affiliation: US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA.
†Contributed equally

**Project Goals: The Quantitative Plant Science Initiative (QPSI) is a capability that aims to bridge the knowledge gap between genes and their encoded function. A central aspect of our strategy is combining genome-wide experimentation and comparative genomics with molecular-level methodologies. Our program leverages the scale of 'omics data and bioinformatic approaches to capture system-level information, while generating sequence-specific understanding of gene and protein function. By incorporating molecular-level experimentation in our workflow, we are addressing the question of how a protein functions and establishing mechanistic insight into how sequence variation impacts phenotype. This knowledge serves as a touchstone for accurate genome-based computational propagation across sequenced genomes and forms the foundation for robust predictive modeling of plant productivity in nutrient-limited environments.**

https://genomicscience.energy.gov/research/sfas/bnlqpsi.shtml

Up to 90% of the iron found in leaves is in the chloroplasts, where every membrane-spanning photosynthetic complex has an absolute requirement for iron cofactors, such as heme. To ensure fidelity of cofactor synthesis and limit potential oxidative stress caused by free heme, specialized biogenesis pathways exist. A few pathways involved in heme trafficking and insertion were previously described for cytochrome *b*- and *c*-type hemoproteins. However, the existence of a generalized heme chaperone that can interact with the labile heme pool in photosynthetic organisms to protect and deliver heme has yet to be identified. We used a phylogenomic based approach to identify a large protein family consisting of uncharacterized and putative heme-binding proteins. Through this analysis we distinguished three distinct subfamilies of phototroph-specific homologs that we hypothesize have evolved discrete functions: i) heme-biosynthesis regulators, ii) heme oxidases, and iii) heme chaperones.

To test our computationally derived hypotheses, we purified protein homologs from green algae, cyanobacteria, and bioenergy feedstocks. We demonstrate that the algal and land plant proteins can bind and degrade heme *in vitro*, suggesting that these chloroplast-localized proteins represent a new family of plant heme oxidases. However, by combining *in vitro* and *in vivo* experiments in cyanobacteria, we hypothesize that the cyanobacteria-specific subfamily may function as either regulators or heme chaperones.

Crystal structures of the cyanobacterial homolog in the presence of heme revealed unprecedented features. The protein forms a dimer where the heme is saddled by two zinc ions. By analogy with the protective axial histidine ligands found in a bacterial heme transporter, we propose that the zinc saddle protects heme from oxidation. By homology modelling, this structure is aiding in understanding the structure-function relationship of plant homologs where the heme binding residues are not conserved. *In vivo*, mutation of key amino acids results in a similar phenotype as for gene deletion (knock-out), which supports their essential role and the relevance of this structure *in vivo*. To decipher biological function, we identified a specific interaction of this heme-binding protein, dependent on these key residues, with a subunit of the succinate dehydrogenase. Together with detailed biochemical and genetic analysis, our results suggest that the heme-binding protein may regulate electron transport. The discovery of this family of novel heme oxidases and putative heme regulators or chaperones provides new key molecular and genomic insights into the evolution of heme regulation in photosynthetic organisms.

*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER) as part of the Quantitative Plant Science Initiative (QPSI) SFA.*

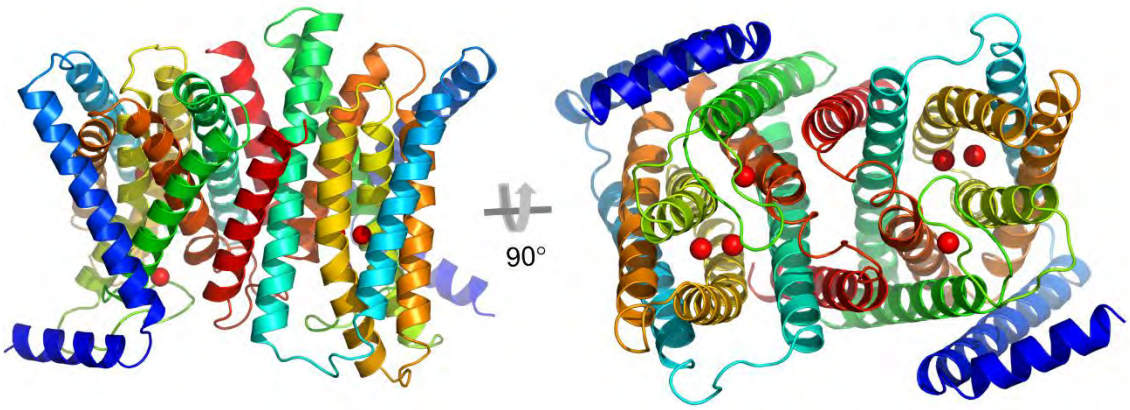
**Title:** Cryo-EM Structure of a Zinc Uptake Transporter in the Closed State

**Authors:** Changxu Pang,[1] Jin Chai,[1] Ping Zhu,[3] John Shanklin,[4] and **Qun Liu**[1]*
(qunliu@bnl.gov)

**Institutions:** [1]Biology department, Brookhaven National Laboratory, Upton, New York

**Project Goals: To improve zinc-stress resilience of bioenergy crops, an understanding of how they have evolved to deal with deficient or excess metal bioavailability is needed to support gene selection, breeding and biosystem design strategies. The goal of this project is to perform structural and functional characterization of zinc uptake transporters to understand the zinc transport process across membranes. The gained knowledge will be used to develop strategies that will promote growth of bioenergy crops in acclimation to zinc availability.**

**Abstract Text:** Zinc is an essential micronutrient and supports all living organisms through regulating biological processes such as gene regulation, structure stability, enzymatic metabolism, immune response, cell division and cell growth. Deficiency of zinc is detrimental to crop yield and bioproduct quality. In US, about sixty percent of farmland is zinc deficient. To understand the physical basis of the zinc uptake process, we have determined a cryo-EM structure of a zinc uptake transporter at 3.1 Å resolution in an inward facing closed conformation. The transporter forms a homodimer, each containing nine transmembrane helices and three zinc ions. Two zinc ions are in the middle of the transporter and form a binuclear pore structure, and the third zinc is located at the transporter's egress site facing cytoplasm. The egress site is covered up by a histidine-rich two-stranded β-sheet which blocks the release of the zinc into cytosol. Two histidine residues on the β-sheet interact to the egress-site zinc and regulate its release. Cell-based functional characterization of the wild-type protein and structure-inspired mutants provide further insights into the understanding of the zinc uptake process.




**Funding Statement:**
This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER).

**Time-series analysis of phenotypic and transcriptomic responses to nutrient stress in two bioenergy crops**

Timothy Paape[1] (tpaape@bnl.gov), Chirag Gupta[2], Daifeng Wang[2], Sunita Kumar[3,4], Doreen Ware[3,4], Dimiru Tedesse[1], Aditi Bhat[1], Jeremiah Anderson[1], Crysten E. Blaby-Haas[1], Meng Xie[1]

[1]Brookhaven National Laboratory, Upton, NY; [2]University of Wisconsin, Madison, WI; [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [4]US Department of Agriculture, Agricultural Research Service, Ithaca, NY,

https://genomicscience.energy.gov/research/sfas/bnlqpsi.shtml

**Project Goals:** The Quantitative Plant Science Initiative (QPSI) is a capability that aims to bridge the knowledge gap between genes and their functions. A central aspect of our strategy is combining genome-wide experimentation and comparative genomics with molecular-level experimentation. In this way, we leverage the scalability of 'omics data and bioinformatic approaches to capture system-level information, while generating sequence-specific understanding of gene and protein function. By incorporating molecular-level experimentation in our workflow, we are addressing the question of how a protein functions and establishing mechanistic insight into how sequence variation impacts phenotype. This knowledge serves as a touchstone for accurate genome-based computational propagation across sequenced genomes and forms the foundation for robust predictive modeling of plant productivity in diverse environments.

**Abstract**

Iron (Fe) and zinc (Zn) are indispensable cofactors for critical aspects of plant growth including numerous metabolic pathways, post-translational modifications, cell-wall biosynthesis, and carbon/nitrogen assimilation. Regulation of function makes a significant contribution to the phenotypic variation observed among different plant species and genotypes. Identification of co-expressed gene modules can provide functional inferences for proteins contained within a defined module, while conserved function of a protein between two or more species can be supported by conserved regulation. We developed hydroponic growth chambers at Brookhaven National Laboratory for time course Fe and Zn deficiency and excess experiments using the *Populus trichocarpa* Nisqually-1 and *Sorghum bicolor* cv. BTx623 genotypes. In both species, we found root and leaf level accumulation of Fe or Zn reflected the excess or deficiency treatments compared with controls. We also used fluorometer and reflectance spectroscopy to estimate leaf responses to Fe or Zn deficiency and found chlorophyll, photosynthesis, and nitrogen estimates were significantly affected by Fe and Zn stress. Time-series analysis of transcriptional responses (using RNA-seq) induced by Fe and Zn stress allowed us to examine both short-term and long-term differential gene expression over a three-week period (7 time points). We are conducting co-expression analysis to characterize the temporal dynamics of gene regulatory networks (GRN's) and we will quantify the extent to which the GRN's are conserved between these two physiologically divergent bioenergy crop species. These large-scale transcriptome datasets will provide a valuable resource for functional genomics studies to identify the roles of proteins involved in resilience to marginal soils and increased biomass production.

**Funding Statement**: This work was supported by the US Department of Energy, Office of Science, Office of Biological, and Environmental Research, as part of the Quantitative Plant Science Initiative at Brookhaven National Laboratory.

**Transcriptional regulatory mechanisms linking secondary cell wall biosynthesis and iron homeostasis in *Populus***

**Meng Xie**[1*](mxie@bnl.gov), Dimiru Tadesse[1], Chirag Gupta[2], Aditi Bhat[1], Sunita Kumari[3], Vivek Kumar[3], Jeremiah Anderson[1], Timothy Paape[1], Doreen Ware[3,4], Daifeng Wang[2], Crysten E. Blaby-Haas[1]

[1]Brookhaven National Laboratory, Upton, NY; [2]University of Wisconsin, Madison, WI; [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [4] US Department of Agriculture, Agricultural Research Service, Ithaca, NY, USA

**Project Goals: The Quantitative Plant Science Initiative (QPSI) is a capability that aims to bridge the knowledge gap between genes and their functions. A central aspect of our strategy is combining genome-wide experimentation and comparative genomics with molecular-level experimentation. In this way, we leverage the scalability of 'omics data and bioinformatic approaches to capture system-level information, while generating sequence-specific understanding of gene and protein function. By incorporating molecular-level experimentation in our workflow, we are addressing the question of how a protein functions and establishing mechanistic insight into how sequence variation impacts phenotype. This knowledge serves as a touchstone for accurate genome-based computational propagation across sequenced genomes and forms the foundation for robust predictive modeling of plant productivity in diverse environments.**

*Populus* is one of DOE's "flagship" plants and is of special interest as a lignocellulosic biomass feedstock. In *Populus*, the major source of lignocellulosic biomass is secondary cell walls, which are mainly composed of cellulose, hemicellulose, and lignin. Iron (Fe) is an essential micronutrient indispensable for photosynthesis and plant growth. Plants have evolved homeostatic mechanisms in Fe uptake, transportation, utilization, and storage to maintain optimal Fe concentration for normal growth. The availability of some nutrients (e.g., nitrogen) has been found to impact plant secondary cell wall biosynthesis. However, the link between secondary cell wall biosynthesis and Fe bioavailability remains poorly studied. Using the hydroponic growth system in the greenhouse, we discovered that Fe deficiency reduces lignin and secondary cell wall biosynthesis in *Populus* stems. Consistently, our transcriptomic study unveiled significant Fe deficiency-induced gene expression changes of transcription factors and biosynthetic genes involved in secondary cell wall formation. BASIC HELIX-LOOP-HELIX (bHLH) transcription factors play key roles in activating Fe homeostatic mechanisms under Fe deficiency in Arabidopsis. By performing chromatin immunoprecipitation-seq (ChIP-seq) analysis in *Populus* protoplast transient expression system, we found that the *Populus* orthologs of bHLH038 and bHLH121 target master regulators and biosynthetic genes of secondary cell walls, as well as genes involved in Fe homeostasis. It's interesting to observe that the two XYLEM NAC DOMAIN1 (XND1) orthologs in *Populus* are targets of bHLH038 and bHLH121. XND1 is a well-known transcription factor that negatively controls lignin and secondary cell wall biosynthesis. The two XND1 orthologs have one amino acid difference in the C-terminal protein-protein interaction domain and exhibited opposite expression patterns under Fe deficiency, suggesting potentially divergent functions. By screening their interacting proteins using the TurboID-based proximity labeling approach in *Populus*

protoplast transient expression system, we demonstrated that the two XND1 orthologs have different protein-interacting preferences. Both secondary cell wall biosynthesis and Fe homeostasis are tightly controlled by transcriptional regulatory networks. Our results suggest that their regulatory networks have crosstalk to coordinate these two essential biological processes.

*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER) as part of the Quantitative Plant Science Initiative (QPSI) SFA.*

**Characterization of a Novel Zinc Chaperone in Arabidopsis**

Lifang Zhang[1*] (zhangl@cshl.edu), Fangle Hu,[1] Audrey Fahey,[1] Miriam Pasquini[2], Meng Xie,[2] and **Crysten E. Blaby-Haas**[2], and **Doreen Ware**[3]
[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [2]Biology Department, Brookhaven National Laboratory, Upton, NY; [3]USDA-ARS;

**Project Goals: Short statement of goals. (Limit to 1000 characters)**
**The Quantitative Plant Science Initiative (QPSI) is a capability that aims to bridge the knowledge gap between genes and their functions. A central aspect of our strategy is combining genome-wide experimentation and comparative genomics with molecular-level experimentation. In this way, we leverage the scalability of 'omics data and bioinformatic approaches to capture system-level information, while generating sequence-specific understanding of gene and protein function. By incorporating molecular-level experimentation in our workflow, we are addressing the question of how a protein functions and establishing mechanistic insight into how sequence variation impacts phenotype. This knowledge serves as a touchstone for accurate genome-based computational propagation across sequenced genomes and forms the foundation for robust predictive modeling of plant productivity in diverse environments.**

Transition metals occupy numerous and often essential positions within the biochemical framework of plant metabolism. As protein cofactors, these elements have expanded the breadth of protein-catalyzed reactions and enabled pivotal energy-intensive reactions. To use metal ions as catalysts, the cell must balance a fundamental dichotomy: nutrient and toxin. Metal homeostasis has evolved to tightly modulate the availability of metals within the cell, avoiding cytotoxic interactions due to excess and protein inactivity due to deficiency. Even in the presence of homeostasis processes, however, low bioavailability of these essential metal nutrients in soils can negatively impact crop health and yield. While research has largely focused on how plants assimilate metals, acclimation to metal-limited environments, such as marginal soils, requires a suite of strategies that are not necessarily involved in metal transport. The identification of these assimilation-independent mechanisms provides an opportunity to improve metal-use efficiency and optimize feedstock yield in low nutrient soils without supplementing with expensive and environmentally damaging fertilizers.

Zinc is a vital micronutrient for plants. It is a constituent of approximately 2400 Arabidopsis proteins and an essential cofactor of enzymes and many regulatory proteins [1]. However, how zinc-dependent proteins are loaded with zinc was previously unknown. By leveraging phylogenomic and data-mining analyses combined with an inter-disciplinary experimental approach, we have discovered a novel metal chaperone that delivers zinc to an essential zinc-dependent enzyme during zinc deficiency. We provide evidence that this function is universally conserved from fungi to plants. In plants, we propose that duplication has resulted in analogous zinc-trafficking pathways in the cytosol and chloroplast. In Arabidopsis, there is one zinc chaperone (ZNG1) in the cytoplasm, and two paralogous chaperones (GMC2A1 and GMC2A2) localized to chloroplast. We have obtained and characterized corresponding mutants for ZNG1, GMC2A1, GMC2A2, MAPs (the target zinc-dependent proteins) and made several

crosses among these mutants. Here we present preliminary data of the characterization of plant ZNG1.Using Yeast two-hybrid (Y2H) and bimolecular fluorescence complementation (BiFc) studies, we show that ZNG1 can interact with the cytosolic target, methionine aminopeptidase, MAP1A, as we have found in yeast. Our hypothesis is that Plant ZNG1 is an activator of cytosolic MAP1, and cytosolic MAP2 can function as a back-up enzyme when cytosolic MAP1 activity is compromised.

**References**

1. Clemens S, 2019. Zn-a versatile player in plant cell biology. In: Cell Biology of Metals and Nutrients. Springer, Berlin, Heidelberg, pp.281-298.

*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER) as part of the Quantitative Plant Science Initiative (QPSI) SFA. Also: USDA-ARS 8062-21000-041-00D*

**The Combined Effect of Abiotic Stresses Reveals Unique Cell Type-Specific Molecular Changes in Poplar**

Vimal K. Balasubramanian [1] (vimalkumar.balasubramanian@pnnl.gov), Maria Del Mar Rubio Wilhelmi [2], Samuel Purvine [1], Allison Thompson [1], Ying Zhu [1], Sarah Williams [1], Hugh Mitchell [1], Ryan McClure [1], Stephen DiFazio [3], Jonathan Cummings [3], C. Neal Stewart, Jr.[4], **Eduardo Blumwald** [2], and Amir H. Ahkami [1]

[1] Environmental Molecular Sciences Laboratory (EMSL), Pacific Northwest National Laboratory (PNNL), Richland, WA; [2] Dept. of Plant Sciences, University of California, Davis, CA; [3]Department of Biology, West Virginia University, Morgantown, WV; [4] Dept. of Plant Sciences, University of Tennessee, Knoxville, TN.

**Project Goals: The main goal of the SyPro project is the development of transgenic trees with sustained photosynthetic activity and increased biomass production under the simultaneous occurrence of water deficit, increased soil salinity, and elevated temperatures. To achieve that, we intend to identify stress-responsive genes and proteins in specific cell types of poplar leaves and roots.**

Plant responses to abiotic stresses have been primarily studied at the whole plant and tissue levels[1], subjected mainly to a single stress condition. However, these studies provide insufficient information about specific cellular regulations[2] in response to multi-environmental stresses, a scenario that plants naturally cope with. Therefore, it is critical to identify the key stress-responsive regulators at cellular levels to better understand the molecular mechanisms of plant responses to individual and combined abiotic stresses.

In the SyPro project, clones of *Populus tremula x alba* (INRA 717 1-B4) were rooted for at least 25 days, grown in the greenhouse for 45 days, and the plant response(s) to individual stresses (salinity, heat, and water deficit) and a combination of two or three stresses were monitored. Leaf and root tissues were collected at different time points, fixed, and embedded for cell-type-specific omics analyses. We targeted distinct poplar cell types and tissues using cryo-sectioning and laser-capture microdissection (LCM) techniques.

Identifying stress-responsive regulatory proteins suffers from technical limitations associated with protein recovery from a low number of isolated cells and mass spectrometry sensitivity. In this work, we employed a novel microfluidic nanodroplet-based sample preparation (nanoPOTS)[3], coupled with laser capture microdissection and ultrasensitive liquid chromatography-mass spectrometry for highly sensitive proteome profiling of a small number of plant cells in a cell-type-specific fashion. We first demonstrated the power of this technology for plant cell-type-specific proteomics by identifying unique poplar leaf- (palisades and vascular) and root- (cortex and vascular) specific proteins in plants grown under normal condition[4]. We then proceeded with the stress-treated samples, in which we examined poplar leaf cell types exposed to individual and combined water deficit, salinity, and heat stresses. Out of 6,172 proteins significantly altered under single and triple stress conditions, 14.2% (palisade) and 33.3% (vascular) of identified proteins were unique to leaf cell types, while 52.4% were abundant in both palisade and vascular cells. Moreover, our biological pathway enrichment analysis revealed triple stress-specific proteins (compared to single stresses) for each cell type.

For example, we identified jasmonic acid biosynthesis and serotonin/melatonin biosynthesis as the most enriched pathways in palisade cells under triple stresses, while acetyl Co-A biosynthesis and chitin degradation pathways were the most active biological processes in leaf vascular cells.

Moreover, a low-input RNA sequencing approach was employed to resolve the plant spatial complexity under the single and simultaneous occurrence of abiotic stresses. Focusing on leaf tissue exposed to individual heat and salinity and a combination of both stresses (salt+heat), 56.3% (palisade) and 26.7% (vascular) of total differentially expressed genes (DEGs) (5899 genes) were regulated in a cell type-specific manner, while 16.9% of total DEGs were active in both cell types. Our functional enrichment analysis identified three categories specifically associated with palisade cells under salt+heat treatment (compared to individual stresses), including photosynthesis, L-methionine, and L-glutamine biosynthesis, while the most enriched biological processes for vascular cells were polysaccharide, glycan, and amino acid metabolism.

The integrative analysis of transcriptomic and proteomic data identified novel cell type-specific stress-responsive regulators highly induced at both transcriptional and translational levels. However, we observed a transcript/protein discordance in specific pathways in both leaf palisade and vascular cell types under heat and triple stress conditions representing a critical layer of cell-type-specific regulatory processes at the post-transcriptional level.

Overall, our findings effectively reveal the underlying molecular mechanisms regulating spatial plasticity under single and multiple stresses, allow future attempts in mapping molecular machinery to the cellular domain, and contribute to the design of poplar trees with enhanced tolerance to abiotic stress combination.

**References:**


1. Terri A Long, Many needles in a haystack: cell-type specific abiotic stress responses. (2011). *Current Opinion in Plant Biology*, 14, 3, pp. 325-33.
2. Dinneny, J.R., Long, T.A., Wang, J.Y., Jung, J.W., Mace, D., Pointer, S., Barron, C., Brady, S.M., Schiefelbein, J., Benfey, P.N. Cell identity mediates the response of Arabidopsis roots to abiotic stress. (2008). *Science*, 320 (5878), pp. 942-945.
3. Zhu, Y.; Piehowski, P. D.; Zhao, R.; Chen, J.; Shen, Y.; Moore, R. J.; Shukla, A. K.; Petyuk, V. A.; Campbell-Thompson, M.; Mathews, C. E. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. (2018). *Nature Communications,* 9, 882.
4. Balasubramanian, V. K., Purvine, S. O., Liang, Y., Kelly, R. T., Pasa-Tolic, L., Chrisler, W. B., Blumwald, E., Stewart, C. N., Zhu, Y., & Ahkami, A. H. (2021). Cell-type-specific proteomics analysis of a small number of plant cells by integrating laser capture microdissection with a nanodroplet sample processing platform. Current Protocols, 1, e153. doi: 10.1002/cpz1.153.



*The SyPro Poplar project is supported by the U. S. Department of Energy, Office of Biological and Environmental Research (BER), Genomic Science Program, Award # DE-SC0018347*

**Title: Transgenic poplar analysis of native and synthetic inducible promoters for sensing abiotic stress and tissue specificity from poplar *cis*-regulatory elements.**

**Authors:** Yongil Yang[1,2] *(yiyang98@utk.edu), Yuanhua Shao[1,2], Timothy Alexander Chaffin[1,2], Jun Hyung Lee[1,2], Magen R. Poindexter[1,2], Vimal Kumar Balasubramanian[5], Amir H. Ahkami[5], Stephen DiFazio[6], Jonathan Cummings[6], **Eduardo Blumwald**[7], and C. Neal Stewart, Jr.[1,2]

**Institutions:** [1]Center for Agricultural Synthetic Biology, University of Tennessee Institute of Agriculture, Knoxville, TN; [2]Department of Plant Sciences, University of Tennessee, Knoxville, TN; [3]Department of Horticultural Science, North Carolina State University, Raleigh, NC; [4]Department of Food Science, University of Tennessee, Knoxville, TN; [5]Environmental Molecular Sciences Laboratory (EMSL), Pacific Northwest National Laboratory (PNNL), Richland, WA; [6]Department of Biology, West Virginia University, Morgantown, WV; [7]Department of Plant Sciences, University of California, Davis, CA

**Project Goals: Sypro Poplar: Improving poplar biomass production under abiotic stress conditions: an integrated omics, bioinformatics, synthetic biology and genetic engineering approach.**

The project goal is to attain robust biomass of trees under abiotic environmental stress conditions via genetic engineering. Abiotic stress-resistance genes, especially those for water deficit, salt, and temperature stress, should optimally be under inducible regulatory control. We are designing, constructing, and testing stress-inducible synthetic plant promoters to drive resistance genes. Omics data are used to discover *cis*-regulatory DNA motifs used to construct synthetic promoters. The synthetic promoters are then tested under appropriate stimuli in engineered plants to use toward the development of environmentally resilient poplar.

**Abstract text:** Abiotic stress can cause significant damage to plants, which is crucial to avoid if production of bioenergy tree crops is to be sustainable. Promoter engineering has been proposed as a promising solution to overcome the challenges caused by

various abiotic stressors: they can be useful to control the expression of stress resistance transgenes. Through previous studies, we synthesized 9 promoters (SD9-1, 9-2, 9-3, 13-1, 18-1, 18-3, SS16-1, 16-2, and 16-3) from poplar transcriptome data using a *de novo* DNA-motif-detecting algorithm. Their stress inducibility was identified in water-deficit and salt treatment assays using poplar leaf mesophyll protoplast transformation and agroinfiltration of *Nicotiana benthamiana* leaves. In this study, these 9 synthetic promoters were stably transformed into *Populus tremula* × *Populus alba* hybrid poplar, by which GFP inducibility was screened under osmotic stress conditions. Of 9 transgenic poplar lines each harboring a different synthetic promoter, SD18-1, SD9-2, SD9-3, and SS16-1, there was significantly induced GFP expression in both salt and PEG treatments.

Tissue-specific native and synthetic promoters were also designed from RNA-seq datasets from laser micro-dissected poplar leaf tissue after water-deficit or salt stress treatment. Three native promoters from water-deficit transcriptome data were separately fused with a GUS reporter gene and used to generate transgenic poplar. All three native promoters induced specific GUS expression in early developmental leaf stages of transgenic poplar. Specifically, the native promoter Potri.003G072800 was induced in leaf blades, but not in vascular tissue. Synthetic promoters were also designed for leaf-tissue-specific induction by constructing heptameric repeats of 10 bases from conserved sequence of native promoters. Additionally, the tissue-specific inducibility of synthetic promoters fused with a GFP reporter was screened in leaf mesophyll protoplasts, and two GFP inducible synthetic promoters were identified for leaf tissue expression.

The tested promoters will be utilized to develop stress-responsive and tissue-specifically-controlled bioenergy tree crops in the future.

**Funding statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018347.

**Climate history dictates microbial metabolic response to drought stress: from semi-arid soils to tropical forest precipitation gradients**

Nick J Bouskill[1], Ulas Karaoz[1], Stephany Chacon[1], Aizah Khurram[1], Lee Dietterich[2], Hoi-Ying Holman[3], Daniela F Cusack[2,4]

[1]Climate and Ecosystems Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA.
[2]Department of Ecosystem Sciences and Sustainability, Colorado State University, Fort Collins, CO.
[3]Berkeley Synchrotron Infrared Structural BioImaging Program, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720.
[4]Smithsonian Tropical Research Institute, Apartado, 0843-03092, Balboa, Ancon, Panama.

The frequency and intensity of environmental fluctuations play an important role in shaping microbial community composition, trait-distribution, and adaptive capacity. We hypothesize here that a communities' climate history dictates it's metabolic response to future perturbation under a changing climate. Such a response is significant as changes in microbial metabolism can, in turn, feedback onto metabolite exudation, the chemical structure of necromass, and the formation and stability of soil organic matter. Here we use laboratory and field experiments to examine the metabolic pathways invoked under osmotic and matric stress within semi-arid and tropical soils. For example, using non-destructive, synchrotron-based Fourier-transform infrared spectromicroscopy we profiled the stress response of phylogenetically similar bacteria isolated from soils with contrasting climate histories subjected to both matric and osmotic stress. We note a strong carbohydrate-based, metabolic response of tropical microbes that is entirely absent in semi-arid organisms. At the field scale, we use metagenomic sequencing and metabolite analysis to demonstrate how four different sites established across a 1 m precipitation gradient from the Caribbean coast to the interior of Panama respond to a 50 % reduction in throughfall. The precipitation gradient permits the development of distinct communities at each site that show clearly divergent response to imposed hydrological perturbation. Our contribution here will discuss how communities adapted to different precipitation regimes respond metabolically to drought conditions, and how these change feedback onto the structure and stability of soil organic matter.

**Title:** Agent-Based Modeling of Algae Reveals Impact of Self-Shading on Metabolism

**Authors:** Alex Metcalf[1] and **Nanette Boyle**[1*] (nboyle@mines.edu)

**Institutions:** [1]Colorado School of Mines, Golden CO

**Project Goals:**

The overall research objective is to develop an experimentally validated multi-paradigm multi-scale modeling framework that will enable the most advanced and predictive metabolic modeling of diurnally grown photosynthetic organisms to date. The genome-scale metabolic model of C. zofingiensis will be embedded into an agent-based modeling framework to allow modeling of diurnal growth; the model will also be able to simulate intracellular fluxes, cell-to-cell interactions, cell-to-environment interactions, metabolite diffusion, and spatial distribution. This modeling approach will allow us to simulate metabolic shifts that occur due to diel cycles and generate rational engineering strategies to design production strains that are not impacted negatively by this natural phenomenon.

**Abstract Text:**

Economical algae production requires growth under outdoor light, but the diel nature of sunlight complicates modeling efforts. This poster documents a solution for that: a fully functional three-dimensional agent-based model, capable of simulating algal growth under diurnal conditions. By rendering light rays on a sub-micron scale, this model captures the self-shading effects of cell density and culture depth. It also includes diffusion, gas transport, and the carbonate equilibrium, making it straightforward to see the impacts of increasing gas-phase carbon dioxide concentrations. The agents themselves use diurnal transcriptomic data to dynamically adjust their individual metabolic constraints, an approach which successfully predicts phenotype from genotype. A typical simulation involves thousands of individual agents, each of which tracks its own metabolism over the length of the simulation. The combination of these advances results in a model capable of simulating physically relevant culture densities of cells, embedded in a water column of physically relevant length and under economically relevant light regimes.

One significant finding of this research is that commonly used culture densities are extremely effective at absorbing light over surprisingly short path lengths. Even under full sunlight, algal ponds need to strike a balance between culture density and functional volume; attempting to increase productivity of a given pond by making it deeper will only produce benefits at very low cell densities. Given that a deeper pond also requires more energy to aerate and mix, it is likely that very shallow ponds can produce algae more effectively.

This phenomena occurs because of the self-shading aspect of photosynthetic growth. Because total light energy scales off area, any given frontal area has a maximum quantity of energy associated with it. The cells nearest the surface receive the largest share of the light, and the energetic availability decreases with depth, as the remaining light is progressively absorbed by cells further down. At a critical depth, the average cell will encounter only a small fraction of the original light – sufficient to keep itself alive, but *not* enough to grow larger. Increasing the depth below that point is therefore counter-productive; the cells occupying that additional volume don't contribute any additional biomass growth to the culture, and indeed may actually *decrease* it.

It is worth noting that this critical depth is a function of both algal density and total light. The more dense the culture is, the shallower the depth will be, as the increased quantity of cells will absorb more light. However, the stronger the light is, the deeper the critical depth will be; more energy will be available at every depth, even if the fractions of total energy remain the same. By rendering light rays and simulating their intersections with individual cells, this model is able to identify the energetic availability as a function of depth for *any* given culture density and light condition.

Since it is built off a metabolic model, this simulation can also be used to model the impacts of all features normally associated with constraint based metabolic models, but on a physically relevant scale. For instance, it can project the total nutrient needs of an algal pond, by summing uptake amounts across all cells – or, it could be used to compare productivity differences between reactor configurations. Adjusting the simulation parameters allows for changes in density, depth, light amount, light composition, and genotype. Finally, the model is – if the data is available – species agnostic; any constraint based model of a similar algal species with the same set of species-specific data can be used.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0019171.

**Title: RCSB Protein Data Bank: Connecting genes to structures to ecosystems**

**Authors:** John D. Westbrook[1], Jasmine Young,[1] Robert Lowe,[1] Christine Zardecki,[1] Jose Duarte,[2] and **Stephen K. Burley***[1,2] (Stephen.Burley@rcsb.org)

**Institutions:** [1]Rutgers, The State University of New Jersey, Piscataway, NJ; [2]University of California San Diego, La Jolla, CA

**Website URL:** http://www.rcsb.org

**Project Goals:** The Vision of the RCSB PDB is to enable open access to the accumulating knowledge of 3D structure, function, and evolution of biological macromolecules, expanding the frontiers of fundamental biology, biomedicine, and biotechnology.

**Abstract**

Protein Data Bank (PDB) was established as the 1st open access digital data resource in biology and medicine in 1971. Today, the PDB is one of two global resources for experimental data central to biological science as a public good (the other key Primary Data Archive being the International Nucleotide Sequence Database Collaboration). PDB currently houses >185,000 atomic level biomolecular structures determined by macromolecule crystallography, NMR spectroscopy, and 3D electron microscopy. It is managed by the Worldwide Protein Data Bank partnership (wwPDB; wwpdb.org) according to the FAIR principles (*i.e.,* Findability, Accessibility, Interoperability, and Reusability).

Through an internet information portal and downloadable data archive, researchers and educators can access 3D structure data for large biological molecules, such as proteins, DNA, and RNA. These are the molecules of life, found in all organisms on the planet. Knowing the 3D structure or shape of a biological macromolecule is essential for understanding the role the molecule plays in health and disease of humans, animals, and plants, food and energy production, and other topics of concern to global prosperity and sustainability.

The RCSB PDB (rcsb.org) operates the US data center for the wwPDB, serves as Archive Keeper for the global PDB Core Archive, and makes PDB data available at no charge to all data consumers worldwide with no limitations on usage. Studies of website usage, bibliometrics, and economics demonstrate the powerful impact of the PDB data on basic and applied research, clinical medicine, education, and the US economy.

During calendar 2021, >1,320 million structure data files were downloaded from the RCSB PDB by Data Consumers working worldwide. During this same period, the RCSB PDB processed >5,680 new atomic level biomolecular structures plus experimental data and metadata coming into the archive from Data Depositors working in the Americas and Oceania. In addition, RCSB PDB served many millions RCSB.org users worldwide with PDB data integrated with >50 external data

resources providing rich structural views of fundamental biology, biomedicine, and energy sciences, and supported >690,000 PDB101.rcsb.org educational website users around the globe.

RCSB PDB provides a rich collection application programmable interfaces (APIs) that enable searching of and access to PDB archival data content. These APIs enable search by key citation, biological and structural features, and retrieval of individual PDB structure entries, reports, and chemical and molecular reference data. Building on this functionality in 2022, RCSB PDB will deploy new API services that enable parallel delivery of both experimental structures and computed structure models (coming from AlphaFold2, RoseTTAFold, ModelArchive) fully integrated with sequence information and functional annotations spanning all features of the RCSB.org research focused web portal. Considerable emphasis will be placed on delivering computed structure models across entire proteomes of interest to the US Department of Energy. 3D structure data (both experimental from the PDB and computed structure models from AlphaFold2, RoseTTAFold, ModelArchive) will also be made available for delivery through the KBase portal (KBase.us).

Access to PDB data and services contribute to patent applications, drug discovery and development, publication of scientific studies, innovations that can lead to new product development and company formation, and STEM education.

**References/Publications**


1. Burley et al. (2021) RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering, and energy sciences. *Nucleic Acids Research 49*, D437-D451. DOI: 10.1093/nar/gkaa1038
2. Rose et al. (2021) RCSB Protein Data Bank: Architectural advances towards integrated searching and efficient access to macromolecular structure data from the PDB archive. *Journal of Molecular Biology 433*, 166704. DOI: 10.1016/j.jmb.2020.11.003
3. Burley et al. (2021) Proteome-scale Protein Structure Prediction with Artificial Intelligence. *New England Journal of Medicine 385*, 2191-2194. DOI: 10.1056/NEJMcibr2113027
4. Goodsell and Burley (2021) RCSB Protein Data Bank Resources for Structure-facilitated Design of mRNA Vaccines for Existing and Emerging Viral Pathogens. Structure. DOI: 10.1016/j.str.2021.10.008.



**Funding Statement:** *RCSB PDB is funded grants to SKB from the National Science Foundation (DBI-1832184), the US Department of Energy (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM133198. Additional funding has been provided by a KBase subcontract from Argonne National Laboratory (P.I.: Chris Henry).*

**Title:** Spatiotemporal Mapping of Perturbations in a Naturally Evolved Fungal Garden Microbial Consortium

**Authors:** Ruonan Wu[1], Margaret W Thairu[2], Yuqian Gao[1], Jennifer E Kyle[1], Marija Velickovic[1], Nathalie Munoz[1], Kirsten Gotting[2], Carrie D Nicora[1], Matthew E Monroe[1], Ronald J Moore[1], Paul Piehowski[1], Cameron R Currie[2], Bobbie-Jo M Webb-Robertson[1] and **Kristin E Burnum-Johnson** (Kristin.Burnum-Johnson@pnnl.gov)[1]*

**Institutions:** [1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA; [2]University of Wisconsin-Madison, Madison, WI

**Project Goals:** This early career research project is dedicated to achieving transformative molecular-level insights into microbial lignocellulose deconstruction through the comprehensive and informative view of underlying biological pathways provided by the integration of spatiotemporal multi-omic measurements (i.e., proteomics, metabolomics, and lipidomics). A focus of this project is to uncover the mechanisms that drive cooperative fungal-bacterial interactions that result in the degradation of lignocellulosic plant material in the leafcutter ant fungal garden ecosystem. This approach will provide the knowledge needed for a predictive systems-level understanding of the fungal-bacterial metabolic and signaling interactions that occur during cellulose deconstruction in an efficient, natural ecosystem.

**Abstract Text:** Predicting the impact of perturbation on microbial communities relies on a mechanistic understanding of how the community members interact with each other and with the environment. Biological samples, however, are often complex and heterogeneous; thus, it is challenging to detect the variation in community composition and activities across spatial and temporal space. Here, we used a naturally evolved leafcutter ant fungal garden consortium carrying out lignocellulose degradation [1] that was gradually infected by a pathogenic fungus, *Escovopsis*, as a dynamic system to study. Across the vertical gradient of infection, we applied multi-omics (metaproteomics, lipidomics, metabolomics) with an integration of microscale imaging to capture and map shifts in microbial community members and detected activities.

We first curated a reference database from 50 million proteins of known members in the consortium, which were grouped into >24 million clusters based on sequence similarity, to annotate the high-resolution tandem mass spectrometry (MS) spectra with stringent matching criteria. These measurements leveraged a Thermo Fisher Orbitrap Eclipse Tribrid mass spectrometer equipped with a front-end High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) interface to enhance peptide selectivity and sensitivity. A total of 40,119 clusters were detected in the metaproteome data, which unveiled a complex community with relatively high representations of fungal, bacterial, and plant proteins. This was followed by archaeal, and insect proteins. The fungal cultivar, *Leucoagaricus*, has naturally co-evolved with the garden consortium to play a dominant role in breaking down lignocellulose. As the pathogenic fungus, *Escovopsis*, proliferated, the normalized spectral counts of *Leucoagaricus* proteins decreased. Additionally, that of bacterial members were generally increased, suggesting an active community response and potential interkingdom interactions under the impact of fungal infection. Among the most dominant protein clusters, a cluster of *Escovopsis*-specific fatty acid synthase was detected and increased with the infection. A principal component analysis of the

global lipidomic data further indicated that an *Escovopsis* infection could shape the lipid profile of the consortium. Across lipid subclasses, including phosphatic acids (PA), diacylglycerophosphocholines (PC), diacylglycerophosphoethanolamines (PE), and diacylglycerophosphoglycerol (PG), significant trends (adjusted p-value < 0.05) were observed due to infection based on individual fatty acid composition. Plant alpha-linolenic acid (18:3) containing lipids, which we previously found to be molecular indicators of lignocellulose degradation [1], were associated with healthy fungal garden consortium. Conversely, oleic acid (18:1) containing lipids (i.e., PA(18:1/18:1), PC(18:1/18:1), PE(18:1/18:1)) and bacterial odd-chain fatty acid containing lipids (i.e., PC(19:1/19:1), PG(19:1/19:1)) increased with infection. Our metabolomic data captured the depletion of myo-inositol, an essential nutrient for fungi, in regions where *Escovopsis* was highly proliferative [2]. Phosphoric acid with antimicrobial activity against bacteria and fungi was one of the most abundant metabolites detected and gradually increased with the fungal infection. This indicated potential interactions among bacteria and healthy and pathogenic fungi. Leveraging our capabilities in microscale imaging and spatial multi-omics (metabolome and metaproteome), we delineated the spatial distribution of the community members and further mapped the detected activities to those that were co-localized within micron-scale regions.

These spatial temporal multi-omics measurements provided an integrated road map to efficiently harness microbiome data for a better understanding of microbial interactions and community response to a perturbation. In addition, it provided a predictive systems-level understanding of how symbiotic fungal-bacterial metabolic and signaling interactions enable the leafcutter ant fungal garden ecosystem to thrive and degrade lignocellulose in dynamic environments.

**References/Publications**


1. Khadempour, L., et al., *From plants to ants: Fungal modification of leaf lipids for nutrition and communication in the leaf-cutter ant fungal garden ecosystem.* Msystems, 2021. **6**(2): p. e01307-20.
2. Reynolds, T.B., *Strategies for acquiring the phospholipid metabolite inositol in pathogenic bacteria, fungi and protozoa: making it and taking it.* Microbiology, 2009. **155**(Pt 5): p. 1386.



**Funding Statement:** The Pacific Northwest National Laboratory is a multi-program national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RLO 1830. This program is supported by the DOE, Office of Biological and Environmental Research (BER), under the Early Career Award Program. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by BER and located at PNNL.

**Direct measure of amino acid distribution across *Populus trichocarpa* roots in a rhizosphere-on-a-chip habitat**

John F. Cahill[1*] (cahilljf@ornl.gov), Courtney Walton[1], Muneeba Khalid[1], Vilmos Kertesz[1], Scott T. Retterer[2], and Jennifer Morrell-Falvey[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge TN; [2]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge TN

**Project Goals: The goal of this bioimaging project is to develop new technologies to image changes in chemistry occurring in the rhizosphere in living biosystems. A major hurdle in existing bioimaging technologies is the lack of broad chemical coverage and specificity in enormously complex chemical environments like the rhizosphere. To tackle this challenge, we developed a non-destructive liquid extraction-based mass spectrometric method that enables *in situ* measure of exogeneous chemical components in a rhizosphere-on-a-chip habitat populated with a *Populus* cutting. We demonstrate empirical measure of the distribution and heterogeneity of several amino acids along the whole of the *Populus* root structure within this environment.**

The rhizosphere is an incredibly complex environment, containing thousands of unique exogenous chemical species oriented in a complex spatial network. Such compounds are known to affect plant-microbe organization, interactions and, ultimately, growth and survivability. Due to its importance, the role of exogenous compounds in the rhizosphere is under much investigation, specifically the relation between plant physiology and the spatiotemporal distribution of molecular components. However, measure of the spatial distribution of exogenous compounds in the rhizosphere is challenging given the complex and dynamic nature of the environment. Compounds include, among others, organic acids, polysaccharides, proteins, and amino acids (AAs) which can exhibit a variety of roles in the rhizosphere including acting as a nutrient source for microbial colonization or a deterrent against pathogenic species. The exudation of AAs is one of the biggest components of plant carbon loss, which when released into carbon-deficient soil can lead to significantly enhanced hot-spots of microbial growth. Therefore, methods capable of directly measuring exudates, including AAs, are necessitated to understand rhizospheric phenomena.

Direct measurement of AA distribution within the rhizosphere is challenging due to the complex and dynamic nature of the environment and the limited accessibility of the rhizosphere for analysis. Nearly every AA can be found within the rhizosphere in a complex chemical cocktail of salts, proteins, and other compounds which present difficulties for chemical differentiation though existing imaging modalities. In addition, most biological pathways are spatiotemporally regulated, which suggests the distribution of exudates and AAs will vary in space and time. However, even for abundant molecules, like AAs, little is known of how they are spatially distributed along plant roots, how their distribution changes over time and how their distribution affects microbial composition and, ultimately, plant health.

Here we show the application of liquid microjuction-surface sampling probe mass spectrometry (LMJ-SSP-MS) for the direct measure of AA distribution within a rhizosphere-on-a-chip habitat populated with a *Populus trichocarpa* cutting. Uniquely, this newly developed imaging technology enables mass

spectrometric imaging of over a dozen AAs across the root structure *in situ*, without destructively impacting *Populus* root growth and development. Our data shows the spatial relationship of AA relative to root structure and finds that AAs exhibit significant heterogenous spatial and compositional structure across the root system. Additionally, differences in AA composition across the root imply heterogenous release of AAs based on their identity, whether by direct exudation or other processes (e.g., cell lysis and sloughing).

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This project was sponsored by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Bioimaging Program under FWP ERKP920.*

**Principles of Fungal Metabolism, Growth and Bacterial Interactions.**

Kevin Tsai[1,3], Jolene Britton[1,3], Alireza Ramezani[2,3], Mark Alber[1,2], Dale Pelletier[4] and **William Cannon***[1,3,5]

[1]Department of Mathematics, [2]Department of Physics, [3]Interdisciplinary Center for Quantitative Modeling in Biology, University of California, Riverside, CA 92521
[4]Oak Ridge National Laboratory, Oak Ridge, TN, 37830
[5]Pacific Northwest National Laboratory, Richland, WA, 99352;

**Project Goals:** The goals of this project is to elucidate fundamental principles of species interactions using hybrid machine learning/simulation models of *fungal-bacterial interactions and dynamics*. These hybrid data-analytic/simulation models are being used to carry out virtual experiments and develop fundamental understanding of the interactions between fungi and bacteria, specifically the mycorrhizal fungus *Laccaria bicolor* and the helper bacteria closely related to *Pseudomonas fluorescens.* At the same time, we carry out experiments aimed at developing and testing quantitative assays to measure the same interactions, and whose data will inform the our view of biology. We are:

- Evaluating the impacts of (1) thiamine and phenazines and (2) trehalose, produced respectively by *P. fluorescens* and *L. bicolor* on the metabolisms of each other. Metabolic exchange is an emerging theme in bacterial-fungal and bacterial-bacterial interactions.
- Characterizing *L. bicolor* -stimulated chemotaxis of pseudomonas by coupling trehalose signaling and metabolism to chemotaxis of pseudomonads.
- Experimentally investigating (1) pseudomonas chemotaxis in response to *L. bicolor* produced metabolites, and growth of *L. bicolor* in response to metabolites secreted by pseudomonads.

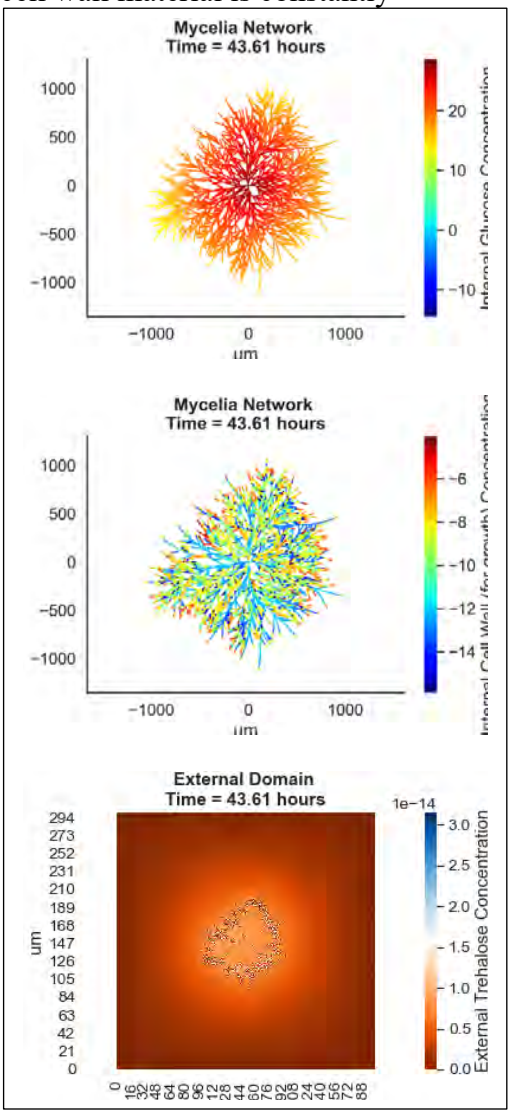
**Abstract:** The exchange of metabolites between microbes is an emergent property that evolves because the exchanged metabolites allow for increased growth of both species by reducing the thermodynamic cost of growth. Instead of each species producing every metabolite needed, metabolite exchange allows each microbe to specialize and efficiently produce a metabolite, such as trehalose, in exchange for one that it cannot produce as cheaply, such as thiamine. In economics this is known as Ricardo's principle of comparative advantage [1]. In order to evaluate the benefits of such microbial trade, physics-based models are needed that are capable of modeling the thermodynamic costs and benefits. The long-term goal is to understand fundamental principles of fungal-bacterial interactions through physics-based models of metabolism, protein expression and gene expression and to couple these models to the mycelial growth and bacterial chemotaxis.

The parameters for the ordinary differential equations (ODEs) needed to model the dynamics of metabolism are obtained by exploiting the natural selection principle that organisms that have the highest entropy production rates as a group will outcompete species with lower entropy production rates. Because we are using detailed models, entropy production includes growth, maintenance and catabolism. Another way of stating the entropy production principle is that the organisms that grow the fastest and most efficient will out compete others with slower and less efficient metabolisms. This perspective of metabolism subsumes many ecological concepts such as the red queen hypothesis and the black queen hypothesis. The maximum entropy production principle has a form that can be explicitly derived at the scale of metabolism.

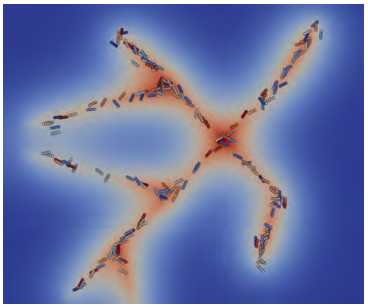
Our *Laccaria bicolor* metabolic model currently includes all reactions of central metabolism, synthesis of all 20 proteogenic amino-acids, synthesis of cofactors, synthesis of ATP, GTP, UTP, CTP and TTP and all deoxy-nucleic acid forms, synthesis of generic proteins, synthesis of fatty acids, sterols, synthesis of generic DNA and RNA strands and growth on minimal media of

glucose, ammonia, sulfate, and phosphate. Regulation of the metabolic activity is carried out by a method that combines reinforcement learning of control with statistical thermodynamics and metabolic control analysis, a branch of control theory [2].

A picture is emerging of natural oscillations within metabolism that are related to the cell cycle and circadian rhythms. In this picture, the NADPH/NADP ratio oscillates between high and low values, driving alternately DNA and fatty acid synthesis (high NADPH/NADP) and protein and RNA synthesis (low NADPH/NADP). Throughout the cycle, cell wall material is constantly produced, leading to an ever growing mycelium.




Concurrently, we are developing structural models of the fungal mycelial growth. The mycelial models (right) take in glucose (top right), convert the glucose to cell wall precursors which are actively transported through the fungal hyphae (middle right), and produce chemo-attractants which are exported and diffuse away in the external environment (bottom right). The internal metabolism of each hyphal segment currently uses a system of coupled Michealis-Menten systems along with equations for diffusive and active transport across hyphal septa. In the coming year, we expect to replace the Michaelis-Menten models with full metabolic models discussed above, allowing us to study the details of the thermodynamic benefits of metabolic exchanges with soil bacteria.



Also concurrently, we are developing models of chemotactic Pseudomonas species that interact with the fungal mycelia described above (left). The bacteria consist of subcellular element models to describe the structure. The bacteria run, flick and then reverse directions in order to navigate toward nutrition sources. The frequency of reversing the direction of motion is controlled by an internal clock. The bacteria have difficulty moving in solid media such as agar (or dehydrated soil) but the water excreted by the fungi due to metabolic activity provides a highway for the bacteria.

**References:**


[1] J. Tasoff, M. T. Mee, and H. H. Wang, "An Economic Framework of Microbial Trade," *PLoS One,* vol. 10, no. 7, p. e0132907, 2015, doi: 10.1371/journal.pone.0132907.

[2] S. Britton, M. Alber, and W. R. Cannon, "Enzyme activities predicted by metabolite concentrations and solvent capacity in the cell," *J R Soc Interface,* vol. 17, no. 171, p. 20200656, Oct 2020, doi: 10.1098/rsif.2020.0656.



**Funding:** *This project is supported by the U.S. Department of Energy's Office of Biological and Environmental Research under contract* 78460.

**Title:** A Bioinformatic Pipeline to Identify and Classify Potential Microbial Signatures from Fungal Sequencing Data Integrated with a Searchable Database of Described Interactions

**Authors:** Michal Babinski[1*] (mbabinski@lanl.gov), Aaron J. Robinson[1*] (arobin@lanl.gov), Karen W. Davenport,[1] and **Patrick S. Chain[1]**

**Institutions:** [1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico

**Website URL:** https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-bacteria-fungal.php

**Project Goals:** The vast taxonomic diversity and the complexity of interactions within the soil microbiome present many challenges. Many of the interactions between soil-dwelling bacteria and fungi are not yet well-understood, and a more comprehensive understanding of these relationships and their response to environmental pressures would lead to substantial agricultural, environmental, and energy-focused advancements. These potential developments align with the foci of the DOE, and would influence multiple scientific disciplines. The aim of this Science Focus Area (SFA) is to better understand the diverse and abundant interactions within the soil rhizosphere, specifically between fungi and bacteria, and decipher the mechanisms behind their communication.

**Abstract Text:** Bacteria and fungi are dominant microorganisms in soil and terrestrial systems and are responsible for an array of ecological functions. Recent research by our group demonstrated that interactions between bacteria and fungi are more widespread and diverse than previously anticipated[1]. Bacterial associations with fungi are largely discovered through sequencing-based approaches, either amplification of 16S sequences directly from fungal isolates or the capture of bacterial sequences during sequencing of fungal genomes. Currently, there exist no standardized bioinformatic pipelines or tools to identify or classify potential BFI from diverse fungal sequencing datasets. In this presentation we will outline our efforts to develop a bioinformatic pipeline to detect and analyze potential bacterial associations from genomic and transcriptomic fungal sequencing datasets. Additionally, we will describe our efforts to expand the capability of this pipeline to identify other organisms and components which comprise the microbiomes of fungal isolates, such as viruses, archaea and even other fungi. After taxonomic classification, binned sequences will also undergo assembly and annotation through sub-pipelines that are optimized for each individual kingdom. The outputs of this pipeline include concise summarizies of the diversity of potential associations detected in the datasets, relevant statistics and metrics of detection and access to all sequences assembled or annotated during the pipeline. While this pipeline can be used to examine data from a single or small number of fungal sequencing projects, we have utilized it to screen all publicly available fungal genomic and transcriptomic datasets available through the NCBI SRA database. This resulted in optimization of the pipeline to allow for examination of large datasets in a reasonable time and manner. Furthermore, we have also developed the first comprehensive database of previously

described or observed BFI, which will be integrated with the pipeline to allow users to compare their results to previous knowledge and allow for the curated addition of newly discovered associations. Both the pipeline and database were designed in a manner that allows for smooth integration to platforms frequently used for microbial and microbiome research such as KBase and the NMDC. We believe this pipeline will provide a valuable resource not only for researchers in the BFI field, but for the larger fields of microbiology and microbial ecology through the provided means to assess potential impacts of microbial interactions in future or previously obtained datasets.

**References/Publications**

1. Robinson, A.J., House, G.L., Morales, D.P., Kelliher, J.M., Gallegos-Graves, L.V., LeBrun, E.S., Davenport, K.W., Palmieri, F., Lohberger, A., Bregnard, D. and Estoppey, A., 2021. Widespread bacterial diversity within the bacteriome of fungi. *Communications biology*, 4(1), pp.1-13.

**Funding Statement:** This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T.

**Development of a model root-associated microbial consortia to investigate the roles of bacterial-fungal-plant interactions in the heat- and drought-tolerant grass *Bouteloua gracilis* (blue grama)**

Buck T. Hanson* (bhanson@lanl.gov), **Patrick S. G. Chain**

Los Alamos National Laboratory, Los Alamos, New Mexico, USA

https://genomicscience.energy.gov/research/sfas/lanlbfi.shtml

## Project Goals

Biological networks in soils are important ecosystem drivers, however little is known about how taxonomic diversity and the complexity of interactions influences biogeochemical processes. Among the two dominant soil-welling microorganisms, bacteria and fungi, their interactions with abiotic and biotic environmental factors remain poorly understood. Increasing our knowledge of these interactions and their responses to environmental pressures would provide substantial advancements in agricultural, environmental, and sustainable energy alternatives, which are in alignment with several DOE priorities. The overarching goal of this SFA program is to improve our understanding of bacterial-fungal interactions (BFI) in soil rhizospheres and their roles in carbon cycling and plant productivity.

## Abstract

Optimizing plant health and resiliency to natural and anthropogenically induced stress plays a critical role in our ability to meet the growing needs for food and sustainable energy alternatives. Plants benefit from interactions with root-associated microorganisms, such as increased pathogen resistance, nutrient acquisition, resilience to abiotic stressors, as well as improved plant growth yields. Root-associated microbiomes have garnered increased attention as a tool to optimize plant health, however characterization and quantification of these microbial populations and the molecular mechanisms associated with plant health remain underexplored. This work aims to further develop blue grama as a model plant and understand the BFI that contribute to its heat and drought tolerance. Towards this goal, a model microbial consortia representative of the core bacteria and fungi associated with blue grama roots is being developed. Additionally, this work demonstrates the feasibility of utilizing blue grama in EcoFABs; a step forward in furthering blue grama as a model for studying bacterial-fungal-plant interactions.

*This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T*

**Title:** Bacterial Communities in Association to Specific Tissues in Wild Morels

**Authors:** Melissa Cravero[1], Guillaume Cailleau[1], Sami Zhioua[1], Patrick Hilpish[1], Celia Ruiz[1], Aaron Robinson[2], Demosthenes Morales[2], La Verne Gallegos-Graves[2], Julia M. Kelliher[2], Karen Davenport[2], Saskia Bindschedler[1], Patrick S. Chain[2], Pilar Junier[1]

**Institutions:**[1]Laboratory of Microbiology, University of Neuchâtel, Neuchâtel, CH-2000, Switzerland, [2]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

**Website URL:** https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-bacteria-fungal.php

**Project Goals:** The goal of this SFA is to obtain mechanistic insights into Bacterial-Fungal Interactions to inform predictive models of their function in soil and their responses to environmental perturbations, such as changing climate and agricultural practices.

**Abstract Text:** The fungal genus *Morchella* is an iconic fungal group with great ecological and economic importance. Because of their widespread consumption, they represent some of the most emblematic fungal groups known to humankind. In this study the bacterial communities associated with different types of tissues (fruiting bodies, mycelium, and sclerotia) in a diverse collection of morels (including wild individuals) were investigated. The fungi collected included representatives of the black and yellow morel clades. Regardless of the species investigated, *Pseudomonas* spp. were detected as the most prevalent associate bacterium in mycelium and sclerotia. Together with *Ralstonia* spp. (mycelium and sclerotia) and *Methylobacterium* spp. (sclerotia), they represented the core associated bacterial community. In contrast, a highly diverse bacterial community was found associated with fruiting bodies, with representatives of *Pedobacter* spp., *Deviosa* spp. and *Bradyrhizobium* spp. constituting the core bacterial community. Multiple strains from Pseudomonas spp. were isolated from mycelia during the cultivation process. Confrontation assays with *Morchella* spp. for these *Pseudomonas* spp. resulted in multiple types of positive or negative interactions. The sequencing of the genomes of these *Pseudomonas* spp. allowed the identification of gene clusters relevant to promote positive interactions with morels, including secretion systems and toxin-antitoxin systems. This study offers the first *in-vivo* evidence linking observations from soils and confrontation studies suggesting the relevance of *Pseudomonas* spp. on the physiology and development of morels.

**Funding Statement:** This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T.

**Title:** Comparative Genomics of Intracellular Mollicutes-related Bacterial Endosymbionts of Fungi

**Authors:** Reid Longley[1*] (longleyr@msu.edu), Aaron Robinson,[2] Julian Liber,[3] Abigail E. Bryson,[1] Kurt LaButti,[4] Robert Riley,[4] Kerrie Barry,[4] Igor V. Grigoriev,[4] Alessandro Desiro,[1] Patrick Chain,[2], **Gregory Bonito[1]**

**Institutions:** [1]Michigan State University, East Lansing, MI; [2]Los Alamos National Laboratory, Los Alamos, NM; [3]Duke University, Durham, NC; [4]Joint Genome Institute, Berkeley, CA

**Website URL:** https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-bacteria-fungal.php

**Project Goals:** The goal of this SFA is to obtain mechanistic insights into Bacterial-Fungal Interactions to inform predictive models of their function in soil and their responses to environmental perturbations, such as changing climate and agricultural practices.

**Abstract Text:** Diverse members of the early-diverging Mucoromycota including arbuscular mycorrhizal fungi and soil fungi in the Mortierellaceae are capable of harboring both Gram-positive Mollicutes-related endobacteria (MRE) and Gram-negative Burkholderia related endobacteria (BRE). Previous work has shown that MRE are dependent on their fungal host and their genomes are thought to be severely reduced in size. Given this, it has been hypothesized that MRE bacterial endosymbionts were acquired early, prior to the diversification of Mucoromycota. Alternatively, these myco-symbionts could have been acquired after the divergence of these lineages and spread horizontally between lineages. To address these hypotheses, we obtained four complete MRE genomes from two genera in the Mortierellaceae: *Linnemannia* (LMRE) and *Benniella* (BMRE). The size of these genomes ranged from 326 to 615 Kbp and includes the smallest known complete bacterial genomes of myco-symbionts. Comparative analyses of these genomes revealed unique content and organization with respect to each MRE lineage and provides insight as to how bacterial genomes may adapt to a particular fungal host. Homology based comparisons of predicted proteins revealed differences in genome reduction as a result of the endosymbiosis. Additionally, MRE protein lengths were significantly shorter on average compared to closely related *Mycoplasma* and *Sprioplasma* relatives. Multigene phylogenetic analysis also indicated that the MRE genomes within *Benniella* were more closely related to MRE from Glomeromycotina compared to MRE in more closely related *Linnemannia*. These results indicate that *Linnemannia* and *Benniella* isolates may have acquired their MRE after divergence from a common ancestor. The outcomes of this work expand upon foundational knowledge of the evolutionary impacts of bacterial-fungal interactions, towards the goal of continued investigations of evolution and impacts of these interactions on host and endosymbiont.

.

**Funding Statement:** This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T.

# Advanced Fluorescence Microscopy Techniques to Measure Bacterial:Fungal Interactions

Demosthenes P. Morales[1*] (dmorales@lanl.gov), Aaron Robinson[2], Julia Kelliher[2], Buck Hanson[2], Junier Pilar[3], Jim Werner[1], **Patrick Chain**[2]

[1]MPA-CINT: Center for Integrated Nanotechnologies, Los Alamos National Laboratory
[2]B-10: Biosecurity and Public Health, Los Alamos National Laboratory
[3]Universite de Neuchatel

https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-bacteria-fungal.php

**Project Goals: In support of the Department of Energy's mission towards energy security and reducing foreign energy dependence, the objective of this Science Focus Area (SFA) is to harness the soil microbiome as a means to achieve ecosystem steering and increase productivity in biofuel crop production. This SFA will focus on acquiring mechanistic insights into bacterial-fungal interactions in response to environmental perturbations to better understand soil dynamics. The challenge will be approached along a concerted effort across length scales from bioinformatics and metagenomic sequencing, to cell-cell interactions and molecular level validations. The insights acquired from these studies will inform predictive models that will work towards managing environmental conditions of soil to impact plant productivity especially for marginal land use.**

The complexity of interactions occurring between the diverse taxa found within the soil microbiome makes elucidating fundamental mechanisms for soil functioning a daunting task. Genomic, metagenomic, and metatranscriptomic sequencing efforts of soil microbes have provided a foundation for identifying key elements to focus experimental efforts and resolve complex mechanisms. Interkingdom interactions such as those shared between bacteria and fungi are especially interesting as these relationships are being understood to be quite cosmopolitan in nature. Ensemble multi-omic and other investigative tools are being explored to dissect these interactions to provide a holistic view of soil microbe behavior. However, there is a strong spatiotemporal contribution that dictates the influence of these behaviors. Here we are investigating the relationships of bacteria and fungi using advanced fluorescence microscopy techniques on cell-to-cell length scales to understand the heterogeneity of molecular responses between microbes that are separated spatially in soil microbiomes. Appreciating the spatial variation of molecular activity among soil microbes and even within single organisms, will provide better contexts to the distribution and diversity found soil environments.

This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T and in part the Center of Integrated Nanotechnologies, a DOE Office of Science user facility.

**Title:** Development of untargeted metabolomics approaches to study bacterial-fungal co-cultures

**Authors:** Baltazar Zuniga[1], Amy Zheng[2], Fabio Palmieri[3], Simone Lupini[4], Berkley Ellis[5], Aaron Robinson[6], Julia Kelliher[6], Demosthenes Morales[6], Debora F. Rodrigues[4], Pilar Junier[3], Jamey Young[1,2], Patrick Chain[6]

[1]Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA, [2]Department of Chemical and Biomolecular Engineering,Vanderbilt University, Nashville, TN, USA , [3]Laboratory of Microbiology, Institute of Biology, University of Neuchâtel, Emile-Argand 11, 2000 Neuchâtel, Switzerland; [4]Civil and Environmental Engineering, University of Houston, Houston, TX, USA,[5]Chemistry, Vanderbilt University, Nashville, TN, USA, [6] Biosecurity and Public Health Group, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM, USA

https://genomicscience.energy.gov/research/sfas/lanlbfi.shtml

**The goal of this project is to understand the complex metabolic interactions between bacterial and fungal partners in mixed culture environments that mimic natural soil microbiomes. Our team is developing a metabolomics pipeline and methods to identify metabolite targets to understand these interactions. These tools are currently being applied to compare the metabolism of multiple different co-culture pairs and identify the impact of growing conditions on their interactions using both targeted and untargeted GC-MS, LC-MS, and NMR methods.**

We aim to solve two significant problems using our metabolomics pipeline. First, our pipeline analyzes metabolites detected using GC-MS and LC-MS to produce high-quality tentative compound identifications. Second, our pipeline helps to generate biological interpretations from the compounds we tentatively identify. Our custom metabolomics data processing software performs relative quantification on these compounds and links the identification output to external databases like KEGG for further analysis. The software highlights conserved metabolites by cross-referencing potential targets against related KEGG metabolomes. This approach streamlines the data analysis workflow by prioritizing biologically relevant identifications and annotating the possible functions of these metabolites in relevant organisms, thus allowing us to visualize trends in pathway activation. By developing new software tools and experimental pipelines, we expect to better understand the microbial metabolic processes in fungal:bacterial co-cultures.

We also aim to identify specific metabolite targets of interest found in co-cultures. One such target is a red pigment produced by certain fungi in our culture collection when grown in co-culture with other fungi, thereby inhibiting their growth. Based on the gene clusters present, we hypothesized that the red pigment was bikaverin. We were able to purify the pigment and analyze its structure with NMR spectroscopy, which confirmed the identity of the pigment as bikaverin. We found that bikaverin is only produced in specific co-cultures, leading us to hypothesize that bikaverin could be produced as a defense mechanism to inhibit the growth of competing fungi.

*This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T.*

# Identification of novel tryptophan-derived metabolites and associated genes of Arabidopsis by integrating PODIUM and mGWAS

Zhiwei Luo[1*] (luo370@purdue.edu), Jeongim Kim[2], Brian Dilkes[1], and **Clint Chapple[1]**

[1]Purdue University, West Lafayette, IN and [2]University of Florida, Gainesville, FL

**Project Goals: A better understanding of plant genomes and metabolism is necessary to discover valuable plant-based compounds, design abiotic and biotic stress-tolerant crops, and improve plant productivity. The difficulties in metabolite identification and associated gene discovery limit the progress to achieve these goals. We developed an approach to address these challenges by integrating the Pathway of Origin Determination in Untargeted Metabolomics (PODIUM) and metabolic genome-wide association studies (mGWAS). The combination of these tools provides origin information for metabolites and identifies metabolite-gene associations. In this project, metabolites generated from amino acids and the genes required for their synthesis or regulation will be identified in Arabidopsis and sorghum.**

Tryptophan is a precursor for valuable compounds that play critical roles in plant growth and defense. Melatonin, a tryptophan-derived compound, is best known for its association with sleep-wake cycle control in mammals. In plants, melatonin has physiological effects on plant growth, photosynthesis, and responses to biotic and abiotic stressors. The biosynthetic route from tryptophan to melatonin in plants remains unclear. Arabidopsis produces other tryptophan-derived defense compounds to deal with herbivore attacks or pathogens, such as indole glucosinolates (IGs), camalexin, and indole-3-carboxylic acid (ICA). The biosynthesis of these compounds is relatively well characterized, but its regulation remains to be fully elucidated. Given the critical functions of these tryptophan-derived metabolites, we explored their biosynthesis and regulation by applying the PODIUM/mGWAS strategy on mutants disrupted in tryptophan metabolism. Differentially accumulated metabolite features (DAFs) generated from tryptophan were identified that may be intermediates in melatonin biosynthesis. By aligning tryptophan-derived DAFs with our mGWAS data, novel indole glucosinolate-gene associations were identified.

*Funding statement. This work was supported by the U.S. Department of Energy, Office of Science (BER), Grant DE-SC0020368.*

# A workflow for the systems-level analysis, design, and engineering of genomically recoded organisms

Anush Chiappino-Pepe[1,2]* (anush_chiappinopepe@hms.harvard.edu), Lisa Dratva[1,3], Alexandra Rudolph[1], and **George M. Church**[1,2]

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA; [2]Wyss Institute for Biologically Inspired Engineering, Boston, MA, USA; [3]School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

http://arep.med.harvard.edu

**We are establishing a computational workflow for the phenotypic analysis, design, and engineering of genomically recoded organisms. This workflow connects genome sequences to cellular phenotypes with improved nucleotide resolution by relying on a set of identified genome design rules. For this purpose, we have developed deep learning models connecting sequence-to-function of genetic features; extended genome-scale models of metabolism, expression, and regulation that predict cellular fitness from genome sequences; and methods for the analysis of recoding mutations.**

Genomically recoded organisms are tightly biocontained and biosolated (*e.g.*, virus resistant) and allow the efficient incorporation of multiple non-standard amino acids, making them attractive platform technologies for biotechnological, industrial, and biomedical applications[1–4]. In the process of recoding, we substitute a set of selected codons by synonymous ones throughout the entire genome[2,5]. Although protein sequences are maintained, we and others have observed considerable fitness reductions upon recoding[2,3,6,7]. To design viable genomically recoded organisms, we need computational methods that can connect genome sequences to cellular phenotypes with enhanced resolution than available methods.

Here, we present a workflow for the rational analysis, design, and engineering of genomically recoded organisms. First, we have developed deep learning models connecting sequence-to-function of genetic features that are mechanistically intractable. Second, we have constructed extended genome-scale models for the integrative analysis of metabolism, expression, and regulation. These models can predict cellular fitness from recoded genome sequences and can integrate large and disparate omics data sets. Finally, we have defined optimization-based methods for the computational analysis of recoding mutations. With this workflow, we can predict at a systems-level the combinatorial effect of recoding mutations on fitness, and we identify sets of mutational bottlenecks constraining growth.

We are applying this workflow to identify fitness-decreasing mutations in a 57-codon *Escherichia coli* genome[6]. In parallel, the models developed together with newly obtained RNA-seq data for strains containing segments of a 57-codon *Escherichia coli* genome are helping characterize the metabolic function arising from synthetic genome segments.

In the future, this workflow will serve to design genomically recoded organisms with minimal growth reduction and will provide a better understanding of genotype-to-phenotype relationships at a cellular level.

**References/Publications**


1. Ostrov, N. *et al.* Synthetic genomes with altered genetic codes. *Curr Opin Sys Biol* **24**, 32–40 (2020).
2. Lajoie, M. J. *et al.* Genomically Recoded Organisms Expand Biological Functions. *Science* **342**, 357–360 (2013).
3. Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514–518 (2019).
4. Robertson Wesley E. *et al.* Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372**, 1057–1062 (2021).
5. Isaacs, F. J. *et al.* Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **333**, 348 (2011).
6. Ostrov, N. *et al.* Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
7. Lajoie, M. J. *et al.* Probing the Limits of Genetic Recoding in Essential Genes. *Science* **342**, 361 (2013).



This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related financial interests: GRO Biosciences, ReadCoor; EnEvolv; and 64-x Bio. For a complete list of Dr. Church's financial interests, see also arep.med.harvard.edu/gmc/tech.html.

# A rapid biosensor engineering platform by translatable fluorogenic amino acids

Erkin Kuru[†,1,2*] (erkin_kuru@hms.harvard.edu), Jonathan Rittichier[†1,2], Isaac Han[2], Helena De Puig[2], Abi Reese[3], Fabio De Moliner[3], Jorge Marchand[1], Sylvie Bernier[2], William Bedell[2], Allison Flores[1], Marc Vendrell[3], **George M. Church[1,2]**

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA.

[2]Wyss Institute for Biologically Inspired Engineering, Boston, MA, USA.

[3]Centre for Inflammation Research, Queen's Medical Research Institute, The University of Edinburgh, Edinburgh, UK.

† Contributed equally

http://arep.med.harvard.edu

**Project Goals: Conditionally fluorescent, or fluorogenic amino acids (FgAAs), can dramatically increase their fluorescence when 'trapped' within macromolecules, for example at the antigen-antibody binding interface. Eventually, we intend to leverage this sensing ability to establish a machine-learning guided and fluorescent in situ sequencing-powered evolution pipeline towards de novo discovery of new binders that will also function immediately as instant biosensors. Toward this goal we begin exploring the molecular rules by which FgAAs can transform protein binders into instant sensors for their targets. We expect these optical biosensors to have immediately useful applications in basic biology research in synergy with the Aim 5 in our renewal grant.**

Modification of existing protein binders with fluorogenic probes can effectively transform the binder into an optical biosensor for its target, generating an easily detectable fluorescence change upon target binding. Variety of robust technologies can quickly evolve new and specific protein binders against virtually any target. However, the transformation of protein binders into fluorogenic sensors has been slow and generally limited to small signal changes. This is likely because current approaches rely on low-throughput screens that are not compatible with many probes, which greatly constrains the space. Such approaches are also not compatible with employment of a molecular evolution pipeline because the protein binder sequence cannot easily be fine-tuned once the fluorogenic residue matures, i.e., after the probe conjugation.

Here we present a two-stage biosensor engineering platform. The first, 'transformation', stage relies on the modular, and simple derivatization of protein binders with many new classes of fluorogenic probes. This also enables cost-effective and scalable (>25 mg from 1 L *E. coli* culture) biosensor manufacturing. The multiplexed exploration of hundreds of optimal probe, linker, and position combinations in parallel streamlines biosensor discovery to ~3 weeks from conception. We demonstrate the generalizable applicability of the platform via construction of multiple biosensors to distinct proof-of-principle antigens with ratiometric, nearly instant readouts. The second, 'evolution', stage relies on a new, highly efficient tRNA charging chemistry which enables streamlined engineering of the biosensor toward further optimization of

residues around the mature fluorogenic residue via compensatory mutations. Together, this platform will allow rapid conversion and fine-tuning of a wealth of available protein binders into optical biosensors that can serve as valuable tools for basic research.

This project has been funded by DOE grant DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related financial interests: ReadCoor; EnEvolv; and 64-x. For a complete list of Dr. Church's financial interests, see also arep.med.harvard.edu/gmc/tech.html.

**Profiling bacterial tRNA 3'-termini *in-vivo*: Establishment of a deep-sequencing approach and a riboswitch based bio-sensor approach.**

Jorge Marchand[1,4], Kamesh Narasimhan[1*] (Kamesh_Narasimhan@hms.harvard.edu), Akos Nyerges[1], Anush Chiappino-pepe[1], Maxwell Kazman[1], Michael C. Jewett[3] and **George Church**[1,2]

[1]Department of Genetics, Harvard Medical School, Boston, MA; [2]Wyss Institute for Biologically Inspired Engineering, Boston, MA; [3]Department of Chemical and Biological Engineering, NorthWestern University, Evanston, IL; [4]Department of Chemical Engineering, University of Washington, Seattle, WA.

http://arep.med.harvard.edu

**Project Goals: Functionally mature tRNAs across all domains of life have a conserved terminal 3'-CCA trinucleotide in their acceptor arm. *In-vitro,* it has been shown that tRNAs with modified 3'-acceptor termini can engage in orthogonal interactions with the ribosome paving the way for two orthogonal genetic codes to operate in parallel. To lay the foundations of an orthogonal genetic code *in-vivo*, we have developed a tRNA-seq approach to profile tRNA 3'-termini and quantitatively determine amino-acylation levels in *E.coli* strains with engineered tRNA processing pathways. Additionally, we are developing assays using a T-box riboswitch-based sensor that can report on the amino-acylation levels of target tRNAs.**

The 3'-CCA terminus of the tRNA engages in highly-conserved interactions with components of the translational machinery (Ribosome, EF-Tu and amino-acyl tRNA synthetases) and with the 3'-end processing and repair machineries (RnaseT, CCA-adding enzyme, RNAse BN, PNPase) [1]. As the 3'-CCA tRNA is a privileged scaffold, mutations in the 3'-CCA region tend to be the subject of surveillance and repair by *E.coli* tRNA processing pathways. Motivated by an *in-vitro* demonstration that ribosomes which carry compensatory mutations in their peptidyl-transferase centre (PTC) interact with variant 3'-tRNA acceptor arms, we are installing an orthogonal genetic code *in-vivo* [2] [3].

We have developed a tRNA sequencing approach that enables us to quantitatively assess 3'-tRNA end processing in genetically engineered *E.coli* strains. This led us to identify *E.coli* genotypes that obviate 3'-tRNA surveillance and repair. Additionally, we utilized the oxidative potential of sodium periodate to quantify the amino-acylation levels of tRNAs with variant 3'-acceptor ends. This method relies on selectively oxidizing 2'-3' *cis* diols of an uncharged tRNA, whereas the 3' end of charged tRNA remains protected from oxidation. We systematically tested various promoters, ribozyme designs and tRNA mini-helix designs to express multiple tRNAs with precisely processed 5' and 3'-termini *in-vivo*.

To further aid our efforts in incorporating 3'-tRNAs with variant acceptor arms into an orthogonal genetic code, we are harnessing a family of riboswitches called the T-box riboswitches[4]. T-box riboswitches are natural RNA sequences found in numerous prokaryotes that control the transcription or translation of downstream genes and have evolved to report on the aminoacylation state of tRNAs. Most importantly, the defining feature of T-box riboswitches is that they directly

base pair to the anticodon and 3'-acceptor end of tRNA, thus sensing the charged state of the 3'-tRNA acceptor end.

Together with the deep-sequencing and an engineered riboswitch based approach, we are operationalizing a fully orthogonal genetic code in *E.coli.*

**References**


1. Wellner, K., Betat, H. & Mörl, M. A tRNA's fate is decided at its 3' end: Collaborative actions of CCA-adding enzyme and RNases involved in tRNA processing and degradation. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1861,** 433–441 (2018).
2. Terasaka, N., Hayashi, G., Katoh, T. & Suga, H. An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. *Nat. Chem. Biol.* **10,** (2014).
3. Orelle, C. *et al.* Protein synthesis by ribosomes with tethered subunits. *Nature* **524,** 119–124 (2015).
4. Marchand, J. A., Pierson Smela, M. D., Jordan, T. H. H., Narasimhan, K. & Church, G. M. TBDB: a database of structurally annotated T-box riboswitch:tRNA pairs. *Nucleic Acids Res.* (2021). doi:10.1093/nar/gkaa721



This project has been funded by DOE grant DE-FG02-02ER63445. Prof. George Church is a founder of companies in which he has related financial interests: GROBiosciences; and 64-x. For a complete list of Dr. Church's financial interests, see also arep.med.harvard.edu/gmc/tech.html.

# Construction of a Synthetic 57-Codon *E. coli* Chromosome and Tools for Microbial Genome-Scale Recoding

Akos Nyerges[1*] (akos_nyerges@hms.harvard.edu), Anush Chiappino-Pepe[1,4], Shirui Yan[1,3], Maximilien Baas-Thomas[2], Nili Ostrov[1], Alexandra Rudolph[2], Jenny Ahn[1], Huseyin Tas[1], Yue Shen[3], and **George M. Church**[1,4]

[1]Department of Genetics, Harvard Medical School, Boston, MA; [2]Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA; [3]BGI Research, Shenzhen, China, [4]Wyss Institute for Biologically Inspired Engineering, Boston, MA

http://arep.med.harvard.edu

**Project Goals: We are finalizing the construction of a fully recoded 3.97 Mb *Escherichia coli* genome that relies on the use of only 57 codons. For this aim, the genome was previously computationally designed, synthesized, and assembled into 87 segments. In the final steps of genome construction, we combine and optimize these segments *in vivo* to assemble the fully recoded, viable genome.**

We present the construction of a fully recoded, 57-codon *Escherichia coli* genome, in which seven codons are replaced with synonymous alternatives in all protein-coding genes. For this aim, the entirely synthetic recoded genome was assembled as 87 50-kb episomal segments, individually tested for functionality, and then integrated into the genome. Developing a specialized integration system and optimizing our workflow enhanced integration efficiency to 100% and resulted in an order of magnitude increase in construction speed. We are now combining recoded clusters with a novel technology that builds on our latest developments in recombineering and CRISPR-associated nucleases[1,2].

In parallel with genome construction, we developed novel experimental methods to identify fitness-decreasing changes and troubleshoot these cases. Leveraging massively parallel genome editing and accelerated laboratory evolution[3] allowed us to correct partially recoded strains' fitness within days.

As we approach the final assembly of a virus-resistant *E. coli* genome, we also implement dependency on non-standard amino acids and encoding modules for stringent biocontainment and enhanced virus resistance.

**In sum, our work will soon I.) demonstrate the first 57-codon organism, II.) establish a tightly biocontained and virus-resistant chassis for new-to-nature protein production, and III.) open a new avenue for the bottom-up synthesis and refactoring of microbial genomes, both computationally and experimentally.**

## References


1. Wannier TM, Ciaccia PN, Ellington AD, Filsinger GT, Isaacs FJ, Javanmardi K, Jones MA, Kunjapur AM, Nyerges A, Pal C, Schubert MG, Church GM (2021) Recombineering and MAGE. *Nature Reviews Methods Primers*, 1(1):1–24. https://doi.org/10.1038/s43586-020-00006-x
2. Wannier TM, Nyerges A, Kuchwara HM, Czikkely M, Balogh D, Filsinger GT, Borders NC, Gregg CJ, Lajoie MJ, Rios X, Pál C, Church GM (2020) Improved bacterial recombineering by parallelized protein discovery. *Proceedings of the National Academy of Sciences*, https://doi.org/10.1073/pnas.2001588117
3. Nyerges Á, Csörgő B, Draskovits G, Kintses B, Szili P, Ferenc G, Révész T, Ari E, Nagy I, Bálint B, Vásárhelyi BM, Bihari P, Számel M, Balogh D, Papp H, Kalapis D, Papp B, Pál C (2018) Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. Proceedings of the National Academy of Sciences, 115(25):E5726–E5735. https://doi.org/10.1073/pnas.1801646115



This project has been funded by DOE grant DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related financial interests: ReadCoor; EnEvolv; and 64x Bio. For a complete list of Dr. Church's financial interests, see also arep.med.harvard.edu/gmc/tech.html.

# Genetic Tools for Photosynthetic Microbes; Toward Genome-scale Engineering

Max G. Schubert[1,4]* (max_schubert@hms.harvard.edu), Tzu-Chieh Tang,[1,4] Kenan Y. Topsakal,[1,4] Himadri Pakrasi,[2] Pramod Wangikar,[3] **George M. Church**[1,4]

[1]Harvard University, Boston, MA; [2]University of California at San Francisco, CA; [2]Washington University, St. Louis, MO; [3]IIT-Bombay, Powai, Mumbai, India; [4]Wyss Institute for Biologically Inspired Engineering, Boston, MA

http://arep.med.harvard.edu

**Project Goals: Expand recombineering techniques into model cyanobacteria, and explore recoding and pooled screening using these techniques. Explore new photosynthetic microbial systems and work to characterize new genetic tools.**

Tremendous genetic variation both exists in nature and is designed by genome refactoring and recoding efforts, but our ability to create and characterize genetic variants *in vivo* remains far more limited in scale. Facile genetic editing along with pooled screening approaches enable characterization of millions of genetic loci[1], but have yet to be applied more broadly outside of our best-characterized prokaryotic hosts. One focus of our efforts to deploy these tools more broadly are photosynthetic microbes, which serve both as capable models for understanding photosynthesis, and as a photosynthetic bioenergy chassis.

To explore recombineering approaches in photosynthetic organisms, we are focusing on UTEX2973 as a fast-growing and well-characterized host organism[2]. By searching both curated protein databases and metagenome datasets, we are able to locate phage recombination systems (similar to the canonical lambda-RED used in *E. coli*) within phage predicted to infect cyanobacteria, overcoming the host-range limitations of these systems[3-5]. We are testing these recombination systems in UTEX2973 to improve the efficacy of genome engineering and enable ssDNA recombineering approaches in cyanobacteria for the first time.

In tandem, we were inspired by other work isolating novel, fast-growing cyanobacterial models[6-8] to prospect for new photosynthetic microbes growing quickly in lab conditions. We have isolated several such new organisms and are further characterizing their genomes and phenotypes. This work can synergize with more detailed exploration of existing isolates, and to this end we are working with the Pakrasi and Wangikar labs to identify elements of interest in fast-growing cyanobacterial isolates, including novel retrons and episomes.

This work aims to pave the way for pooled screening in phototrophs, recoding efforts in phototrophs, and to inform future efforts to bring recombineering techniques into new prokaryotic hosts.

## References

1. Schubert, M. G. et al. High throughput functional variant screens via in-vivo production of single-stranded DNA. Cold Spring Harbor Laboratory 2020.03.05.975441 (2020) doi:10.1101/2020.03.05.975441.

2. Yu, J. et al. Synechococcus elongatus UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and $CO_2$. Sci. Rep. 5, 8132 (2015).

3. Filsinger, G. T. et al. Characterizing the portability of phage-encoded homologous recombination proteins. Nat. Chem. Biol. (2021) doi:10.1038/s41589-020-00710-5.

4. Wannier, T. M. et al. Improved bacterial recombineering by parallelized protein discovery. Proc. Natl. Acad. Sci. U. S. A. 117, 13689–13698 (2020).

5. Wannier, T. M. et al. Recombineering and MAGE. Nature Reviews Methods Primers 1, 1–24 (2021).

6. Jaiswal, D. et al. Genome Features and Biochemical Characteristics of a Robust, Fast Growing and Naturally Transformable Cyanobacterium Synechococcus elongatus PCC 11801 Isolated from India. Sci. Rep. 8, 16632 (2018).

7. Jaiswal, D. *et al.* A Novel Cyanobacterium Synechococcus elongatus PCC 11802 has Distinct Genomic and

Metabolomic Characteristics Compared to its Neighbor PCC 11801. *Sci. Rep.* **10**, 191 (2020).

8. Włodarczyk, A., Selão, T. T., Norling, B. & Nixon, P. J. Newly discovered Synechococcus sp. PCC 11901 is a robust cyanobacterial strain for high biomass production. *Commun Biol* **3**, 215 (2020).

**Funding Statement:** *This project has been funded by DOE grant DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related financial interests: ReadCoor; EnEvolv; and 64-x. For a complete list of Dr. Church's financial interests, see also arep.med.harvard.edu/gmc/tech.html.*

**The Global Prevalence and Biogeochemical Impact of Ancient Phosphorus-Oxidizing Bacteria**

Sophia D. Ewens (sdewens@berkeley.edu)* [1, 2], Tyler P. Barnum[1], Kyle Metcalfe[1], Mikayla A. Borton[3], Hans K. Carlson[3], Kelly C. Wrighton[4], **John D. Coates[1, 2]**

[1]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

[2]Energy & Biosciences Institute, University of California, Berkeley, CA, USA

[3]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA

[4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Lab, Berkeley, CA, USA

**Project Goals:** This project aims to investigate the role of microbial dissimilatory phosphite oxidation (DPO) in the global phosphorus and carbon biogeochemical cycles. We are examining the prevalence of DPO and phosphite ($HPO_3^{2-}$) in a broad range of geochemical environments, and to examine fundamental physiological and biochemical aspects of DPO. To achieve this goal we test three specific hypotheses:

1. DPO is an environmentally prevalent metabolism that co-occurs with global phosphite reserves.
2. DPO metabolism is universally conferred by the conserved *ptx-ptd* operon.
3. DPO is universally associated with $CO_2$ fixation

Phosphite is the most energetically favorable chemotrophic electron donor known, with a half-cell redox potential ($E^{o\prime}$) of −650 mV for the $PO_4^{3-}/PO_3^{3-}$ couple. Dissimilatory phosphite oxidizing microorganisms (DPOM) can harness the free energy from phosphite oxidation for cellular growth and concomitantly use that energy to fix $CO_2$. At the time of this work, the prevalence of this unique metabolism remained largely uncharacterized since its discovery in 2000, as only two DPOM species had been identified in three discrete locations. A false notion of rarity consequently limited our understanding of the diversity, environmental distribution, and biogeochemical impact of DPOM. However, phosphite has been detected in several environments at concentrations that suggest a contemporary P redox cycle that might sustain a greater diversity of DPOM than is currently recognized.

To survey the prevalence and diversity of DPOM, selective enrichments were inoculated with wastewater sludge from six different wastewater facilities around the San Francisco Bay area. Ion chromatography was used to monitor metabolic activity and identify active DPO enrichments, and metagenomic sequencing of active enrichments allowed for the identification of 21 DPOM. These DPOMs span six classes of bacteria, including the *Negativicutes*, *Desulfotomaculia*, *Synergistia*, *Syntrophia*, *Desulfobacteria*, and *Desulfomonilia_A*. Evolutionary analyses of all binned DPOM genomes suggest that modern DPOM are relics of an ancient ancestor whose capacity to perform DPO originated ~3.2 Gya. We compared the DPO

marker genes from these enriched genomes with over 17,000 publicly available metagenomes and found that DPO metabolism exists globally in diverse anoxic environments, including wastewaters, sediments, and subsurface aquifers. We subsequently tested for this metagenomic diversity in an expanded sampling regime that included the Sacramento River, San Francisco Delta watershed, San Francisco Bay, and coastal Pacific Ocean. We found that DPO activity was stimulated in 30% of environmental samples and that the 16S rRNA gene taxonomy mirrored the diversity that was identified in the original wastewater metagenomes. Relatives of *Phosphitivorax anaerolimi* Phox-21 dominated most sampled environments, but we hypothesize that divergent environmental conditions are likely to select for phylogenetically diverse DPOM.

We performed metabolic analyses of metagenome-assembled genomes and found that most DPOM are specialists that use phosphite as their sole electron donor and $CO_2$ as their preferred electron acceptor. This metabolic niche would poise DPOM to be primary producers while providing nutritional phosphate to their local microbial community, signifying that DPO can sustain geographically isolated microbial communities using exclusively phosphite and $CO_2$ to generate energy and fixed carbon. However, we find that this pivotal role may be mutually dependent on the activity of the local microbial community, as physiological data suggests that DPOM require symbiotic nutrient exchange to support growth. We explore the potential microbial interactions that could support DPOM activity and find that DPOM enrichments generate diverse corrinoids. One of those corrinoids is novel, and its characterization may lead to crucial insights about DPO activity.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0020156

**Title:** Defining the Influence of Environmental Stress on Bioenergy Feedstocks at Single-Cell Resolution

**Authors:** Margot Bezrutczyk[1], Danielle Goudeau[1], Rex Malmstrom[1], **Benjamin Cole[1]***

**Institutions:** [1]DOE-Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA

**Project Goals:** This Early Career Research Project aims to build comprehensive single-cell resources for bioenergy grasses, characterizing sorghum, switchgrass, and Brachypodium cell types at unprecedented scale and resolution. These resources will be used to investigate cell type-specific environmental stress responses under agriculturally relevant conditions. In addition, we aim to understand how microorganism influence these responses. These efforts will ultimately lead to enhanced understanding of the multifaceted ways that plants react to changing conditions.

**Abstract Text:** Biomass from plants is an important resource that enhances energy independence and promotes good environmental stewardship. Poor growing conditions (such as drought or low soil nutrient composition) hinders optimal performance of these bioenergy crops. Therefore, the development of new strategies to improve plant biomass production will require a better understanding of how plants respond to (and cope with) stress imposed by the environment. Plant responses to drought are complex and involve the coordinated action of many different types of cells with specialized functions (cell types). For example, cells that compose stomata (pores in the leaf that open and close to exchange carbon dioxide and oxygen, potentially leading to water loss) will respond very differently to drought than cells of the plant vasculature.

This project aims to use cutting-edge single-cell characterization technologies to measure how individual cells and cell types respond to drought and nutrient limitation in two prominent bioenergy crops, sorghum and switchgrass. This will require the construction of large curated datasets detailing the regulation of genes in hundreds of thousands of individual plant cells. In addition, we will quantify how genes are turned on or off under drought and nutrient stress when grown under conditions very similar to those in agricultural plots using sophisticated plant growth chambers. Lastly, the we will investigate the impact of beneficial microorganisms in the soil to plant growth under stress. The results of this project will significantly advance our foundational knowledge of how plants coordinate responses to environmental stress, and will ultimately enable us to target genes in specific cell types for crop improvement. Here, we will present our experimental design, expected outcomes, and preliminary data from our single-cell atlas construction efforts.

**Funding Statement:** The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

**Title: Understanding the Effects of *Populus*—Mycorrhizal Associations on Plant Productivity and Resistance to Abiotic Stress**

**Authors:** Alyssa A. Carrell[1], Brandon Kristy[1], David McLennan[1], Miranda Clark[1], and **Melissa A. Cregger[1*]**(creggerma@ornl.gov)

**Institutions:** [1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

**Project Goals:** The overarching goal of this project is to develop sustainable, multipurpose bioeconomies whereby globally important feedstocks can be produced while simultaneously maximizing soil health and mitigating adverse impacts of climatic change. In this project, I will leverage the unique ability of *Populus sp*. to associate with both ecto- and arbuscular mycorrhizal fungi to examine how variation in these associations alters plant productivity, abiotic stress response, and belowground soil carbon cycling.

**Abstract text:** Within the myriad of possible plant–microbe interactions occurring belowground, plant–mycorrhizal associations are widespread with the two most common mycorrhizal types being ectomycorrhizae (ECM) and arbuscular mycorrhizae (AM). Belowground plant interactions with these fungi have been shown to increase water uptake and nutrient acquisition and alter soil carbon storage. It is unclear how these two dominant mycorrhizal fungal types differ in their abilities to offer benefits to the plant and change belowground carbon and nutrient cycling. Most plants associate with one type of mycorrhizal fungi, and individual plant species are less likely to associate with both AM and ECM fungal species. Contrary to this, *Populus sp.* uniquely associate with AM and ECM simultaneously in natural settings, thus providing an ideal experimental system in which I will examine how mycorrhizal fungal types confer benefits to their host. Using high-throughput plant phenotyping, neutron imaging, greenhouse and field experiments I will first characterize how variation in *Populus* – mycorrhizal associations alter the plant response to drought, which I will then use to inform plant–mycorrhizal interaction manipulation experiments to ultimately influence plant productivity, increase plant drought tolerance, and enhance soil health.

**Funding statement:** Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR45678. This research is supported by the U.S. Department of Energy Office of Science, through the Office of Biological and Environmental Research (BER) Early Career Research Program. Preliminary data was collected through the BioEnergy Science Center and Center for Bioenergy Innovation, Bioenergy Research Centers funded by BER.

**Genetic mapping of sugarcane aphid resistance in sorghum line SC112-14**

**Hugo Edgardo Cuevas**, USDA-ARS, Mayaguez, PR,
Joseph E. Knoll, Crop Genetics and Breeding Research Unit, USDA-ARS, Tifton, GA, Jason

Abstract Text:

Sugarcane aphid [*Melanaphis sacchari* (Zehntner)] is a destructive pest of sorghum [*Sorghum bicolor* (L.) Moench] worldwide. Since 2013, sugarcane aphids (SCA) began to infect grain sorghum in south Texas, spread across North America and has become a significant pest in all sorghum growing areas of the continent. The economic impact on the sorghum industry have led to multiple research efforts directed at identifying different resistance sources and the most adequate strategy to control the pest. Today, a limited number of resistance sources have been identified in sorghum germplasm adapted to temperate regions. In this study, a recombinant inbred line (RILs) population derived from crossing the sugarcane aphid (SCA) resistant line SC112-14 with the susceptible line PI609251 were evaluated for their SCA resistance response in Georgia during two consecutive years. The resistance response was determined based on the aphid population size (APS) and plant damage (APD). Each RIL plot was rated twice each year before flowering, approximately two weeks apart, to generate two APS (APS_1 and APS_2) and APD (APD_1 and APD_2) scores. Segregation for SCA resistance was observed for the first APS and both APD scores, and the broad-sense heritability estimate ranged from 0.71 to 0.76, respectively. A genome scan using a high-density linkage map of 3,838 SNPs detected an 81 kb genomic region in chromosome 6 with 2.5 cM of length (Figure 1). This locus explained 50 - 55% of the phenotypic variation and includes seven genes including two (Sobic06G.015200 and Sobic06G.015250) that have protein kinase and leucine rich repeats domains, both features are common among *R*-genes. Comparative mapping analysis found that the resistance source in SC112-14 is new and is located 8 and 10 cM downstream of the Henong 16 (*RMES-1*) and Tx2783 resistance loci, respectively (Figure 2). Previous studies found that the SC112-14 anthracnose resistance response was controlled by a locus in chromosome 5, therefore, this line is becoming an important germplasm that can be immediately used in breeding programs for the development of new sorghum hybrids and varieties with resistant to anthracnose and sugarcane aphid.

**Figure 1**. Genome scan for sugarcane aphid (SCA) resistance response in recombinant inbred lines derived from the cross of SC112-14 and PI609251 evaluated in Tifton, Georgia, U.S. in 2018 and 2019. Inclusive composite interval mapping (ICIM) using the logarithm value of the first SCA population size and two scores of plant damage. The first and second score were taken two weeks apart.

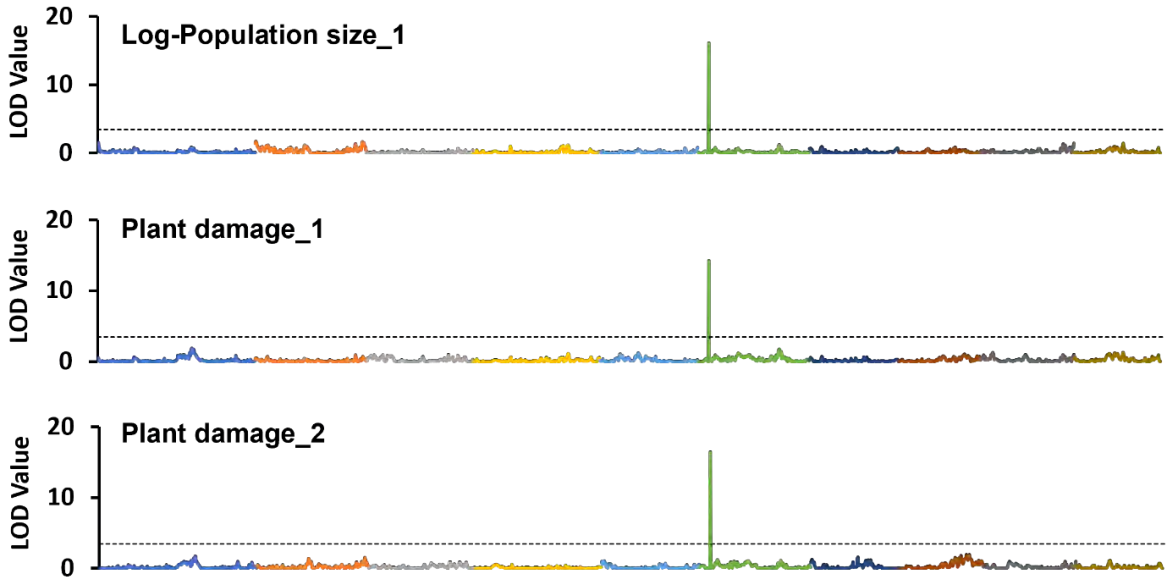
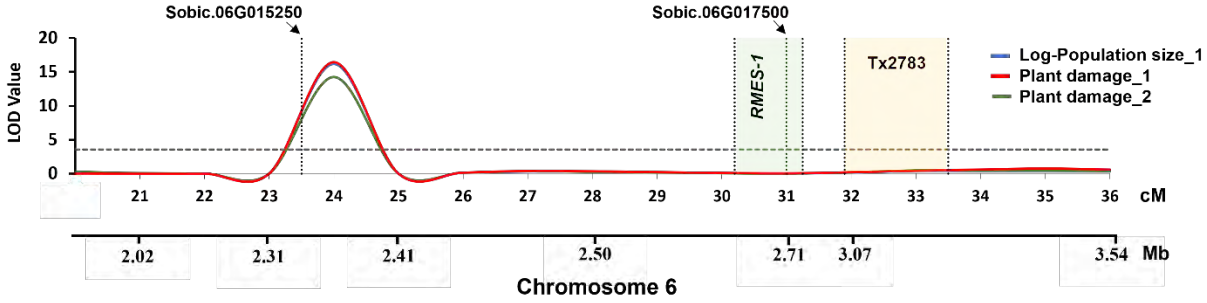



**Figure 2** Genomic regions associated with the sugarcane aphid (SCA) resistance response in sorghum line SC112-14. Candidate genes and genomic regions associated with the SCA resistance response observed in sorghum line Henong 16 (*RMES-1*; Wang et al. 2013), 407B (Zhang et al. 2020) and Tx2783 (Wang et al. 2021) are delimited within the associated region.




## Funding statement

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0014171 and DE-SC000019

**Applying new models to describe biomembrane structure and solvent partitioning in living cell membranes and membrane mimics.**

Jonathan D. Nickels[1]* (nickeljo@ucmail.uc.edu), Micholas Dean Smith[2,3], Luoxi Tan[1], Haden L. Scott[3], John Katsaras[3], Hugh M. O'Neill[3], Sai Venkatesh Pingali[3], Jeremy C. Smith[2,3], James G. Elkins[3], Brian H. Davison[3]

[1]University of Cincinnati, Cincinnati, Ohio; [2]University of Tennessee, Knoxville, Tennessee; [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee

https://cmb.ornl.gov/dynamic-visualization-of-lignocellulose/

**Project Goals: The development of renewable biofuels is a key mission of the DOE Genomic Science program. Lignocellulosic biomass has the potential to be an abundant, renewable source material for production of biofuels and other bioproducts. The use of organic solvents to optimize biomass pretreatment has shown considerable promise, but their disruption of microbial membranes is key to toxic effects limiting fermentation titers. The Oak Ridge National Laboratory (ORNL) Scientific Focus Area (SFA) Biofuels Program utilizes multi-length scale imaging with neutron scattering complemented by high performance computer simulations, NMR, biochemistry and targeted deuteration to provide fundamental knowledge about the molecular forces that drive solvent disruption of the critical assemblies of that comprise plant cell walls and microbial biomembranes.**

Amphiphilic co-solvents have a significant impact on the structure, organization, and physical properties of lipid bilayers. The cell membrane is defined by its transverse structure, an approximately five nanometer thick selectively semi-permeable lipid membrane; but it is so much more. Compositionally complex, dynamic, and organized in both the transverse and lateral dimensions, understanding the cell membrane structure – and the role that structure plays in cellular function, communication, and environmental sensing is an active scientific effort. Describing the mutual impact of partitioning and induced structure changes is therefore a crucial consideration in bioenergy research for microbial solvent tolerance in the production of biofuels and other fermentation products where molecules such as ethanol, butanol, or acetic acid might be generated by fermenting microbes; or when residual solvents such tetrahydrofuran (THF) are present from cellulose extraction procedures. Small angle neutron scattering (SANS) is a key method for studying lipid and polymer bilayer structures, with many models for extracting bilayer structure (thickness, area per lipid, etc.) from scattering data in use. However, the molecular details of co-solvent partitioning are conflated with induced changes to bilayer structure, making interpretation and modeling of the scattering curves a challenge. To address this issue, we present a model of bilayer structure which includes a two-term partition constant accounting for the localization of the co-solvent within the bilayer. We validate this model using

a series of SANS measurements of lipid vesicles in the presence of the co-solvent THF, showing several strategies of how to deploy the two-parameter partition coefficient model to describe scattering data and extract both structure and partitioning information from the data. The associated code will be publicly deposited pending publication. Molecular dynamics (MD) simulations are used to both evaluate underlying assumptions of the new data fitting model and illustrate its complementary approach to the data fitting procedure for our model membrane standard (phosphatidylcholine lipids) and a lipid mixture mimicking the *B. subtilis* cell membrane extract. We subsequently demonstrate the use of MD derived estimates of solvent partitioning to refine our SANS modeling. The new structure/partitioning model has been applied to solvent partitioning in the cell membrane of *Bacillus subtilis.* An updated and improved method of isotopic labelling has been developed. Previously, we have devised a novel isotopic labelling approach to enable direct *in vivo* structural study of the cell membrane of the gram-positive organism, *B. subtilis*, using neutron scattering. This was accomplished through a genetic inhibition of fatty acid degradation (*ΔyusL*) and a chemical inhibition of fatty acid biosynthesis through cerulenin. Here, we improve upon the previous system by introducing a dCas9/sgRNA-*fabF* complex that blocks transcription of the essential *fabF* gene when under xylose induction. This leads to greater sensitivity to cerulenin and more robust cell growth when supplementary fatty acids are introduced. A subtle change in fatty acid uptake is noted which manifests as an increase in the membrane thickness determined via neutron scattering. This enables improved investigations of cellular uptake and utilization of fatty acids, cell membrane structure and organization as a phenotypic response to metabolic and environmental changes. SANS observations of live cells and lipid extracts in the presence of co-solvents reveal bilayer thinning and estimates of partitioning of the co-solvent. These are currently being analysed using the new model. Initial analysis of complementary MD simulations compares favorably to both structure (thinning) and solvent partitioning from SANS.

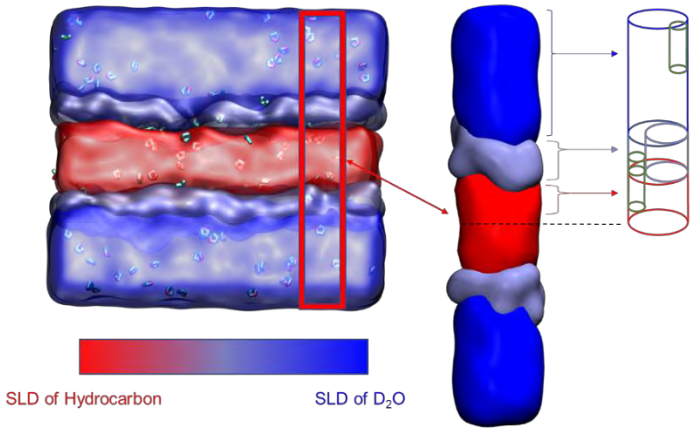



Figure 1. The two-parameter partition constant model presented for the case of a lipid bilayer (DMPC) in water with THF as a co-solvent. The bilayer structure is approximated as a three-slab model: the hydrophobic core containing a portion of the lipid and co-solvent – central symmetric outer layers containing a portion of the lipid, water, and co-solvent – and the bulk solvent containing water and co-solvent. Given a knowledge of the bilayer chemistry, atomic scattering length and molecular volumes, these relationships can define the partitioning of the co-solvent and bilayer structure using the area per lipid ($APL$), number of water molecules per lipid headgroup ($N_w$), the partition constant ($K_p$) and the *co-solvent localization constant* ($P_s$).

**Funding Statement:**
This research is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP752. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

**Noncellulosic biopolymer morphology and structural changes during real-time reaction studies.**

Zhi Yang,[1,2] Marcus Foston,[3] Hugh O'Neill,[1] Volker S. Urban,[1] Arthur Ragauskas,[1] Barbara Evans,[1] Sai Venkatesh Pingali[1]* (pingalis@ornl.gov), and **Brian H. Davison[1]**

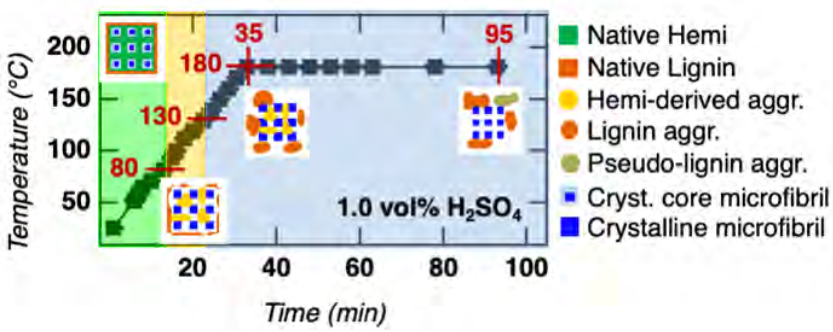
[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Massey University, Auckland, New Zealand (current); and [3]Washington University at St. Louis, St. Louis, Missouri.

https://cmb.ornl.gov/dynamic-visualization-of-lignocellulose/

**Project Goals: The development of renewable biofuels is a key mission of the DOE Genomic Science program. Lignocellulosic biomass has the potential to be an abundant, renewable source material for production of biofuels and other bioproducts. The use of organic solvents to optimize biomass pretreatment has shown considerable promise, but their disruption of microbial membranes is key to toxic effects limiting fermentation titers. The Oak Ridge National Laboratory (ORNL) Scientific Focus Area (SFA) Biofuels Program utilizes multi-length scale imaging with neutron scattering complemented by high performance computer simulations, NMR, biochemistry and targeted deuteration to provide fundamental knowledge about the molecular forces that drive solvent disruption of the critical assemblies of biomolecules that comprise plant cell walls and microbial biomembranes.**

Plant cell wall structure of biomass is an intricate design of several carbohydrate polymers encased in the hydrophobic lignin polymer to protect against degradation. Production of second-generation bioethanol from lignocellulosic biomass requires thermochemical pretreatment to open this complex plant cell wall structure and consequently improve enzyme access. However, the recalcitrant nature of lignin negatively affects efficient enzymatic access. Several different thermochemical pretreatments have been extensively developed and employed, but the exact nature of plant cell wall recalcitrance and the most efficient and economical approach to alter plant cell wall structure via pretreatment remains elusive. To understand the role of noncellulosic switchgrass polymers on the overall efficiency of pretreatment, the structural evolution of the noncellulosic polymers of the plant cell wall was investigated during dilute acid pretreatment (DAP) by employing in-situ small-angle neutron scattering (in-situ SANS). In this study, we observed real-time structural changes not possible to observe by any other technique. To deconvolute the structural evolution of lignin and hemicellulose polymers during DAP, native switchgrass (NATV), and isolated holocellulose (HOLO) and cellulose (CELL) fractions from NATV were studied. Our results show that aggregate particles first appear around 80 °C for NATV and HOLO samples. Due to the low temperature and pretreatment severity condition, these particles are likely derived from hemicellulose. The formations of much larger aggregate particles, only observed in the NATV sample, were attributed to lignin. For the HOLO sample, as the temperature and pretreatment severity condition increased, hemicellulose-derived aggregate particle sizes increased, suggesting this process was the nucleation and early stage formation of pseudolignin particles. Consistent with our interpretation of structural evolutions in NATV and HOLO samples, no formation of aggregate particles was observed in CELL samples for the

entire duration of the pretreatment. These results suggest that not only lignin but also hemicellulose can form aggregate particles within plant cell walls during pretreatment.




## References/Publications


1. Yang, Z., Foston, M.B., O'Neill, H., Urban, V.S., Ragauskas, A., Evans, B.R, Davison, B.H., and Pingali, S. V. "Structural Reorganization of Non-Cellulosic Polymers Observed *in situ* during Dilute Acid Pretreatment by Small Angle Neutron Scattering." *ACS Sust Chem Eng* (2021). doi: 10.1021/acssuschemeng.1c06276


## Funding Statement:


This research is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP752. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

**Solid state NMR characterization of lipid membrane and organic solvent induced effects**

Yunqiao Pu[1]* (puy1@ornl.gov), Haden Scott,[1] James G. Elkins,[1] John Katsaras,[1] Jonathan Nickels,[2] Arthur J. Ragauskas,[3] and **Brian H. Davison[1]**

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2] University of Cincinnati, Ohio; and [3]University of Tennessee Knoxville, Tennessee

https://cmb.ornl.gov/dynamic-visualization-of-lignocellulose/

**Project Goals: The development of renewable biofuels is a key mission of the DOE Genomic Science program. Lignocellulosic biomass has the potential to be an abundant, renewable source material for production of biofuels and other bioproducts. The use of organic solvents to optimize biomass pretreatment has shown considerable promise, but their disruption of microbial membranes is key to toxic effects limiting fermentation titers. The Oak Ridge National Laboratory (ORNL) Scientific Focus Area (SFA) Biofuels Program utilizes multi-length scale imaging with neutron scattering complemented by high performance computer simulations, NMR, biochemistry and targeted deuteration to provide fundamental knowledge about the molecular forces that drive solvent disruption of the critical assemblies of biomolecules that comprise plant cell walls and microbial biomembranes.**

Organic solvent is widely acknowledged as a major determinant of product titer limitations of fermentations due to its disruption and induced destabilization of microbial cellular membranes. Amphiphilic solvents disrupt the molecular packing and organization of lipid membranes, causing fluidity alteration and reducing structural integrity. In this study, solid state nuclear magnetic resonance (NMR) was employed to investigate the effects of organic solvents on phase and dynamics of lipid membranes. Model lipid 1-palmitoyl-2-oleoylglycerophosphoglycerol (POPG) was used to prepare a multilamellar lipid vesicle (MLV) and two organic solvents, i.e., n-butanol and tetrahydrofuran (THF), were used. High resolution solid state $^{1}H$, $^{31}P$ and $^{13}C$ NMR experiments with various pulse sequences including cross-polarization (CP), direct-polarization (DP), and insensitive nuclei enhancement by polarization transfer (INEPT) under magic-angle spinning (MAS) were used to investigate the phase behavior of membrane in the presence of solvents. The interaction of POPG with the solvent molecules was examined by using two-dimensional nuclear Overhauser effect spectroscopy (NOESY) spectra. The static solid state $^{31}P$ NMR spectra showed that the phospholipids had a combination of phases with powder distribution and isotropic peaks. The lipid membrane was also observed to adapt to various orientations depending on the solvents. For example, butanol addition caused the lipid molecules to orient both parallel and perpendicular to the magnetic field, while the presence of THF changed the orientation of the lipid molecules to be primarily perpendicular to the magnetic field. In this work, we are extending the use of Solid state NMR to provide fundamental insights on solvent-membrane interaction and dynamics and phase behavior of membranes.

**Funding Statement:**
This research is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP752. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

**Effect of Cyrene Pretreatment on Switchgrass Lignin Structure**

Yun-Yan Wang[1*] (ywang226@utk.edu), Luna Liang[1], Xianzhi Meng[1], Yunqiao Pu[2], Micholas Dean Smith[1], Arthur Ragauskas[1,2], and **Brian H. Davison[2]**

[1]University of Tennessee Knoxville, Knoxville, Tennessee; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee

https://cmb.ornl.gov/dynamic-visualization-of-lignocellulose/

**Project Goals: The development of renewable biofuels is a key mission of the DOE Genomic Science program. Lignocellulosic biomass has the potential to be an abundant, renewable source material for production of biofuels and other bioproducts. The use of organic solvents to optimize biomass pretreatment has shown considerable promise, but their disruption of microbial membranes is key to toxic effects limiting fermentation titers. The Oak Ridge National Laboratory (ORNL) Scientific Focus Area (SFA) Biofuels Program utilizes multi-length scale imaging with neutron scattering complemented by high performance computer simulations, NMR, biochemistry and targeted deuteration to provide fundamental knowledge about the molecular forces that drive solvent disruption of the critical assemblies of biomolecules that comprise plant cell walls and microbial biomembranes.**

To satisfy sustainability criteria for the biorefinery, the "closed-loop" biorefinery process using biomass-derived solvents has been proposed as a new concept for the realization of profitable liquid fuels and bioproducts produced from lignocellulosic biomass. A cellulose derived green solvent, Cyrene, has shown high potential in poplar biomass pretreatment. However, due to its high dynamic viscosity, at high concentrations Cyrene could cause negative effects on the sugar release of the pretreated biomass as well as driving up the operational cost of the lignin recovery. Using a combination of molecular simulations and experimental optimization, we examined impact of Cyrene concentration on lignin release. Simulations of lignin pretreated in neat Cyrene, 80:20 Cyrene:water, and 70:30 ethanol:water mixtures indicate that while lignin aggregate disruption does occur under all Cyrene containing conditions, in a manner similar to traditional organosolv (ethanol-water) pretreatments, long relaxation times associated with Cyrene reorganization substantially slows this release process compared to ethanol-water controls. Additionally, simulations do indicate that the addition of water does decrease these reorganization times. Subsequently, experimental studies of Cyrene pretreatment with three different Cyrene concentration were performed on switchgrass in dilute acidic aqueous under mild conditions. Results indicated that loss of pretreatment efficacy caused by low Cyrene concentration could be compensated by prolonged pretreatment time and high catalyst dosage. The switchgrass lignin extracted by Cyrene pretreatment possessed high preservation of β–O–4 ether inter-unit linkage, which could provide versatility in the integration of downstream lignin valorization into the modern biorefinery industries.

**Funding Statement:**
This research is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP752. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

**Title:** Probabilistic Annotation and Ensemble Metabolic Modeling in KBase

**Authors: Patrik D'haeseleer[1*]** (dhaeseleer2@llnl.gov), Jeffrey Kimbrel[1], Ali Navid[1], Chris Henry[2], Rhona Stuart[1]

**Institutions:** [1]Lawrence Livermore National Laboratory , Livermore, CA; [2]Argonne National Laboratory, Lemont, IL

**Website URL:** https://www.kbase.us/research/stuart-sfa/

**Project Goals:** We are developing tools for the DOE Systems Biology Knowledgebase (KBase) to give users a principled way to weight multiple sources of functional annotation against each other, enable better metabolic modeling of hard-to-annotate organisms and pathways, allow analysis of uncertainty in the resulting models network structure or behavior, and provide an infrastructure on which to build more sophisticated machine learning techniques in KBase.

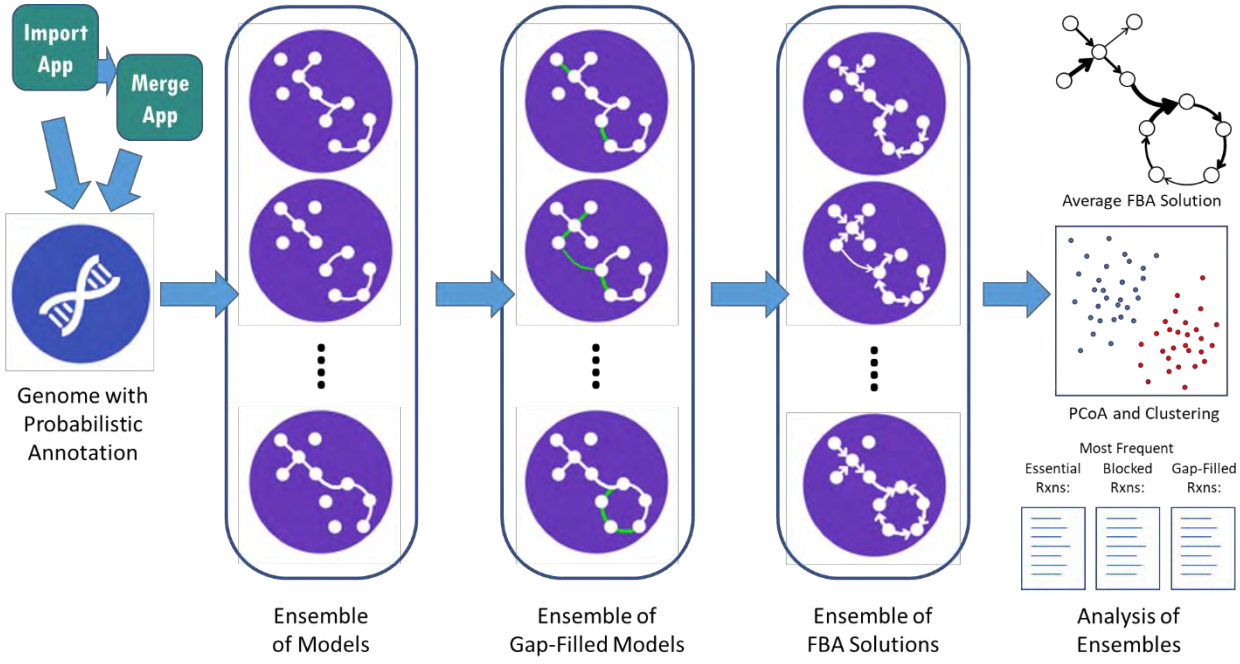
**Abstract Text:**

Our μBiospheres SFA investigates metabolic interactions in bioenergy-relevant microbial communities. A critical part of this research is development of genome-scale models of metabolism, which requires well-annotated genomes. We have found that combining annotations from multiple sources results in a more complete metabolic network reconstruction, greatly reducing the effort required to curate quality metabolic models (1). In a previous round of plus-up funding from DOE, we implemented a set of KBase Apps that allow users to upload metabolic annotations from multiple functional annotation tools, compare and merge these annotations, and to use them for model building and gapfilling to achieve significantly improved metabolic models. These Apps have proven to be very useful and are currently in daily use in our own SFA and several other research groups using KBase.

However, it is quite common for functional annotation tools to disagree on the function that should be assigned to certain genes, and this uncertainty can have significant consequences on the resulting metabolic networks and the behavior they predict for the organism. We are developing a set of KBase apps – and the underlying infrastructure to enable them – to allow the user to take advantage of these multiple inputs and explore the uncertainty and incompleteness of the functional annotations in their organism of interest in a probabilistic modeling framework. We will provide Apps to import annotation probabilities where available, implement Bayesian methods for merging annotation probabilities, generate an ensemble of models by sampling from the underlying probability distribution, sample alternative gapfilling solutions using the Medusa COBRApy package by Medlock and Papin (2), and then analyze the results of ensemble modeling. This ensemble approach will allow us to leverage all the existing KBase tools for importing and merging metabolic annotations, metabolic modeling, and gapfilling.

This work will provide our SFA and other KBase users a principled way to weight annotation sources against each other, enable better metabolic modeling of hard-to-annotate organisms and pathways, allow analysis of uncertainty in the resulting models network structure or behavior, and provide an infrastructure on which to build more sophisticated machine learning techniques in KBase.

These Apps will be broadly applicable to a wide range of users interested in model building using multiple annotation sources in KBase, as well as groups working on modeling challenging organisms and pathways, and those who would like to have access to state-of-the-art ensemble modeling tools.




Proposed workflow diagram, starting from imported annotation probabilities or probability estimates generated by our Merge app.

**References/Publications**


1. Griesemer M, Kimbrel JA, Zhou CE, Navid A, D'haeseleer P. Combining multiple functional annotation tools increases coverage of metabolic annotation. BMC Genomics. 2018 Dec 19;19(1):948.
2. Medlock GL, Moutinho TJ, Papin JA. Medusa: Software to build and analyze ensembles of genome-scale metabolic network reconstructions. PLoS Comput Biol. 2020 Apr 29;16(4):e1007847.



**Funding Statement:** This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL μBiospheres SFA, FWP SCW1039.

**Title: Mining Innovative Strategies for Stress-resilient Growth in Extremophyte Relatives of Brassicaceae Oil Crops**

**Authors:** Ying Sun,[1] Dong-Ha Oh[2*] (ohdongha@lsu.edu), Kieu-Nga Tran[2], Lina Duan[1], Prashanth Ramachandran[1], Andrea Ramirez[1], Guannan Wang[2], Maheshi Dassanayake[2], **José R Dinneny[1]**

**Institutions:** [1]Stanford University, Stanford, CA; [2]Louisiana State University, Baton Rouge, LA

**Project Goals:**

Objective 1: Discover how extremophytes and stress sensitive species differ in the cell-type functions of roots and those triggered downstream of ABA.
Objective 2: Define how changes in the wiring of gene regulatory networks produce innovations in transcriptional regulation in extremophytes and how bioenergy crops have diverged.
Objective 3: Establish a data driven, predictive framework for accelerating functional testing of stress resilience genes using *Arabidopsis* and *Camelina* as a chassis for engineering.

**Abstract Text:**

Wild relatives of crops, especially "extremophytes" naturally adapted to harsh environments, harbor traits that may have been lost during crop domestication, as well as novel innovations that enabled survival and growth under stresses. Two extremophyte oil crop-relatives, *Schrenkiella parvula* and *Eutrema salsugineum* in Brassicaceae, offer models to define gene functions contributing to physiological mechanisms that enable survival and growth under salt stresses and sub-optimal nutrient conditions. We used a comparative analysis of stress-responsive transcriptomes, ionomes, metabolomes, and the gene regulatory networks (GRN) between the extremophyte models and their stress-sensitive relatives, including *Arabidopsis thaliana*, to identify innovations in stress response strategies and genetic variations behind them.

We found that, while achieving the similar level of salt tolerance, *S. parvula* and *E. salsugineum* take distinct strategies [1]. While *S. parvula* allows tissue $Na^+$ accumulation similar to *A. thaliana* upon salt stresses, *E. salsugineum* appeared to limit the net $Na^+$ entry. Still, both extremophytes maintained ion nutrient levels including $K^+$, which *A. thaliana* failed to do so. Both extremophytes accumulate metabolites that can serve as osmo-protectants and antioxidants in response to salt treatment by following different trajectories. *E. salsugineum* metabolome showed stress readiness and pre-adaptation, coupled with its overall basal-level growth rates compared to both *A. thaliana* and *S. parvula*. On the other hand, *S. parvula* metabolome was characterized with rapid response to salt treatment, dynamically moving from pre-stress state more similar to that of *A. thaliana* to a state converging to that of *E. salsugineum* upon salt treatment. Transcriptomic responses in *S. parvula* supported the observed growth uninterrupted by salt treatment [2], seemingly governed by genes associated with auxin and ABA pathways.

To elucidate the innovations in stress-associated gene function in extremophyte lineages, we compared the GRNs responsive to the stress hormone ABA between the two extremophytes and two stress-sensitive relatives, *A. thaliana* and *Sisymbrium irio* [3]. *S. parvula*, mirroring its uniquely low disturbance of primary growth in the presence of salt [1, 2], showed an enhanced root growth upon ABA treatment, in contrast to the inhibition of growth in the other three species. We characterized the genomic landscape of transcription factor (TF)-binding for a key stress-response TF family, ABA-RESPONSIVE ELEMENT BINDING FACTORS (AREB/ABF), by performing DNA Affinity Purification followed by sequencing (DAP-Seq) for all ABF orthologs in the four target species, as well as comparative RNA-seq analyses of ABA-responses in both root and shoot tissues. By performing Phylogenetically informed Profiling (PiP) and comparative GRN analysis, we found prevalent modifications in both transcriptome ABA responses as well as the AREB/ABF-binding landscapes in the *S. parvula* lineage. For example, expression of orthologs related to translation machinery showed strong attenuation as an early response to ABA in roots, but this response was strikingly absent in *S. parvula*. While the core ABA signaling GRN was relatively conserved, GRNs associated with the growth hormone auxin were highly divergent especially between *S. parvula* and the other species through gain-and-loss of AREB/ABF-binding. We identified *TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS 1* (*TAA1*) as a key candidate whose ortholog in *S. parvula* showed a modification in ABA-responsive GRN, which may be at least partially responsible for the distinct growth response to ABA treatment. We are currently expanding the comparative analyses to ten Brassicaceae species including the oil crop *Camelina sativa*, and developing pipelines that enable GRN comparisons across multiple species with polyploidy and lineage(s)-specific gene duplications. Our study will provide both genetic resources and strategies for developing more stress-resilient oil crops, as well as identify gene functions that cannot be explored in stress-sensitive model plants.

**References/Publications**


1. Tran KN, Wang G, Oh DH, Larkin JC, Smith AP, **Dassanayake M** (2021) Multiple paths lead to salt tolerance - pre-adaptation vs dynamic responses from two closely related extremophytes. *bioRxiv* preprint https://doi.org/10.1101/2021.10.23.465591
2. Tran KN, Pantha P, Wang G, Kumar N, Wijesinghege C, Hong H, Johnson J, Kelt R, Matherne MG, Clement A, Crain C, Tran D, Oh DH, Adhikari P, Foroozani M, Finnegan P, Longstreth D, Larkin JC, Smith AP, **Dassanayake M** (2021) Balancing growth amidst salinity stress-lifestyle perspectives from the extremophyte model *Schrenkiella parvula*. *bioRxiv* preprint https://doi.org/10.1101/2021.08.27.457575
3. Sun Y, Oh DH, Duan L, Ramachandran P, Ramirez A, Bartlett A, Dassanayake M, **Dinneny JR** (2020) Divergence in a stress regulatory network underlies differential growth control. *bioRxiv* preprint https://doi.org/10.1101/2020.11.18.349449



**Funding statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0020358.

**Title: Cell type specific abiotic stress responses among closely related Brassicaceae species reveal how stress tolerance in extremophytes differ from stress-sensitive models**

**Authors:** Guannan Wang (gwang23@lsu.edu),[1*] Kook Hui Ryu[2], Jiyoung Lee[3], Dong-Ha Oh[1], Chathura Wijesinghege[1], Prashanth Ramachandran[4], Andrea Ramirez[4], Andrea Dinneny[4], **José R Dinneny**[4], Song Li[3], John Schiefelbein[2], Maheshi Dassanayake[1]

**Institutions:** [1]Louisiana State University, Baton Rouge, LA; [2]University of Michigan, Ann Arbor, MI; [3]Virginia Polytechnic Institute and State University, Blacksburg, VA; [4]Stanford University, Stanford, CA

**Project Goals:**
**Objective 1:** Discover how extremophytes and stress sensitive species differ in the cell-type functions of roots and those triggered downstream of ABA. ***(Following abstract is focused on this objective)***
**Objective 2:** Define how changes in the wiring of gene regulatory networks produce innovations in transcriptional regulation in extremophytes and how bioenergy crops have diverged.
**Objective 3:** Establish a data driven, predictive framework for accelerating functional testing of stress resilience genes using Arabidopsis and Camelina as a chassis for engineering.

**Abstract:**
Environmental stress tolerance is a complex trait that involves spatio-temporal regulation of large gene networks. Our current understanding of how plants regulate gene expression to achieve stress tolerance is largely limited to model plants and crops that are highly sensitive to environmental stresses. Wild plants adapted to harsh environments developed as extremophyte models offer a unique molecular genetic resource to investigate how millions of years of natural selection has resulted in innovative ways of achieving stress resilient growth. With ongoing environmental challenges to biofuel focused agriculture often compounded by climate change, new designs of smart crops are critically needed. Improved insight into genetic mechanisms regulating stress responsive gene networks for multiple abiotic stresses is needed as input for such smart crop designs to optimize balance between plant growth and stress survival. The core abiotic stress responsive networks operating at the cell-type level have been determined for *Arabidopsis thaliana* [1, 2]. However, these networks are limited in explaining how stress tolerance limits can be expanded to levels we observe in stress-adapted plants in the wild. Therefore, in our current research, we have used a comparative platform to explore and deduce flexible gene regulatory networks that respond to abiotic stresses across closely related Brassicaceae genomes that includes both extremophytes and their closely related stress-sensitive species [3, 4].

We have selected the extremophyte models, *Eutrema salsugineum* and *Schrenkiella parvula* compared to the stress-sensitive models, *Arabidopsis thaliana* and *Sisymbrium irio* to study single cell transcript profiles of root tissues in response to ABA and salt treatments. We have generated scRNAseq datasets for ~15,000 cells per condition from each species. Individual cells were assigned to ten defined root cell types separated into clusters based on their similarity to reference transcriptome profiles, presence of cell-type marker genes, and their alignment with published root scRNAseq profiles. Differential expression profiles within the same cell type across the species at control conditions allowed us to identify cell type specific basal expression level differences between extremophytes and their stress-sensitive models, undetected with bulk RNAseq data. Surprisingly, orthologs of many well-established *A. thaliana* cell type marker genes did not show the expected preferential expression in specific cell types in the other species. We identified new cell type markers that were shared among all target species previously not documented as marker genes in *A. thaliana* as well as markers that were preferentially expressed in selected cell types only at a species specific level. Ortholog expression between closely related species within an in-group in a clade generally showed higher similarity compared to those outside the clade. However, a significant proportion (~30 - 40%) of ortholog expression profiles did not match the phylogenetic signal supported by the transcript sequence similarity of a common ancestor. Clustering of transcripts into expression

modules across all cell types, conditions, and species revealed that cell identity was better predicted by both positive and negative contributors of expression instead of clusters of expressed marker genes alone. These expression modules that distinguished cell types could be supported by enriched cell-type specific functions such as xylem or endodermis development and highlighted developmental processes that differed among the species and further diverged when examined under stress treatments. These results collectively provide insights into spatio-temporal regulation of stress responsive genes at a single cell resolution unexplored before using extremophytes in a comparative study.

**References/Publications**


1. Dinneny, J.R., Long, T.A., Wang, J.Y., Jung, J.W., Mace, D., Pointer, S., Barron, C., Brady, S.M., Schiefelbein, J., Benfey, P.N. Cell identity mediates the response of Arabidopsis roots to abiotic stress. (2008). Science,320 (5878), pp. 942-945.
2. Ryu KH, Huang L, Kang HM, Schiefelbein J. Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. (2019). Plant Physiology, 179(4):1444-1456. https://doi.org/10.1104/pp.18.01482
3. Tran KN, Wang G, Oh DH, Larkin JC, Smith AP, Dassanayake M. Multiple paths lead to salt tolerance - pre-adaptation vs dynamic responses from two closely related extremophytes. (2021) bioRxiv preprint https://doi.org/10.1101/2021.10.23.465591
4. Sun Y, Oh DH, Duan L, Ramachandran P, Ramirez A, Bartlett A, Dassanayake M, Dinneny JR. Divergence in a stress regulatory network underlies differential growth control. (2020) bioRxiv preprint https://doi.org/10.1101/2020.11.18.349449



**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0020358.

**Plant-Microbe Interfaces: Development of an experimental approach to achieve spatially resolved plant root-associated metaproteomics using an agar-plate system**

Manasa R. Appidi[1,2*] (appidim@ornl.gov), Amber N. Bible,[1] Sara S. Jawdy,[1] Dana L. Carper,[1] Richard J. Giannone,[1] Robert L. Hettich,[1] Jennifer Morrell-Falvey,[1] Paul E. Abraham,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; and [2]Genome Science and Technology graduate program, University of Tennessee, Knoxville, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

The plant-microbe interface is a region where microbes respond to and interact with their plant hosts. Plants secrete root exudates that shape the microbiome, yet it is poorly understood how microbiome functions change with respect to space and time across plant roots. Methodologies that facilitate spatio-temporal metaproteomic studies of root-associated microbiomes are yet to be realized. To address these challenges, we developed a novel extraction method that facilitates spatial characterization of microbiomes with bottom-up proteomics using a plant-agar culture system under controlled laboratory conditions. Spatially defined agar 'plugs' of interest near or distant from the plant roots were excised and subsequently processed using a novel peptide extraction method that uses magnetic microparticle beads for agar removal from the samples allowing for metaproteomic measurements. As a proof-of-principle, a previously studied 10-member bacterial community constructed from a *Populus* root system was grown in an agar-plate with a 3-week-old *P. trichocarpa* plant. High-resolution metaproteomic measurements that are used to infer both bacterial community composition and function were obtained across two time points (24-h and 48-h) for three distinct agar plugs collected at the root base, root tip, and a region distant from the root. The spatial resolution of these measurements provide evidence that the root-associated microbiome structure and proteome expression changes across the plant root interface. Interrogation of the individual bacterial proteomes revealed functional profiles related to their behavioral associations with the plant root, in which chemotaxis and augmented metabolism likely supported predominance of the most abundant member. Obtaining spatio-temporally resolved microbial communities along different regions of the plant root was enabled using this novel peptide extraction method, which is a critical step towards understanding how plants select for or

against specific microbes. More broadly, we anticipate this novel extraction approach will benefit other soft-agar plate studies that study nutrient exchange among microbes.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Temporal dynamics of the *Populus* microbiome across scales**

Nicholas C. Dove[1]* (dovenc@ornl.gov), Alyssa A. Carrell,[1] Dawn M. Klingeman,[1] Nancy L. Engle,[1] Timothy J. Tschaplinski,[1] Melissa A. Cregger,[1,2] and **Christopher W. Schadt**[1,3]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN; and [3]Department Microbiology, University of Tennessee, Knoxville, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Our understanding of the plant microbiome is clouded by the fact that the majority of studies on the plant microbiome represent "snapshots", as they present data from a single time point. It is well known, however, that microbiomes are temporally-dynamic resulting from external forcing factors and intra-community interactions. Hence, a predictive understanding of the relationship between the plant and its microbiome, and the ways these events manifest themselves across the various temporal scales relevant to natural systems, is a challenge that requires long-term fundamental research. Using amplicon and metagenomic sequencing of the plant microbiome in combination with metabolomic measurements, we are characterizing how the *Populus* microbiome changes over time in three projects. Studying temporal changes in the microbiome of a long-lived plant species, such as *Populus*, can give us unique insights into the ecological processes shaping microbiomes when compared to annual plant species (e.g., *Arabidopsis*, *Maize*, etc.) that often serve as models plant-microbe interaction studies.

The first project leverages a multi-year common garden experiment planted with 10 genotypes of *P. deltoides* and *P. trichocarpa* with varying degrees of disease resistance. In 2018 and 2019, we collected soils, roots, and leaves from these genotype four times each year representing the major seasons to determine intra- and inter-annual variations in in the *Populus* microbiome. We found that in the rhizosphere, the microbial community changed both seasonally and annually, suggesting a non-cyclical pattern of rhizosphere microbiome composition. However, while the taxonomic composition of ectomycorrhizal (EM) fungi (important plant symbionts) changed, the relative abundance of fungi followed a cyclical pattern and was highest in the spring. Functionally, there was a greater genetic potential for nitrate and nitrite assimilation over time regardless of the season in the rhizosphere, suggesting that these microbial populations are rapidly growing. Analysis for of leaf and root microbiomes is currently underway. We hypothesize that these endospheric microbiomes will vary with time, but to a lesser extent than the rhizosphere due to relatively stronger plant genetic effects.

The second project further characterizes how the broader *Populus* holobiome is moderated following infection by *Sphaerulina musiva*, a well-characterized fungal pathogen, which infects *P. trichocarpa* and *P. trichocarpa x deltoides* hybrids. For this project, we investigate the microbiome and metabolome of *Populus* leaves and roots and the microbiome of their associated surfaces (to understand the impact *S. musiva* abundance (quantified by qPCR) and infection systematically in the same common garden field setting as above. We found that *S. musiva* is present in all trees and tissues, but *S. musiva* abundance as measured by qPCR was unrelated to stem canker onset and development. We also find that the leaf and root metabolomes significantly differ between the two *Populus* species and that certain leaf metabolites, particularly the phenolic glycosides salirepin and salireposide, are diminished in canker-infected *P. trichocarpa* trees compared to their uninfected counterparts. Furthermore, we found significant associations between the metabolome, *S. musiva* abundance, and microbiome composition, particularly in *P. trichocarpa* leaves. This suggests that the effects of *S. musiva* on *P. trichocarpa* trees are widespread and not confined to the site of canker infection.

The third project extends the temporal scale of our previous work to the lifespan of a tree. To accomplish this, we are taking advantage of the availability of large (10s to 100s of acres) clonal stands of *P. tremuloides* in the intermountain west, including the grove nicknamed "Pando", which is one of the largest organisms on earth. By using defined chronosequences within these clonal stands, we are able to isolate the effects of tree (ramet) age, from the soil and genotypic factors that we know from previous work also exert large controls on microbiome composition. During the summer of 2021, we collected leaf, stem and root tissues as well as associated soils from 140 trees/ramets representing four aspen clones, including Pando, ranging in age from 1 to 140 years old. Microbiome and metabolomic analysis is currently underway. We hypothesize that differences in the microbiome with tree age will attenuate in older trees in concert with changes in the metabolome. Furthermore, we hypothesize that microbiome and metabolomic differences will be greater in older tissues (e.g., heartwood) compared to younger tissues that regenerate each year (e.g., leaves).

By integrating these three projects we are gaining an understanding of the temporal dynamics in the *Populus* microbiome from seasons to centuries (Figure 1).

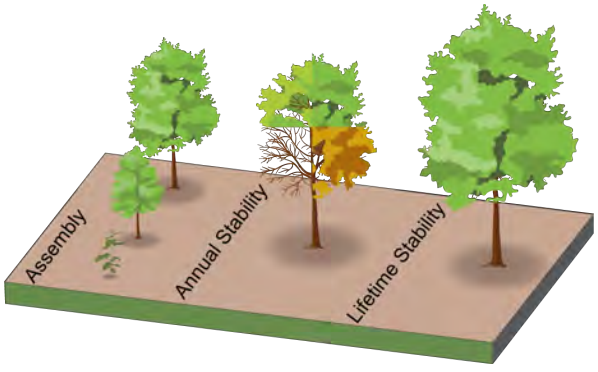



Figure 1: Temporal scales of the *Populus* microbiome.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Plant mechanisms shaping poplar root microbiota**

Felix Fracchia[1*] (Felix.Fracchia@inrae.fr), Nancy Engle,[2] Timothy Tschaplinski,[2] Francis Martin,[1] Aurélie Deveau,[1] and **Claire Veneault-Fourrey**[1]

[1]Université de Lorraine, INRAE, IAM, Nancy, France; and [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Trees interact with complex microbial communities that contribute to their extended phenotype. How early fungal and bacterial root colonization occur and how these interactions take place remain poorly understood. The woody perennial model *Populus* has the ability to interact with distinct trophic guilds, such as endophytes and mycorrhizal fungi, and can help decipher the organization of the early establishment of microbial root communities. We determined that the early establishment of the fungal and bacterial root communities of axenic *Populus tremula x alba* transplanted in environmental soil is dynamic in time and space over 50 days of culture. This microbial colonization involves three successional waves of colonization, for both bacterial and fungal community structures, harboring distinct trophic abilities. Even though bacterial community structures were more stable in time than the fungal community structures, the early colonizers were saprotrophs, fast-growing species that were progressively replaced by endophytes and mycorrhizal fungi. The composition of root exudates and host-based selection through distinct hormonal and metabolic profiles could explain the evolution of these microbial communities. To test these hypothesis, transgenic poplar trees altered in ethylene perception and biosynthesis were grown in sterile condition, transplanted into environmental soil and grown for 30 days in natural soil. We examined fungal root colonization using confocal laser scanning microscopy and combined this analysis with metabarcoding and metabolomics of roots and their exudates. Microbial enzymatic activities associated with rhizospheric soil differed depending on the poplar transgenic and wild type lines. Taken together, these results support the importance of plant genotype in the structuring of microbial communities and their importance in soil functional traits.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific*

*Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Network integration uncovers gene-targets involved in plant-microbial interactions**

Piet Jones[1]*(jonespc@ornl.gov), J. Izaak Miller,[2] David Kainer,[2] Anna Furches,[1,2] Manesh Shah,[2] Wellington Muchero,[2] Jay Chen,[2] Doug Hyatt,[2] Mirko Pavicic,[2] Hari B. Chhetri,[2] Dan Jacobson,[2] and **Mitchel J. Doktycz**[2]

[1]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN; and [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

The interactions between plants and their constituent microbiomes are complex. Thus, it is important to delineate potential host factors that may shape the composition of the microbiome to bring us one step closer to understanding this complex set of interactions. Here, we investigate the relationship between microbial relative abundance and host genomic variation within *Populus trichocarpa* for both leaf and xylem tissues. Leveraging a kmer-based pipeline to extract putative taxa from bulk total-RNA samples, we identified hundreds of taxa. We focused on four leaf and four xylem genera-level taxa that overlap with the PMI culture collection. We observed *Acinetobacter*, *Bradyrhizobium*, *Enterobacter*, and *Pantoea* in xylem tissue; and *Bacillus*, *Chryseobacterium*, *Paenibacillus*, and *Streptomyces* in leaf tissue. The relative abundance of each of these taxa were then used as phenotypes in a genome wide association study (GWAS). Significant single nucleotide polymorphism associations are annotated with nearby genes. Utilizing novel network integration approaches that leverage extant knowledge from multiple 'omic data sources, called random walk with restart filter (RWR-filter) and random walk with restart lines of evidence (RWR-LOE), we refine the GWAS gene set to obtain a set of high confidence genes with biological context for each respective phenotype. We used gene annotation and network topological context around the respective genes to create a set of conceptual models that provide hypotheses to explore for future CRISPR/Cas9 experiments. In xylem, by integrating the annotations of the gene sets for each taxa to identify associations present in two or more phenotypes, we find a putative signal transduction cascade. In particular, we find associations to members of various transcription factor families, genes involved in signaling, transport, cell-

division, cell-wall cellulose synthesis, the abscisic acid phytohormone pathway, and post-translational modification involving ubiquitin protein degradation. In addition, our analysis allows for investigation of the host-factors associated with individual phenotypes. For the *Chryseobacterium* found in leaf tissue, for example, we find associations to genes involved in the auxin and salicylic-acid phytohormone pathways, abiotic stress, flavonoid biosynthesis, nitrogen metabolism, and protein degradation. Together our approach allows for a thorough investigation of the host-microbial interactions, and thereby, identifies target host genes that may play an important role in shaping the microbial community.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Determining the rate and consequences of horizontal gene transfer in the rhizosphere by simulating lateral spread of salicylate catabolism genes**

Stephan Christel[1], Joshua K. Michener[1]* (michenerjk@ornl.gov), Paul E. Abraham,[1] Alyssa A. Carrell,[1] Sara S. Jawdy,[1] Tao Yao,[1] Julie E. Chavez,[1] Leah Burdick,[1] Dawn M. Klingeman,[1] Timothy J. Tschaplinski,[1] Jin-Gui Chen,[1] Wellington Muchero,[1] Melissa A. Cregger,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Plants affect the composition of the rhizosphere microbiome through the secretion of biomolecules. For example, *Populus* trees produce high levels of salicylic acid conjugates that are thought to play a role in structuring its microbiome, inhibiting the growth of some microbes, while serving as a carbon source to others. However, rhizosphere microbes are continually evolving to maximize their own fitness, perhaps changing the growth effects of compounds like salicylates. Horizontal gene transfer (HGT) is one of the main drivers of prokaryotic diversity and is hypothesized to play a significant role in how microbes compete in the rhizosphere and evade host controls. In turn, mutations in the plant host can change the abundance and composition of secreted small molecules. The functional consequences of these dynamic evolutionary interactions are key to understanding and predicting the stability of plant-microbe interactions in the rhizosphere.

To assess the potential consequences of successful HGT events in the *Populus* rhizosphere, *Pseudomonas* isolates from *Populus* were engineered to express a pathway to metabolize salicyl alcohol. These modified strains successfully used salicyl alcohol and salicylic acid as the sole carbon and energy source. Global proteomic measurements showed minimal disruption to the native physiology after pathway acquisition, suggesting few barriers to pathway transfer in the rhizosphere. To test the effects on abundance and localization in the rhizosphere, genomically barcoded populations of wild-type and engineered strains from one isolate, *Pseudomonas* sp. GM17, were inoculated onto otherwise sterile plant roots. In the absence of synergistic strains and on roots of unaltered *P. trichocarpa* BESC819, DNA barcode sequencing revealed no discernable effect of acquiring the salicyl alcohol catabolic pathway.

However, concentrations of salicyl alcohol and salicylic acid are low in native *Populus* roots since these aromatic compounds are generally secreted as glycosylated salicylates. Therefore, we tested whether the fitness effects of salicyl alcohol catabolism depend on epistatic interactions with the host or with other members of the rhizosphere microbiome. Experiments were conducted with *Rahnella* sp. OV744, a species that increases availability of salicyl alcohol by deglycosylating salicin, and with a genetically-modified *Populus* variant overexpressing XBAT35. The roots of the XBAT35 line contain high concentrations of salicylic acid-related conjugates, including a 5- to 15-fold increase in tremuloidin and a 3- to 7-fold increase in salicin. We expect that the effect of catabolic pathway acquisition will be increased when its substrate is more prevalent. These experiments will aid in understanding how plant-microbe-community interactions modulate microbial composition, localization, and ultimately function.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Investigating how interactions and local environments in the rhizosphere influence microbial responses and behavior**

Jennifer Morrell-Falvey[1*] (morrelljl1@ornl.gov), Amber N. Bible,[1] Manasa Appidi,[2] Paul Abraham,[1] Sara Jawdy,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; and [2]Genome Science and Technology graduate program, University of Tennessee, Knoxville, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

The rhizosphere is a dynamic and heterogeneous environment and the behavior of organisms within the rhizosphere depends on local environmental conditions and spatial organization. Developing methods to monitor community organization and map local chemical and physical environments and to correlate these data with microbial responses are necessary to understand, and ultimately predict, microbiome function. To this end, we have visualized the dynamics of plant-microbe interactions using engineered habitats and found that many plant-associated bacteria show reproducible root colonization patterns. We hypothesize that differing chemical profiles around the root contribute to the recruitment and growth of bacteria at these preferred sites. To better understand how the chemical environment around the root influences microbial responses and behavior, we examined proteome profiles of poplar-associated bacterial strains grown in the presence of a poplar cutting. For these studies, we examined the proteome profiles of individual bacterial strains grown in the presence of a plant compared to the proteome profiles when all ten strains were grown as a synthetic community in the presence of the plant. Common to all strains were significant changes in the proteome profiles involving proteins predicted to be involved in nutrient transport and metabolism, sensing, and motility. Current efforts are underway to test hypotheses derived from these proteomic experiments in *Pantoea* sp. YR343 using mutant strains, growth assays, and plant colonization experiments.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Modulation of microbial community dynamics by contrasting regulation of salicylic acid and jasmomic acid-ethylene signaling pathways.**

Meng Xie,[1] Kuntal De,[1] Debjani Pal,[1] Alyssa A. Carrell,[1] Sara S. Jawdy,[1] Tao Yao,[1] Leah Burdick,[1] Dawn M. Klingeman,[1] Paul E. Abraham,[1] Timothy J. Tschaplinski,[1] Jin-Gui Chen,[1] Wellington Muchero[1*] (mucherow@ornl.gov), and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Interactions between salicylic acid (SA) and jasmonic acid-ethylene (JA/ET) signaling pathways have been found to be important to activate the correct defense responses to biotrophic and necrotrophic pathogens. The understanding of the molecular regulation of SA and ET/JA signaling interactions is essential to rationally improve plant responses against different pathogens or in recruitment of beneficial microbes. Molecular studies focusing on the crosstalk between SA and ET/JA signaling have demonstrated the critical regulatory role of transcriptional mechanisms, however the exact mechanism remains poorly understood. Using *Arabidopsis* knockout and overexpression mutants, we showed that one nuclear protein TYROSYL-DNA PHOSPHODIESTERASE 1 (TDP1) physically interacts with ANGUSTIFOLIA C-terminus Binding Protein (AN) and enhances the nuclear accumulation of AN. In addition, AN was found to directly target and repress the transcription of *MYB46*, which encodes a master regulator of the phenylpropanoid pathway. Moreover, AN displayed the capability of releasing transcriptional repression on *WRKY33* by altering chromatin association of TDP1, which negatively affects *WRKY33* expression. The transcriptional effects of AN on the expression of *MYB46*, *WRKY33*, and their downstream genes were validated in *an* knockout mutant and *AN* overexpression plants. The antagonistic transcriptional regulation of *MYB46* and *WRKY33* by AN suggests a transcriptional co-regulation mechanism of SA and ET/JA signaling. Consistent with this notion, the alteration of *AN* expression in transgenic plants was shown to oppositely affect defenses against the biotrophic pathogen *Pseudomonas syringae* and the necrotrophic pathogen *Botrytis cinerea*. Our results demonstrate a transcriptional co-regulatory mechanism of the crosstalk of SA and ET/JA signaling and suggest that it may play a key role in the coordination of defense responses towards biotrophic and necrotrophic pathogens. Importantly, changes observed in defense responses also affected biomass accumulation suggesting a trait-tradeoff between defense and

growth. Work has now been initiated to establish the implication of these genetic modifications on microbial community dynamics.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-microbe interfaces: Interrogating poplar fungal microbiome interactions using meta-transcriptomics and constructed communities**

Jake Nash[2]* (jan58@duke.edu), Keaton Tremble,[4] Bryn Dentinger,[4] Brian Looney,[2] Corbin Bryan,[2] Khalid Hameed,[2] Yi-Hong Ke,[2] Lotus Lofgren,[2] Melissa Cregger,[1] Nicholas Dove,[1] Christopher Schadt,[1] Tomás Rush,[1] Igor Grigoriev,[3] Vivian Ng,[3] Deborah Wendt,[3] Francis Martin,[5] Claire Veneault-Fourrey,[5] and **Rytas Vilgalys[3]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Duke University, Durham, NC; [3]JGI, Berkeley, CA; [4]Univ. of Utah, Salt Lake City, UT; and [5]INRAE IAM, Nancy, Champenoux, France

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Poplar trees (genus *Populus*) are host to diverse root fungal microbiomes that include ectomycorrhizal, arbuscular mycorrhizal, and endophytic fungi. These fungi perform services for the plant host including growth promotion, nutrient acquisition, protection from pathogens, and conferral of abiotic stress tolerance. Meta-transcriptomics can provide large amounts of data on the function and taxonomic composition of the poplar root fungal microbiome. We developed an RNA-seq method using a synthetic spike-in standard curve that allows for the calculation of absolute abundances of fungal transcripts on poplar roots. We implemented a bioinformatics workflow that provides taxonomic and functional annotations of assembled fungal contigs from meta-transcriptomic data. These methods were applied to an ecosystem-scale, time-series field experiment to document taxonomic and functional shifts of the poplar fungal microbiome in response to a historic drought in the semi-arid American West during the summer of 2021. We identified transcripts from a previously isolated dark septate endophyte in the genus *Hyaloscypha* as a highly active root colonizer across our field sites. Dark septate endophytes are a functionally diverse group of root associates that have been described as either mutualists, commensalists, or latent pathogens. We conducted further work to understand the characteristics of the *Populus*-*Hyaloscypha* association. *In vitro* inoculations with this fungus demonstrated compatibility with both *Pinus* and *Populus*, suggesting that it engages in antagonistic interactions with arbuscular mycorrhizal fungi during plant host colonization. We were also able to establish simplified constructed communities with this fungus and three common ectomycorrhizal fungi, ranging in diversity from one to four species. These constructed communities will allow us to identify interactions between fungi during root colonization and evaluate the effects of fungal diversity on plant performance and nutrient uptake. Future work will also 1) dissect the molecular mechanisms of the antagonistic interaction with arbuscular mycorrhizal fungi, 2) evaluate the ability of this fungus to confer drought tolerance to *Populus*, and 3) identify common and unique symbiosis-induced genes when colonizing different plant hosts.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Quantification of *Populus* transcriptomic response to colonization by select bacterial symbionts.**

Bryan Piatkowski[1]* (piatkowskibt@ornl.gov), Xiaohan Yang,[1] Lead H. Burdick,[1] Sara S. Jawdy,[1] Amy Schaefer,[2] Yasuhiro Oda,[2] Caroline Harwood,[2] Dale A. Pelletier,[1] David J. Weston,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; and [2]Department of Microbiology, University of Washington, Seattle, WA

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Poplar species (*Populus* spp.) associate with a diverse array of bacteria that are selected from the environment. However, the molecular mechanisms by which poplars recognize and establish symbioses with these microbes are largely unknown. To determine how these plants respond to colonization with bacteria, we inoculated germ-free *P. trichocarpa* root tissue with select bacterial strains isolated from field-collected roots and quantified plant gene expression using RNAseq. Colony forming units (CFU) were quantified for each bacterial strain and ranged over three orders of magnitude, demonstrating wide variation in the ability to colonize *Populus* root tissues. The number of differentially expressed poplar genes also varied among treatments but did not correlate with CFU count. Plant transcriptomic responses were largely strain-specific with no single gene induced or repressed among all treatments. Across all genes that showed differential expression relative to control tissues, those induced were enriched for biological processes including cell wall biosynthesis, vesicle trafficking, and root development while those repressed were enriched for processes including stress response and carbon metabolism. Comparisons between plants inoculated with pathogenic (*Pseudomonas syringae* isolate NP28-5) and non-pathogenic strains (including *Pseudomonas fluorescens* isolate GM79) revealed potential mechanisms by which poplar defends against pathogens, including the induction of genes coding for anti-microbial peptides and pathogen-associated microbial pattern (PAMP) receptors. Inoculation with *P. fluorescens* isolate GM79 compared to a knock-out strain missing the microbial gene pipA, involved in bacterial recognition of a plant-derived effector molecule, elicited the induction of ten poplar genes largely involved in transmembrane transport and signal transduction. Together, these data provide insight into the molecular mechanisms involved in the establishment of symbiosis between *Populus* and its microbiome. Future directions include qPCR validation of candidate loci

and leveraging metabolomic, proteomic, and network-based approaches to further interrogate poplar response to select bacteria.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Towards a rhizosphere on a chip for understanding physical and chemical transitions in multi-kingdom systems**

Scott T. Retterer[2]* (rettererst@ornl.gov), Muneeba Khalid,[1] Yi-Syuan Guo,[2] Amber B. Webb[1], Sara Jawdy,[1] John F. Cahill,[1] Courtney L. Walton,[1] Vilmos Kertesz,[1] Julian A. Liber,[3] Gregory Bonito,[4] Jessy L Labbé,[1,5] Jennifer Morrell-Falvey,[1] Dale A. Pelletier,[1] and **Mitchel J. Doktycz[1]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN; Duke University, Durham, NC; [4]Department of Plant Soil and Microbial Sciences, Michigan State University, East Lansing, MI; and [5] Invaio Sciences, Cambridge, MA (current address).

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

The mechanisms by which plants, bacteria, fungi and other microorganisms negotiate spatial networks defined by the soil to ultimately shape and define interactions in the rhizosphere are multifaceted and complex. It is clear that a combination of physical and chemical interactions affect movement through such networks, but their relative importance and the molecular signals that dominate such behaviors are yet to be elucidated. Microfluidics, as engineered habitats, provide a path for experimentally teasing apart the answers to such questions, and provide a means by which scientists can carefully explore the impact of environmental structure and composition on multi-kingdom interactions, niche establishment, and complex community dynamics. In effect, tractable systems that combine known quantities of community members in well-defined initial conditions, can be observed and quantified over extended periods of time to track the position and activity of different biological species, while simultaneously mapping the chemical composition of the environment in which they are interacting. In this study, we have begun to map and image the chemical environment of *Populus* cuttings grown over extended periods of time within engineered habitats that mimic soil structure in two dimensions. These rhizosphere-on-a-chip platforms provide optical access to capture fine morphological changes and growth in living plant and microbial systems. This enables local and global assessments of structure and chemical composition. In complementary efforts, we have refined methodology and completed the baseline assessment of fungal growth and hyphal elongation rates in microfluidic networks with varied levels of confinement and network complexity. Distinct differences in growth rates, branching,

penetration potential, and exploration motifs were evident across the species that were examined and quantified.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Beyond symbiosis, fungi can specifically perceive and use lipochitooligosaccharides to organize and modulate the development of microbial communities**

Erica Teixeria Prates,[1] Tomás A. Rush[1]*(rushta@ornl.gov), Omar Demerdash,[1] Manuel I. Villalobos Solis,[1] Joanna Tannous,[1] Muralikrishnan Gopalakrishnan Meena,[2] Matthew J. Lane,[3] Manesh Shah,[3] Nancy L. Engle,[1] Alyssa A. Carrell,[1] Margaret K. Spangler,[1] Jean-Michel Ané,[4,5] Nancy Keller,[4,6] Sylvain Cottaz,[7] Sébastien Fort,[7] David Kainer,[1] Dale A. Pelletier,[1] Tim Tschaplinski,[1] Robert L. Hettich,[1] Richard J. Giannone,[1] Paul E. Abraham,[1] Jesse L. Labbé,[1,#] Dan A. Jacobson,[1] and **Mitchel J. Doktycz[1]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2] National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN; [3] Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN; [4] Department of Bacteriology, University of Wisconsin-Madison, Madison, WI; [5] Department of Agronomy, University of Wisconsin-Madison, Madison, WI; [6] Department of Medical Microbiology and Immunology, University of Wisconsin-Madison, Madison, WI; [7] Université Grenoble Alpes, CNRS, CERMAV, Grenoble, France; and [#]Invaio sciences, Cambridge, MA (current address)

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Lipochitooligosaccharides (LCOs) are signaling molecules produced by rhizobial bacteria and fungi. The molecular mechanisms of perception of LCOs by host plants are fairly well characterized in their role as symbiotic molecules. Lysin motif receptor-like kinases (LysM-RLKs) are identified as the LCO-receptors in various plants, and this interaction elicits phenotypes like oscillations in nuclear calcium, and promote root hair branching, or curling in legumes. LCOs are produced by most fungi, including plant and opportunistic human pathogens. However, the role of LCOs outside of symbiosis with plants remains unknown. We test for alternative roles of LCOs to address this question. We examined two species of fungi, namely, *Aspergillus fumigatus*, a soilborne fungus that can be an opportunistic human pathogen, and *Laccaria bicolor*, a symbiont of poplar. We determined significant changes in fungal behavior in response to exogenous applications of various types of LCOs, including changes in fungal physiology, metabolomics output, and differential regulation of genes and proteins. Both fungi displayed fewer hyphae branching formation, increased sexual propagation, and delayed growth, indicating that LCOs could be fungistatic compounds. Moreover, upon exposure to LCOs, early transcriptomic and

proteomic changes are observed, as well as the regulation of secreted metabolites that can inhibit or promote the growth of specific bacteria known to inhabit the poplar rhizosphere. Lastly, structural modeling and analysis identified fungal LysM proteins that carry key structural features of known plant LCO-receptors. We describe a molecular dynamics and machine learning-based workflow to predict the relative binding affinity of LysM-LCO complexes. We demonstrate its remarkably high accuracy using validation sets, and therefore, its promising application to identify LysM proteins that can effectively bind to LCOs in fungi and other organisms. Overall, we propose that LCOs are fungistatic compounds produced and used by fungi to organize microbial communities. We hypothesize that these signaling molecules produced by some microbes may be sensed by multiple organisms via membrane-attached LysM proteins. Therefore, the role of LCOs seems to greatly surpass microbe-host plant symbiosis as a trans-kingdom communication signal.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Bio-Scales: Defining gene function and its connection to ecosystem processes**

Christopher W. Schadt,[1*] (schadtcw@ornl.gov), Melanie A. Mayes,[2] Priya Ranjan,[1] Brandon D. Kristy,[1] Alyssa Carrell,[1] Dawn Klingeman,[1] Sara Jawdy,[1] Wellington Muchero,[1] David J. Weston,[1] Stanton L. Martin,[1] J. Christopher Ellis,[1] Daniel A. Jacobson,[1] Julie Mitchell,[1] **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; and [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

https://www.ornl.gov/content/bio-scales-0

**Project Goals: The Bio-Scales project focuses on understanding how genes influence traits and ecosystem-level processes. Our initial use case examines specific combinations of host and microbial traits that influence nitrogen transformation patterns and fluxes across the coupled plant-soil-microbial system and consists of two objectives. In our first objective, we are sampling a subset of *Populus* genotypes with differential phenotypes related to N transformation and biomass production, characterizing their associated microbiomes using multi-omic approaches (in collaboration with the Joint Genome Institute), and collecting extensive environmental metadata. An important second objective is to prepare data and metadata for integration and analysis using the National Microbiome Data Collaborative, KBase and other tools and approaches (e.g., GWAS, network analysis).**

The Bio-Scales project aims to rapidly determine gene functions and traits and how they scale to influence ecosystem-level processes. A current use case hypothesizes that specific combinations of plant and microbial traits influence rhizosphere elemental cycling patterns and fluxes across the coupled plant-soil-microbial system, and that these transformations will have important ramifications for nutrient cycling and availability within the rhizosphere. Key plant chemical traits that influence the plant microbiome, rhizosphere biogeochemical processes, and subsequent rhizosphere nitrogen cycling are being examined. Using *Populus* as a model host system, we are assessing key plant genotypic traits known to influence nitrifying and denitrifying microbial activity across a genome-wide association mapping study population. The three plant chemotypic traits initially targeted involve production of alpha-linolenic acid, para-coumeric acid and ferulic acid, that are hypothesized to influence rhizospheric microbial communities and activities leading to altered N cycle transformations. Target plant genotypes with high and low expression levels for these secondary metabolites, as well as rhizosphere and bulk soils, were collected in Oregon in the fall of 2020 from two field study sites near Corvallis and Clatskanie Oregon. Metagenomic analyses of soil and root microbiomes, plant transcriptomics, and metabolomics datasets are being generated in collaboration with the Joint Genome Institute (JGI) to be integrated together with soil environmental metadata and used to identify the effects of these host-microbiome-environment relationships. Metagenomic and biogeochemical data on nitrification and denitrification capacity of the soils collected and analyzed to date show that the effects of tree genotypic and chemotypic *Populus* traits scale differently across the two study sites analyzed, and that these data are related to overall nitrogen pools, nitrification and denitrification potential, and overall site soil characteristics.

The data and metadata resulting from the use case above have been developed in collaboration with the National Microbiome Data Collaborative (NMDC) to generate a highly integrated microbiome data set with rich and validated metadata. This high-dimensional dataset will: formalize how to collect and curate field microbiome data related to nutrient cycling; provide a dataset from which to build predictive models of gene function within a plant microbiome; and illustrate the Bio-Scales concept of determining how gene functions scales across biological systems and influences broader ecosystem properties and processes.

Our next step in Bio-Scales will integrate data from the Advanced Plant Phenotyping Laboratory (APPL) and continue our focus on rigorous data management. APPL is capable of growing up to 500 plants on a conveyor system that advances individual plants through a suite of imaging technologies. These images include a 3D point cloud along with 2D chlorophyll fluorescence imaging, hyperspectral imaging and more. The hyperspectral images are known to vary by plant chemotype, and our goal will be to observe chemotypic differences among plants based on their chemotype and nitrogen uptake ability. The APPL Public Interface (or APPL-PI) will be developed to host data for our Bio-Scales use case and other APPL experiments.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKPA12.

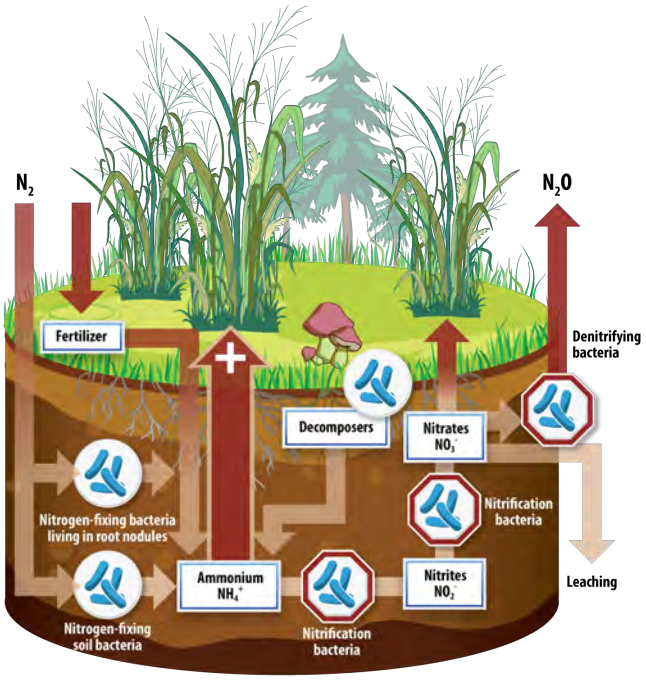



**Fig 1.** Conceptual illustration of the Bio-Scales concept to understand how plant chemotypic traits may influence rhizosphere biological nitrification and denitrification processes and microbiomes, and how these in turn may feedback to affect plant N availability and other critical ecosystem N cycle processes.

**Plant-Microbe Interfaces: Identification of genes controlled by a plant-responsive transcription factor in the *Populus* endophyte *Pseudomonas* GM79**

Amy L. Schaefer[1*](amyschae@uw.edu), Tingting Jin,[2] Yasuhiro Oda,[1] Jiahong Ren,[2] Robert L. Hettich,[3] Paul E. Abraham,[3] Dale A. Pelletier,[3] **E. Peter Greenberg**,[1] **Caroline S. Harwood**,[1] and **Mitchel J. Doktycz**[3]

[1]Department of Microbiology, University of Washington, Seattle, WA; [2]Faculty of Biological Science and Technology, Changzhi University, Shanxi, China; and [3]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

A number of plant-associated proteobacteria, including members of the *Populus* microbiome, have LuxR-family transcription factors that we refer to as PipR subfamily members. PipR proteins play roles in interactions between bacteria and their plant hosts, and some are important for bacterial virulence in plants. We identified a plant-derived ethanolamine derivative, *N*-(2-hydroxyethyl)-2-(2-hydroxyethylamino) acetamide (HEHEAA), as a potent effector of PipR-mediated gene regulation in the *Populus* root endophyte *Pseudomonas* GM79. To better understand which genes are controlled by PipR in GM79, we performed RNAseq transcriptomic analyses and identified a regulon of approximately 40 genes. The positively regulated genes (n=13) included those encoding N-terminal peptidases, a HEHEAA signal transporter, and a two-component regulator, which is inactive under laboratory growth conditions. Surprisingly, the majority of PipR-controlled genes exhibited lower gene expression in the presence of HEHEAA. The PipR-HEHEAA repressed genes included those involved in hydrogen cyanide production and anaerobic respiration on nitrate. Additional experimentation confirmed PipR-HEHEAA regulation of cyanide production and nitrate respiration. We were surprised to find that the influence of PipR-HEHEAA on cyanide production and nitrate respiration was mediated solely through the activity of the PipR-controlled peptidase, PipA. We are attempting to understand the mechanism of this regulatory pathway through proteomics experiments.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: Identification of PtLecRLK1-based signaling cascade in *Laccaria bicolor* root colonization**

Tao Yao[1]* (yaot@ornl.gov), Jay Chen,[1] Zhenzhen Qiao,[1] Yali Sun,[1] Him K. Shrestha,[1,2] Timothy B. Yates,[1,3] Nancy L. Engle,[1] Paul E. Abraham,[1] Robert L. Hettich,[1] Timothy J. Tschaplinski,[1] Wellington Muchero,[1] and **Mitchel J. Doktycz[1]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN; and [3]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Soil-borne microbes can establish mutualistic relationships with host plants, providing a large variety of nutritive and protective compounds in exchange for photosynthesized sugars. However, the molecular events mediating the establishment of these beneficial relationships have yet to be fully characterized. Our previous genetic mapping and whole-genome resequencing studies identified a gene deletion event of a lectin receptor-like kinase gene *PtLecRLK1* that is associated with differential root colonization by the ectomycorrhizal fungus *Laccaria bicolor* among different *Populus* species. We introduced *PtLecRLK1* into the model annual plant *Arabidopsis* and the model perennial plant switchgrass (*Panicum virgatum*), subsequently converting these non-host plants to host plants, allowing colonization by *L. bicolor.* These results have established PtLecRLK1 as a key regulator of *L. bicolor* colonization. Among all proteins currently identified thus far as regulators of *Populus*-*L. bicolor* interactions, PtLecRLK1 is the most promising receptor candidate, responsible for perceiving and transducing signals from *L. bicolor*, which leads to molecular and physiological responses required for root colonization in the host plant. We wanted to define the molecular mechanism of action of PtLecRLK1 in *L. bicolor* root colonization. We hypothesize that PtLecRLK1 perceives signals from *L. bicolor* resulting in phosphorylated downstream components. We applied phospho-proteomics to identify proteins with differential abundance between *L. bicolor*-inoculated and un-inoculated switchgrass *PtLecRLK1* transgenic plants. We have identified several promising phosphorylation targets, including a leucine-rich receptor-like kinase, a cGMP-dependent protein kinase, and a splicing factor. Biochemical assays

are being conducted to validate protein-protein and protein phosphorylation between PtLecRLK1 and these candidate targets. Currently, we are taking a similar phospho-proteomics approach by using recently generated *Populus PtLecRLK1* transgenic plants. We will further examine the biological significance of these protein-protein interactions and protein phosphorylation through genetic validation. Collectively, our studies will help construct the entire PtLecRLK1-based signaling cascade that is responsible for specific molecular and physiological responses leading to *L. bicolor* root colonization, advancing our fundamental understanding of the molecular mechanism underlying the selection and maintenance of a mutualistic relationship between *Populus* and *L. bicolor*.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

**Plant-Microbe Interfaces: A high-throughput bioassay to investigate bacterial-provided benefits to heat stress**

Jun Hyung Lee[1]* (leej1@ornl.gov), Dana L. Carper,[1] Dale A. Pelletier,[1] David J. Weston,[1] Leah H. Burdick,[1] Sara S. Jawdy,[1] Alyssa A. Carrell,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://pmiweb.ornl.gov/

**Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.**

Heat is a major environmental factor that negatively affects plant growth and development. Along with breeding efforts to develop thermotolerant plants, exploiting beneficial plant microbes is a promising alternative approach. In our previous study, we demonstrated that the plant thermotolerant phenotype can be transmitted through the microbiome. To understand the underlying molecular genetics by which plants receive benefits from microbes to heat, we developed a high-throughput bioassay using *Arabidopsis thaliana* to screen individual bacterial strains that confer enhanced thermotolerance to the plant host. Multiple culture systems and heat treatment strategies were tested. The final system used hydroponically cultured seedlings with and without bacteria and was subjected to heat shock followed by recovery. Plants were then harvested to measure total chlorophyll content as an indicator of heat-induced damage. Initial screening results with several genera of rhizobacterial isolates from our inhouse collection found that *Variovorax* strains provided thermal benefit to the host plant. Subsequent screening results with a larger panel of 26 *Variovorax* found 6 strains that confer enhanced host plant thermotolerance. To investigate promising strains further, we examined a set of treatments that includes no heat shock control, non-lethal thermal exposure (i.e., thermoprimed) prior to heat shock, and heat shock with and without bacteria. Thermoprimed plants showed no significant difference ($p$=0.0587) in chlorophyll content compared to no heat shock control. Chlorophyll degradation by heat shock was observed in both inoculated plants (18%) and uninoculated plants (32.5%), whereas the level of damage was significantly reduced in inoculated plants ($p$<0.001), indicating bacterial-induced resilience to heat. Not only is this assay amenable to high-throughput screening of individual

bacterial isolates, but it can also scale to include multiple *Arabidopsis* genetic variants and microbial consortia.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research under FWP ERKP730.*

# Improving Iron-Sulfur Cluster Stability in *Zymononas mobilis* to Increase Terpenoid Production

Isabel Askenasy[1,2*] (askenasyflor@wisc.edu), Jyotsna Misra[1,2], and **Patricia Kiley**[1,2]

[1] DOE Great Lakes Bioenergy Center, University of Wisconsin-Madison, Madison, WI;
[2] Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI

https://www.glbrc.org

**Project Goals: Our goal is to increase the production of terpenoid precursors in the bacterium *Zymomonas mobilis.* These molecules are synthesized from glucose via the MEP pathway. This pathway has known rate-limiting steps catalyzed by the iron-sulfur dependent enzymes IspG and IspH. Our work focuses on increasing the activity of these steps to engineer *Z. mobilis* strains for terpenoid production.**

Terpenoids can substitute for petroleum in the production of compounds of economic value, including drugs, flavoring agents, and biofuels. The bacterium *Zymomonas mobilis* uses glucose to produce the terpenoid precursors dimethylallyl diphosphate (DMADP) and isopentenyl diphosphate (IDP) via the methyl erythritol phosphate (MEP) pathway. Thus, *Z. mobilis* has the potential to become an important engineering platform for terpenoids. However, recent data showed that $O_2$ exposure reduces terpenoid precursor production, accompanied by an accumulation of intermediates prior to the iron-sulfur (FeS) cluster-dependent enzymes IspG and IspH. These enzymes have been previously identified as the pathway's limiting steps. Since Fe-S clusters are known targets of oxidative damage, we are investigating if Fe-S cluster lability explains the effect of $O_2$ on flux through the enzymes IspG and IspH in *Z. mobilis* and accordingly, develop approaches to improve these enzymes' $O_2$ stability. We have taken two strategies to achieve this goal. First, we have taken advantage of a well characterized *E. coli* platform to examine *Z. mobilis* IspG and IspH activity under aerobic conditions compared to the native *E. coli* enzymes. Our results suggest that *Z. mobilis* IspH is more $O_2$ sensitive than either *E. coli* IspH or *Z. mobilis* or *E. coli* IspG. Further we found that the $O_2$ sensitivity of IspH function can be rescued by coexpression of IspG, indicating IspG and IspH may have co-evolved for optimal $O_2$ stability. In a second complementary approach, we are testing if flux through the MEP pathway can be improved by co-overexpressing IspG and IspH orthologs in *Z. mobilis.* To successfully accomplish this goal, we will need to provide strains with sufficient Fe-S cluster biogenesis machinery to assemble active proteins. Surprisingly, overexpression of the *Z. mobilis suf* genes, encoding the Suf Fe-S biogenesis machinery induces an $O_2$ sensitive phenotype in *Z. mobilis*. To have more refined control over *suf* operon expression, we are investigating its native regulation. We have found that *suf* expression is controlled by a transcription factor, RsuR, that is a homolog to [2Fe-2S]-IscR from *E. coli,* which regulates Fe-S cluster biosynthesis. Unlike E. coli IscR, RsuR binds a [4Fe-4S] cluster. Using DNaseI footprinting, we have identified the DNA sequence that RsuR binds to in the *suf* operon promoter region. DNA binding is cluster dependent. $O_2$ eliminates DNA binding and induces cluster degradation, which implies a sensing mechanism for the transcription factor. In summary, our diverse strategy will generate new knowledge relative to the Fe-S cluster dependent enzymes in the MEP pathway and insight into engineering strategies for boosting synthesis of Fe-S cluster enzymes in *Z. mobilis.* This new knowledge will allow us to generate a more robust strain of *Z. mobilis* with improved terpenoid production.

This material is based upon work supported in part by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409.

**Title:** A High Solids Field-To-Fuel Research Pipeline to Identify Interactions Between Feedstocks and Biofuel Production

**Authors:** Meenaa Chandrasekar[1,2], Leela Joshi[1,2], Karleigh Krieg[1,2], Sarvada Chipkar[1,2], Emily Burke[1,2], Derek J Debrauske[3], Kurt D Thelen[4], **Trey K Sato[3], Rebecca G Ong[1,2]**

**Institutions:** [1] DOE Great Lakes Bioenergy Research Center, Michigan Technological University, Houghton, MI, USA, [2] Department of Chemical Engineering, Michigan Technological University, Houghton, MI, USA, [3] DOE Great Lakes Bioenergy Research Center, Univ. of Wisconsin-Madison, USA, [4] DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, USA

**Project Goals:**

- To develop a laboratory-scale high solids field-to-fuel platform to evaluate fermentation performance of diverse feedstocks
- To achieve a higher throughput to process the large number of samples from agronomic studies

**Abstract Text:**

Environmental factors, such as weather extremes, have the potential to cause adverse effects on plant biomass quality and quantity. Beyond adversely affecting feedstock yield and composition, which have been extensively studied, environmental factors can have detrimental effects on saccharification and fermentation processes in biofuel production. Only a few studies have evaluated the effect of these factors on biomass deconstruction into biofuel and resulting fuel yields. This field-to-fuel evaluation of various feedstocks requires rigorous coordination of pretreatment, enzymatic hydrolysis, and fermentation experiments, which led to the need for a laboratory-scale high solids experimentation platform. A field-to-fuel platform was developed to provide sufficient volumes of high solids loading enzymatic hydrolysate for fermentation. AFEX pretreatment was conducted in custom pretreatment reactors, followed by high solids enzymatic hydrolysis. To accommodate enzymatic hydrolysis of multiple samples, roller bottles were used to overcome the bottlenecks of mixing and reduced sugar yields at high solids loading, while allowing greater sample throughput than possible in bioreactors. The roller bottle method provided 42–47% greater liquefaction compared to the batch shake flask method for the same solids loading. In fermentation experiments, hydrolysates from roller bottles were fermented more rapidly, with greater xylose consumption, but lower final ethanol yields and $CO_2$ production than hydrolysates generated with shake flasks. The entire platform was tested and was able to replicate patterns of fermentation inhibition previously observed for experiments conducted in larger-scale reactors and bioreactors, showing divergent fermentation patterns for drought and normal year switchgrass hydrolysates. Thus, pipeline of small-scale AFEX

pretreatment and roller bottle enzymatic hydrolysis can be effectively utilized to compare divergent feedstocks and diverse process conditions.

**Funding Statement:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409 and DE-FC02-07ER64494.

# Characterizing fungal inhibitors in drought-stressed switchgrass

Sarvada Chipkar[1,2], Katie Smith[1], Larkin Hooker-Moericke[1], Marissa Gallmeyer[1], Elizabeth Whelan[5], Derek Debrauske[5], Katherine Overmyer[4,5], Annie Jen[4,5], Arthur Daniel Jones[3,6], Joshua Coon[4,5], Trey Sato[5], Rebecca G. Ong[1,2]

[1]Michigan Technological University, Houghton, MI;
[2]DOE-Great Lakes Bioenergy Research Center (GLBRC), Michigan Technological University, Houghton, MI;
[3]Mass Spectrometry and Metabolomics Core, Michigan State University, East Lansing, MI;
[4]Coon Metabolomics lab, University of Wisconsin-Madison, Madison, WI;
[5]DOE-GLBRC, University of Wisconsin-Madison, Madison, WI; [6]DOE-GLBRC, Michigan State University, East Lansing, MI.

**Project goals: The overarching goal of this project is to identify fungal inhibitors from drought-stressed switchgrass using extraction techniques and use this information to develop strategies to overcome biomass variability and ensure consistent biofuel generation.**

Development of economically viable and greener pathways to synthesize renewable energy has become an important research theme in recent years. Lignocellulosic biomass is a promising group of feedstocks that can be used for second-generation biofuel production. Recent research has showed that biomass characteristics are altered by environmental growth conditions, and directly influence the extent of biomass conversion to fuels. Specifically, it was reported that drought experienced during the growth of switchgrass led to complete inhibition of yeast growth during fermentation [1]. Our goal in this project was to characterize specific compounds that led to this inhibition and to determine whether the microbial-inhibitors are plant-generated compounds, by-products of the pretreatment process, or a combination of both. Switchgrass harvested in drought (2012) and non-drought (2010) years were pretreated using Ammonia Fiber Expansion (AFEX). Untreated and AFEX processed samples were then extracted using solvents (i.e. water, ethanol, and ethyl acetate) to selectively remove potential inhibitory compounds and determine whether pretreatment affects the inhibition. High solids loading enzymatic hydrolysis was performed on all samples followed by fermentation using yeast strain *Saccharomyces cerevisiae* (GLBRCY945). Cell growth ($OD_{600}$), sugar consumption, and ethanol production were used to evaluate fermentation performance. Extracts were analyzed using liquid chromatography-mass spectrometry (LC-MS) to identify potential inhibitory compounds. Tandem MS analysis was conducted to annotate the identities of unknown inhibitors. We found that water extraction of drought-year switchgrass before AFEX pretreatment was most effective in overcoming yeast inhibition. We also identified numerous saponins, a class of plant-generated triterpene glycosides, which were significantly more abundant in the water extracts from drought-year (inhibitory) switchgrass. These compounds are commonly known as natural laundry detergents and have been reported as toxic to eukaryotic cells by various researchers [2]. Our analysis showed that plant generated compounds inhibited the conversion and that including a water extract step might reduce the inhibition and increase the biofuel yield from biomass.

***This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409***

## References

[1] Ong, R. G., et al. (2016). Biotechnol Biofuels 9: 237.
[2] Heng W., et al (2014). *Industrial Crops and Products*, 61:422-429.

**Title:** Genome-scale CRISPRi Enables Functional Annotation of the *Zymomomas mobilis* Genome

**Authors:** Amy B. Banta,[1,2] Amy L. Enright[1,2,3*] (alenright@wisc.edu), and **Jason M. Peters**[1,2,4,5]

**Institutions:** [1]DOE Great Lakes Bioenergy Research Center, Madison, WI; [2]Pharmaceutical Sciences Division, University of Wisconsin-Madison; [3]Microbiology Doctoral Training Program, University of Wisconsin-Madison; [4]Department of Bacteriology, University of Wisconsin-Madison; and [5]Department of Medical Microbiology and Immunology, University of Wisconsin-Madison

**Project Goals:** Enable rational engineering of the promising biofuel producer *Zymomonas mobilis* by using genome-wide CRISPR interference (CRISPRi) screens for gene function discovery and annotation, including for essential and/or biofuel-relevant genes.

**Abstract:** The emerging model bacterium, *Zymomonas mobilis*, is a promising biofuel producer, but rational engineering of the genome to improve biofuel yields relies on gene function information that is currently lacking. To probe gene function in *Z. mobilis*, we previously developed a CRISPRi (clustered regularly interspaced short palindromic repeats interference) system that caused robust gene knockdown and enabled phenotyping of metabolic and stress genes. Here, we utilize *Z. mobilis* CRISPRi for systematic gene phenotyping at the genome scale and identify genes that are conditionally essential for growth in aerobic or anaerobic conditions. Our screen uncovered expected (e.g., superoxide dismutase) and unexpected (e.g, ssDNA-specific exonuclease RecJ) players in oxygen tolerance. Further, we found a surprising role for the ATP synthase in maintaining the electrochemical gradient during anaerobic growth. Future work will focus on genes that are crucial for growth in the presence of biofuel-relevant stresses such as plant-derived toxins and accumulation of fermentation products. Identification of bioenergy-relevant genes will enable informed genetic engineering of stress-tolerant *Z. mobilis* strains with increased yields, closing the economic gap between biofuels and fossil fuels and paving the way toward mitigation of climate change.

**References/Publications**

1. Banta AB, Enright AL, Siletti C, Peters JM. A High-Efficacy CRISPR Interference System for Gene Function Discovery in *Zymomonas mobilis*. *Appl Environ Microbiol*. 2020;86(23):e01621-20. Published 2020 Nov 10. doi:10.1128/AEM.01621-20

**Funding Statement:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409.

**Title:** Economic and Environmental Analysis of Bioenergy with Carbon Capture and Sequestration (BECCS) Systems

**Authors:** Caleb H. Geissler[1,2*] (geissler@princeton.edu) and **Christos T. Maravelias**[1,2,3]

**Institutions:** [1]Princeton University, Princeton, NJ; [2]DOE Great Lakes Bioenergy Research Center; [3]Andlinger Center for Energy and the Environment, Princeton, NJ

**Project Goals:** This project aims to examine the impact of carbon sequestration credits, biorefinery capacity, and energetic self-sufficiency on the cost optimal conversion technology for a biorefinery with carbon capture, and how this is expected to change into the future. Performance metrics include cost per gallon gasoline equivalent and the greenhouse gas balance of the biorefinery and supply chain system.

**Abstract Text:** Bioenergy with carbon capture and sequestration (BECCS) involves simultaneous fuel production with permanent carbon dioxide removal (CDR) to achieve net-negative emissions. A variety of fuels can be produced by a range of biological and thermochemical conversion technologies. Conventional liquid fuels can be produced via fermentation, pyrolysis, or gasification. Additionally, hydrogen can be produced via gasification and the water-gas-shift reaction, or electricity can be produced by via either direct combustion or integrated gasification and combustion. Each conversion technology to fuel combination results in point sources of $CO_2$ emissions with different flow rates and purity from which $CO_2$ could be captured.

Previous studies are typically limited to a single conversion technology and capture source, or a small subset of the potential options, at consider a fixed biorefinery capacity. We use a mixed-integer nonlinear programming (MINLP) model of a lignocellulosic biorefinery and supply chain system that includes biorefineries utilizing fermentation, pyrolysis, gasification, or combustion, capture from any point source, and options to purchase hydrogen or electricity to increase fuel yield or capture rates as appropriate. We examine the impact of input parameters such as carbon sequestration credit, biorefinery capacity and energetic self-sufficiency on the economic and environmental performance of BECCS systems. In addition to determining the expected greenhouse gas (GHG) mitigation based on the current production mix, we compare the expected GHG mitigation into the future as the energy production mix changes.

**Funding statement:** This research was supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018409.

**Multiple-approaches to engineer mixed-linkage glucan in sorghum**

Sang-Jin Kim (sjkim@msu.edu)[1,2*], Starla Zemelis-Durfee[1], Magie Williams[1], Rylee Sokoloski[1], Maureen Cronin[1], **Federica Brandizzi**[1,2,3]

[1]Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI; [2]MSU-DOE Plant Research Lab, Michigan State University, East Lansing, MI; [3]Department of Plant Biology, Michigan State University, East Lansing, MI.

https://www.glbrc.org

**Project Goals:** The goal of Bioenergy Plant Design team in Great Lakes Bioenergy Research Center (GLBRC) is to increase the quantity and quality of bioenergy crop biomass per hectare of land, which is crucial for the sustainable and economically viable production of lignocellulosic-derived bioproducts.

Mixed-linkage glucan (MLG), a β-(1,3;1,4)-glucose polymer, is one of the most abundant cell wall matrix components in grass species. Due to its simple structure and composition, which make MLG a highly digestible polymer, it has become the target of cell wall manipulation to improve bioenergy feedstocks for microbial conversion. Therefore, we are aiming to accumulate MLG in bioenergy sorghum, which exhibits high biomass yield and elevated content of soluble sugars. To increase the levels of MLG in this crop, we are focusing on MLG synthesis and degradation with the goal to overproduce MLG and prevent degradation of MLG in the stem. Overproduction of MLG hampers growth in some species, such as Brachypodium, rice and barley. Nevertheless, we established a technology to generate transgenic sorghum overexpressing a major MLG synthase (CSLF6) from either Brachypodium or sorghum (constructs named BdCSLF6 and SbCSLF6, respectively), which are known to generate a different frequency of β1-3 linkages in MLG, resulting in different types of fine structure of this polymer. The transgenic sorghum grown in a greenhouse showed higher MLG levels compared to wild type (WT) without any noticeable growth defects. In summer 2021, we conducted field trials (East Lansing, MI) with the T3 generation of CSLF6-sorghum transgenics and found no defects in growth and photosynthesis. MLG analyses from the field-grown sorghum were performed during development within single stem internodes. We observed that the transgenic lines have higher levels of MLG compared to WT, even in the fully elongated region of the internode where MLG degradation normally occurs. Furthermore, to modulate and reduce MLG degradation in sorghum stem and lead to higher levels of MLG, we identified and characterized three sorghum MLG endoglucanases (also called lichenases). We established that the three lichenases have similar enzymatic activity, with an optimal at pH5, between 25˚C and 45˚C. We also found that the three enzymes localize in the apoplast, where lichenase activity is expected to take place. In addition, we established the lichenases' expression pattern during development and light conditions, which led us to identify the lichenase that is likely responsible for MLG degradation in sorghum. We are now generating CRISPR-Cas9-mediated suppression of all three lichenases

in sorghum for crossing with the CSLF6-overexpressing lines to maximize the levels of MLG in the stem.

*This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409.*

**Title:** Building a Molecular Understanding of Biomass Deconstruction

**Authors:** Nathaniel J. Kuch[1,2], Elise Gilcher[1,3], Craig A. Bingman[2,4], Theodore W. Walker[1,3], Alex Parker[1,2,5], Mark Kutschke[1,2], James A. Dumesic[1,3], **Brian G. Fox[1,2]**

**Institutions:** [1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison; [2]Department of Biochemistry, University of Wisconsin-Madison, Madison; [3]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison; [4]Collaborative Crystallography Core, University of Wisconsin-Madison, Madison; [5]Dane County Youth Apprenticeship Program, Dane County School Consortium, Monona, WI

**Project Goals:**

The goal of this project is to combine insights from solid-state nuclear magnetic resonance (ssNMR), x-ray crystallography, and enzymology to expand understanding of how biomass impacts and is impacted by both thermochemical pretreatment and enzymatic hydrolysis. Understanding biomass deconstruction on a molecular scale enables more detailed pretreatment and enzyme improvements and clearer, more specific identification and addressing of process bottlenecks.

**Abstract Text:**

Identifying bottlenecks in the deconstruction of plant biomass is an important aspect of developing increasingly effective deconstruction strategies. Because of the heterogeneity of biomass, both between plant species and within a single sample, the specificity of bottleneck identification has traditionally been a limitation. For example, it is well known that crystalline cellulose is less amenable to enzymatic hydrolysis than amorphous cellulose; and lignin is a potent inhibitor of cellulases, but correlations between these different molecular states and their impact on deconstruction is less clear.

We previously showed a correlation between enzymatic hydrolysis and the intensity of the split C4 resonance in bulk cellulose [1], which holds across plant species and co-solvents used in pretreatment. To better understand the reactivity of a model purified enzyme (*Rumiclostridum thermocellum* CelR) with cellulose produced via γ-valerolactone pretreatment (GVL cellulose), CelR variants consisting of the catalytic domain alone, the native enzyme (catalytic domain and a single carbohydrate binding module (CBM)), and an engineered native construct containing an additional CBM were constructed along with variants of each inactivated by mutation of the catalytic glutamate residue. Studies of the kinetics, binding affinities, and activity profiles of these variants on amorphous, crystalline, and GVL cellulose showed that the catalytic domain alone is nearly inactive; restoration of the native CBM improves specific activity by 10-fold and addition of the engineered CBM doubles again the specific activity.

In this work, we have extended our prior ssNMR studies by using spectral deconvolution and fitting to quantify contributions of different cellulose substructures within GVL cellulose and to correlate changes in the proportion of these substructures during thermochemical treatment and enzyme hydrolysis. Results show that the $I_\beta$ form of crystalline cellulose accumulates during the GVL treatment and remains unhydrolyzed by CelR. $I_\beta$ cellulose is therefore a major bottleneck for deconstruction by CelR and can be targeted to improve deconstruction overall at both the pretreatment and enzymatic hydrolysis levels. These findings provide a molecular-level understanding of the progress of biomass deconstruction, and insight into how to improve this essential process.

**References/Publications**

1. Walker TW, Kuch N, Vander Meulen KA, Clewett CFM, Huber GW, Fox BG, Dumesic JA. Solid-state NMR studies of solvent-mediated, acid-catalyzed woody biomass pre-treatment for enzymatic conversion of residual cellulose. *ACS Sustain Chem Eng*. 2020 Apr 27; 8(**16**): 6551-6563. doi: 10.1021/acssuschemeng.0c01538.

**Funding Statement:**

This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409. Nathaniel J. Kuch was supported by the UW-Madison Biotechnology Training Program, grant number NIH 5 T32 GM135066.

**Title:** Development and Function of Bioenergy Sorghum's Deep Roots

**Authors:** Austin Lamb*[1,2] (al17@tamu.edu), Brock Weers[1], Brian McKinley[1,2], William Rooney[2,3], Cristine Morgan[4], Amy Marshall-Colon[5,6], and **John Mullet[1,2]**

**Institutions:** [1]Department of Biochemistry, Texas A&M University, College Station, TX; [2]DOE Great Lakes Bioenergy Research Center, Texas A&M University; [3]Department of Soil and Crop Science, Texas A&M University, College Station, TX; [4]Soil Health Institute, Morrisville, NC; [5]Department of Plant Biology, University of Illinois, Champaign-Urbana, IL; and [6]DOE Center for Advanced Bioenergy and Bioproducts Innovation – University of Illinois

**Project Goals:** To investigate the bioenergy sorghum's root system development, morphology, anatomy, and gene expression and adaptations that help confer drought resilience and nutrient use efficiency, and total root biomass distribution within the soil profile.

**Abstract Text:** Bioenergy sorghum has low input requirements, good water and nutrient use efficiencies, high yield potential, and drought and heat tolerance. Under field conditions this study analyzed bioenergy sorghum`s root development, morphology, anatomy, and transcriptome profiles during a 155-day growing season. Bioenergy sorghum`s root system grew continuously, produced ~175 nodal roots, accumulated ~7Mg of biomass per hectare, and reached ~3m deep into the soil profile. Nodal roots within 20cm of the culm were between one and ten millimeters in diameter whereas outside of this region, where lateral roots predominate, the root diameters ranged from 30 to 500µm enabling growth through soil macropores. Nodal and lateral roots produced lysigenous aerenchyma capable of fully degrading the root cortex leaving the endodermal, vascular, and inner root tissues intact. Nodal, surface (0-20cm), and deep (180-240cm) roots were subjected to transcriptomic analysis; differentially expressed genes involved in root growth, transport, adaptation, defense, and AMF-interactions were identified. Bioenergy sorghum`s root system grew for a longer time, deeper into the soil, and accumulated more biomass than many annual grain crops; these traits could help restore annual cropland soil organic carbon and improve soil productivity. Deep roots active in nitrogen and phosphate uptake are poised to take up leached surface nutrients helping to mitigate nutrient runoff and eutrophication. Bioenergy sorghum's expansive root system is critical for the sustainable production of biomass on annual cropland while helping to mitigate carbon emissions.

**References/Publications**

1. Lamb, A., Weers, B., McKinley, B., Rooney, W., Morgan, C., Marshall-Colon, A., & Mullet, J. (2021). Bioenergy sorghum's deep roots: A key to sustainable biomass production on annual cropland. *GCB Bioenergy*, 00, 1– 25. https://doi.org/10.1111/gcbb.12907

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018409, and the DOE Advanced Research Projects Agency-Energy, grant no. DE-AR0000823.

# Role of the Phyllosphere microbiome in Sorghum Resilience

**Authors:** Marco E. Mechan-Llontop[1,2]* (mechanll@msu.edu), John Mullet[2,3], and **Ashley Shade**[1,2]

**Institutions:** [1]Michigan State University, East Lansing, MI; [2]Great Lakes Bioenergy Research Center, Michigan State University, MI; [3]Texas A&M University, College Station, TX.

**Project Goals:**

The research goal of this project is to determine the taxonomic, phylogenetic, and functional diversity of the microbiome of the sorghum aerial root mucilage and epicuticular wax, and how it changes under plant stress. Our overarching objective is to understand how sorghum interacts with the mucilage and wax microbiome and to determine plant-interactions that can be leveraged to support sorghum resilience and productivity.

**Abstract:**

During development, sorghum (*Sorghum bicolor)* produces several exudates on its external aerial surfaces (epiphytic phyllosphere). In particular, sorghum accumulates elevated levels of epicuticular wax on its stems and leaves to prevent water loss. Also, sorghum produces sugar-rich mucilage on its aerial roots to likely facilitate nutrient acquisition, including nitrogen fixation as has been reported in landrace maize. Here, we hypothesize that the epicuticular wax structure selectively filters for microbiome members that confer drought tolerance to the host and that the aerial root mucilage harbors a diazotrophic microbiota that supports nitrogen acquisition. Thus, we assessed the microbiome associated with the aerial root mucilage from N-fertilized and non-fertilized sorghum plants at two points in the growing season, and the epicuticular wax from plants under non-limiting and limiting water conditions.

Our initial results showed that Proteobacteria, Bacteroidetes, and Firmicutes are the dominant phyla in the aerial root mucilage compartment regardless of the fertilization treatment. The relative abundance of *Pantoea* was higher in the mucilage at an earlier stage of sorghum growth, whereas *Flavobacterium* and *Pseudomonas* were enriched in the mucilage later in the growing season. The dominant epicuticular wax microbial phyla were Proteobacteria and Firmicutes. We also identified *Sphingomonas* as a major taxon associated with the wax. To capture microbiome members hypothesized to be important for host resilience, we built a large bacterial collection by targeting a wide range of traits expected to be beneficial for the host plant during stress.

The outcome of this work will inform understanding of the interplay between the phyllosphere microbiome and plant exudates for sorghum resilience and productivity, and supports our long-term goal of translating our findings into sorghum cultivation for biofuel production and crop adaptation to stress.

**Funding statement:** This work is supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409. Support for field research was provided by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological

and Environmental Research (Awards DE-SC0018409 and DE-FC02-07ER64494), by the National Science Foundation Long-term Ecological Research Program (DEB 1637653) at the Kellogg Biological Station, and by Michigan State University AgBioResearch.

**Title:** Can Perennial Bioenergy Cropping Systems Promote Negative $N_2O$ Fluxes?

**Authors:** Ekrem Ozlu,[1,2]* (ozluekre@msu.edu), Kevin Kahmark[2], Sven Bohm[2], and **G. Philip Robertson**[1,2]

**Institutions:** [1]Department of Plant, Soil and Microbial Sciences and [2]W. K. Kellogg Biological Station, Michigan State University, Hickory Corners, Ml

**Website URL:** glbrc.org

**Project Goals:**

(i) Determine the occurrence and frequency of negative $N_2O$ fluxes from bioenergy cropping systems using automated flux chamber technology

(ii) Identify environmental and soil characteristics that might favor $N_2O$ consumption and determine their predictive capacity

**Abstract Text:** Atmospheric nitrous oxide ($N_2O$) concentrations reached a new high of 331 ppb in 2019 primarily due to agricultural intensification and soil disturbance. Efforts to mitigate $N_2O$ from agricultural soils have thus far focused on reducing fluxes from high-emitting systems such as fertilized row crops. The potential for promoting negative fluxes (denitrification of atmospheric $N_2O$) in low-emitting systems such as perennial bioenergy feedstocks has been unexplored, largely because of the technical difficulty of identifying small negative fluxes within a background of more frequent and easier-to-detect high or net-zero fluxes. Automated flux chambers deployed in situ offer the opportunity to better resolve the potential for negative fluxes with near-continuous measurements and low analytical error. We examined sub-daily $N_2O$ fluxes from 2012 to 2017 (six years) under nine different bioenergy cropping systems in the GLBRC Biofuel Cropping System Experiment at the W. K. Kellogg Biological Station in southwest Michigan USA. Daily, weekly, monthly, and annual flux measurements were calculated individually for positive and negative sub-daily data after first determining a minimum flux detection limit of ±1.029 g $N_2O$-N $ha^{-1}$ $d^{-1}$. Of 37,000 sub-daily fluxes, 34% were statistically less than zero, ranging from -1.029 to -926 g $N_2O$-N $ha^{-1}$ $d^{-1}$. Weather transitions lowered the magnitude of negative $N_2O$ fluxes. Minimum daily temperature and volumetric soil moisture content (0-10 cm depth) together explained 45% of negative $N_2O$ fluxes. The variability of $N_2O$ fluxes and the preceding seven days' average precipitation were strongly associated. Higher plant diversity and lower inputs also promoted the frequency and magnitude of negative fluxes. Results suggest a potential for designing bioenergy cropping systems with a capacity to consume atmospheric $N_2O$.

**Funding Statement:** Support for this research was provided by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Award DE-SC0018409), by the National Science Foundation Long-term Ecological Research Program (DEB 1832042) at the Kellogg Biological Station, and by Michigan State University AgBioResearch.

# Effects of Chloramphenicol Treatment on Cellular Storage Granules and Membrane Structures in *Rhodobacter sphaeroides*

Daniel Parrell[1,2,3]* (dparrell@wisc.edu), Rachelle A.S. Lemke,[2,3] Joseph Olson,[1] Timothy J. Donohue,[2,3] and **Elizabeth Wright**[1,2,3]

[1] Department of Biochemistry, University of Wisconsin, Madison, WI
[2] DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI
[3] Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI

**Project Goals:** This project seeks to define the physiological responses of bacterial cells to antibiotic treatments. This information will be used to improve our understanding of bacterial stress responses, including the induced accumulation of cellular structures that may be beneficial for bio-industrial and bioenergy purposes. Bacterial storage compartments, such as polyhydroxybutyrate (PHB) granules and polyphosphate (PP) granules are cellular structures highly enriched in specific chemical compounds. Bacterial membranes are a source for a variety of lipids. All of these structures may serve as important sources of primary material for a number of pharmaceutical, industrial, and biofuel applications. In this work, we use the photoheterotrophic bacterium *Rhodobacter sphaeroides* to study the accumulation of storage granules and membranes using cryo-electron tomography (cryo-ET), fluorescence microscopy, and biochemical purifications. Addressing these goals will provide a significant benefit to the development of renewable biofuels and bioproducts.

*Rhodobacter sphaeroides* is an excellent model organism for the studies of cellular structures and bioenergy. It is a facultative photoheterotrophic bacterium that generates diverse cellular structures, including various storage granules and intracytoplasmic membranes, depending on its growth conditions. In this study, we observed the behavior of three cellular structures in response to the translation blocking antibiotic, chloramphenicol (Cm). The storage granules we studied were: PHB, a subcellular compartment where butyrate monomers are polymerized and are stored under nutrient rich conditions; PP, a chain of inorganic phosphate residues linked together to serve as storage for reducing potential; and internal and external cell membrane derived structures. An *R. sphaeroides* culture, grown in Sistrom's medium, was treated with 200 μg/mL Cm and samples were collected for cryo-ET, fluorescence microscopy, and gas chromatography mass spectrometry (GCMS). Upon Cm treatment, the occurrence of PHB granules inside of individual cells changed. Using cryo-ET, it was observed that the average radius of individual PHB granules increased nearly 7-fold, and the corresponding volume of PHB per cell increased approximately 5-fold, despite this volume being comprised of fewer granules per cell. The accumulation of PHB in the cell was further verified by GCMS analysis. In addition to PHB granules, *R. sphaeroides* cells accumulate PP in granules that are smaller, more electron dense, and less labile to electron dose than PHB. These observations will be presented. The presence of PP granules can be controlled by growth in a medium lacking excess phosphate, further corroborating their identity as PP granules. Accumulation of PP is also increased by Cm treatment. In addition to the observation of changes to these nutrient storage granules, it was also noted in cryo-ET volumes that aggregations of cellular membranes were present in the wild-type strain of *R. sphaeorides*. Highly irregular membrane aggregations in the periplasmic space formed both at the pole and along the cell body. These membrane aggregations were apparent by

fluorescence microscopy using membrane staining. Progress toward analyzing the changes in fatty acid and lipid profiles following Cm treatment will be presented.

Our observations lead to an understanding of the cellular responses to antibiotic treatment. Since chloramphenicol blocks translation, thereby inhibiting cellular processes, it is possible that accumulation of nutrient and energy stores may occur as a method of mitigating and preparing for Cm induced stress. Accumulation of irregular membrane structures demonstrates that perhaps misregulation of intracellular membrane production occurs. Our findings lead to a better understanding of mechanisms bacteria may use to respond to, and possibly cope with, antibiotic stress, thus having the potential to impact studies of antibiotic resistance. Additionally, these analyses demonstrate conditions under which *R. sphaeroides* can be encouraged to produce stores of carbohydrates, lipids, and reducing power – all of which may have utility toward creating better chemicals for industrial purposes and biofuels.

**Funding statement:**
This material is based upon work supported in part by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409. Additional support is provided by National Institutes of General Medical Sciences F32 fellowship 1F32GM143854 to DP and by the University of Wisconsin, Madison, the Department of Biochemistry at the University of Wisconsin, Madison. We are grateful for the use of facilities and instrumentation at the Cryo-EM Research Center in the Department of Biochemistry at the University of Wisconsin, Madison.

**Title:** Engineering *Novosphingobium aromaticivorans* to Stoichiometrically Convert S-type Aromatics into PDC

**Authors:** Jose M. Perez[1,4,5]* (perez8@wisc.edu), Wayne S. Kontur,[4,5] Carson Gehl,[2,4,5] Derek M. Gille,[4,5] Yanjun Ma,[4,5] Alyssa V. Niles,[4,5] German Umana,[4,5] Timothy J. Donohue,[3,4,5] and **Daniel R. Noguera**,[1,4,5]

**Institutions:** Departments of [1]Civil and Environmental Engineering, [2]Biomedical Engineering, [3]Bacteriology and [4]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI; and [5]Wisconsin Energy Institute, Madison, WI.

**Project Goals:** The overall project aims to valorize the lignin fraction of plant biomass via chemical fractionation and depolymerization followed by conversion of the resulting mixtures of aromatic compounds into single valuable chemicals by genetically engineered bacteria. The goal of this study was to identify the key O-demethylases and ring-opening dioxygenases involved in the degradation of plant-derived aromatic compounds in *Novosphingobium aromaticivorans*.

One of the major components of plant biomass is lignin, a heterogeneous and recalcitrant aromatic polymer. A strategy to make value from lignin is to use chemical techniques to depolymerize it into mixtures of phenolic compounds and then funnel these mixtures into a single product using engineered bacteria. *Novosphingobium aromaticivorans* DSM12444 can naturally degrade multiple lignin-derived phenolic compounds and we have previously engineered it to produce 2-pyrone-4,6-dicarboxylic acid (PDC) from a variety of compounds that contain S-, G-, or H-type aromatic structures (two, one, or no methoxy groups attached to the aromatic ring, respectively). However, conversion of S-type aromatics into PDC by this engineered strain is below stoichiometric, suggesting the existence of competing metabolic routes that do not involve the production of the target product. One potential competing pathway involves aromatic O-demethylation followed by oxidative aromatic ring opening, with gallic acid as an intermediate metabolite. Since enzymes performing these type of reactions commonly have broad substrate specificity and their inactivation can potentially affect the conversion of other aromatic compounds, we investigated enzymes hypothesized to be involved in O-demethylation of syringic acid, vanillic acid, and 3-methoxygallic acid (3-MGA), and enzymes hypothesized to participate in ring opening of 3-MGA, gallic acid, and protocatechuic acid (PCA). Our results confirmed the existence of an alternative pathway for the degradation of S-type aromatics and the generally broad substrate specificity of the tested enzymes. For instance, the O-demethylase DesA was active on syringic and vanillic acids, whereas LigM was involved in O-demethylation of vanillic acid and 3-MGA. We also found evidence of a new O-demethylase transforming 3-MGA into gallic acid. Among aromatic ring opening dioxygenases, LigAB had activity with 3-MGA, gallic acid, and PCA. In addition, LigAB2 was identified as a new aromatic ring opening dioxygenase active with the same substrates.

The data obtained in this study revealed a previously uncharacterized route for metabolism of S-type aromatic compounds in *N. aromaticivorans* that involves O-demethylation of 3-MGA into gallic acid followed by aromatic ring opening. We predict that in wild-type *N. aromaticivorans* this pathway carries ~15% of the carbon flux from syringic acid, explaining the below stoichiometric transformation of S-type aromatics to PDC in engineered strains that did not block this pathway. By further inactivating O-demethylation of 3-MGA in the PDC-producing strain, we created a new engineered strain (PDC2) that stoichiometrically converts S-, G-, and H-type aromatics to PDC.


1. Jose M. Perez et al. (2019) Funneling aromatic products of chemically depolymerized lignin into 2-pyrone-4-6-dicarboxylic acid with *Novosphingobium aromaticivorans*. *Green Chemistry*, DOI:10.1039/c8gc03504k.
2. Jose M. Perez et al. (2021) Redundancy in aromatic O-demethylation and ring opening reactions in *Novosphingobium aromaticivorans* and their impact in the metabolism of plant derived phenolics. *Applied and Environmental Microbiology*. DOI:10.1128/aem.02794-20



This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409.

**Title**: Lipid Membrane Remodeling during Ethanol, Isobutanol, and Lignocellulosic Hydrolysate Stress in *Zymomonas mobilis*

**Authors:** Julio Rivera Vazquez[1,2]*(RiveraVazque@Wisc.Edu), Melanie Callaghan[1,2], Tyler Jacobson[1,2], Edna Trujillo[1,3], Joshua J Coon[1,3], **Daniel Amador-Noguez**[1,2]

**Institutions:** [1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53726; [2]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706; [3]Genome Center of Wisconsin, Madison, Wisconsin, USA.

**Project Goals:** The lipid membrane is the first line of defense for bacteria and evidence shows that microorganisms tend to remodel their lipid membrane in response to stressors in the environment. The response, which aids in the defense, often consists of increasing certain lipid classes and/or changing the composition of the fatty acid chains. Understanding a microorganism's universal lipid membrane remodeling response could lead to the data necessary to engineer strains that are capable of withstanding various stressors.

**Abstract Text:** *Zymomonas mobilis* is an ethanologenic gram-negative bacterium currently being developed for production of isobutanol. Despite being a proficient alcohol producer, *Z. mobilis* experiences growth inhibition at high ethanol titers and is highly sensitive to isobutanol. It is known that bacteria can modulate lipid membrane composition to increase their tolerance against environmental stressors. In this study, we used LC-MS/MS-based lipidomics to measure changes in lipid membrane composition that occur when *Z. mobilis* is exposed to increasing concentrations of ethanol, isobutanol, or ammonia fiber expansion (AFEX) lignocellulosic hydrolysate. Exposure to ethanol and isobutanol resulted in significant but distinct changes to the lipid and fatty acid composition. Affected lipid classes included ceramides, cardiolipins, phosphatidylcholines, and phosphatidylethanolamines. The fatty acid composition was also significantly affected. Most notably, we observed a substantial increase in C19 cyclopropane fatty acid content when cells were grown at high ethanol concentrations, suggesting that the changes comprise a defense mechanism in response to solvent stress. Previous evidence showed that cyclopropane ringed fatty acids modify membrane fluidity and act as a barrier to prevent detrimental molecules from entering the cell. To test the hypothesis that C19 cyclopropane fatty acids and derived lipids contribute to solvent resistance in *Z. mobilis*, we engineered a strain that overexpressed the Cyclopropane Fatty Acyl Synthase (CFA-Synthase) protein (ZMO1033) responsible for transforming unsaturated fatty acids into cyclopropane fatty acids. Analysis of the lipid membrane composition of the CFA-synthase overexpressing strain showed a significant increase in C19 cyclopropane fatty acid content for all lipid classes. This increase correlated with significantly improved growth rates in the presence of high ethanol and isobutanol concentrations. These data demonstrate the importance of cyclopropane fatty acids to solvent stress resistance in *Z. mobilis* and will allow us to engineer strains that are more resistant to high ethanol and isobutanol concentrations.

**Funding Statement:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409.

**Title:** Crabtree/Warburg-like aerobic xylose fermentation by engineered *Saccharomyces cerevisiae*

**Authors:** Sae-Byuk Lee[1,2,3,4], Mary Tremaine[1], Michael Place[1,3,4], Lisa Liu[1], Austin Pier[1], David J. Krause[1,2,3,4], Dan Xie[1], Yaoping Zhang[1], Robert Landick[1,5,6], Audrey P. Gasch[1,3,4], **Chris Todd Hittinger**[1,2,3,4], and **Trey K. Sato**[1,*] (tksato@glbrc.wisc.edu)

**Institutions:** [1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, WI; [2]Wisconsin Energy Institute, J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI; [3]Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI; [4]Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI; [5]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI; [6]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI

**Project Goals:** We aimed to identify genetic modifications that enhance metabolic flux and the conversion of xylose into fermentative end-products by yeast through directed evolution.

**Abstract:**

Bottlenecks in the efficient conversion of xylose into cost-effective biofuels have limited the widespread use of plant lignocellulose as a renewable feedstock. The yeast *Saccharomyces cerevisiae* ferments glucose into ethanol with such high metabolic flux that it ferments high concentrations of glucose aerobically, a trait called the Crabtree/Warburg Effect. In contrast to glucose, most engineered *S. cerevisiae* strains do not ferment xylose at economically viable rates and yields, and they require respiration to achieve sufficient xylose metabolic flux and energy return for growth aerobically. Here, we evolved respiration-deficient *S. cerevisiae* strains that can grow on and ferment xylose to ethanol aerobically, a trait analogous to the Crabtree/Warburg Effect for glucose. Through genome sequence comparisons and directed engineering, we determined that duplications of genes encoding engineered xylose metabolism enzymes, as well as *TKL1*, a gene encoding a transketolase in the pentose phosphate pathway, were the causative genetic changes for the evolved phenotype. Reengineered duplications of these enzymes, in combination with deletion mutations in *HOG1*, *ISU1*, *GRE3*, and *IRA2*, increased the rates of aerobic and anaerobic xylose fermentation. Importantly, we found that these genetic modifications function in another genetic background and increase the rate and yield of xylose-to-ethanol conversion in industrially relevant switchgrass hydrolysate, indicating that these specific genetic modifications may enable the sustainable production of industrial biofuels from yeast. We propose a model for how key regulatory mutations prime yeast for aerobic xylose fermentation by lowering the threshold for overflow metabolism, allowing mutations to increase xylose flux and to redirect it into fermentation products

**Funding Statement:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409 and DE-FC02-07ER64494.

**Title:** Linking Microbial Funneling to Hydrogenolysis-Based Lignin Depolymerization to Produce 2-Pyrone-4,6-Dicarboxylic Acid from Phenolic Monomers

**Authors:** Canan Sener[1,2*] (csener@wisc.edu), Jose M. Perez,[1,2] German E. Umana,[1,2] Shamik Misra,[1,2] Yaoping Zang,[1,2] Christos Maravelias,[1,2,3] Steven D. Karlen,[1,2] John Ralph,[1,2] Timothy J. Donohue,[1,2] and **Daniel R. Noguera**[1,2]

**Institutions:** [1]DOE Great Lakes Bioenergy Research Center, Madison, WI; [2]University of Wisconsin-Madison, Madison, WI; [3]Princeton University, Princeton, NJ

**Project Goals:** To develop a biomass-to-bioproduct processing chain to produce 2-pyrone-4,6-dicarboxylic acid (PDC) through microbial funneling of the phenolic monomers derived from catalytic depolymerization of lignin.

**Abstract Text:** When designing viable biorefineries, it is imperative to maximize the value obtained from all the components of lignocellulosic biomass. Lignocellulosic biomass is composed of 70-85 wt% polysaccharides (cellulose and hemicellulose) and 15-30 wt% lignin, a heteropolymer of aromatics. As a bountiful source of renewable carbon for the sustainable production of fuels and chemicals, the polysaccharide fractions have been extensively studied and various processes have been developed to produce valuable liquid fuels and commodity chemicals. The lignin fraction, being difficult to fractionate and process, often is burned for its energetic value to generate process heat and electricity. However, lignin is the largest source of renewable aromatics and, as such, strategies to valorize lignin are required for a bio-refinery to produce aromatic-derived chemicals currently derived from fossil fuels.

"Lignin-first" strategies have been developed to extract value from lignocellulosic biomass in the form of commodity chemicals and liquid fuels from both the lignin and polysaccharide fractions.[1] These strategies liberate the lignin in "native-like" form from the plant cell wall and convert it through catalytic processing[2] or by protection-group chemistry[3] before it can be degraded under the deconstruction processing conditions.

In our previous work, we developed a lignin-to-bioproduct processing chain to produce 2-pyrone-4,6-dicarboxylic acid (PDC) through microbial funneling of the phenolic monomers obtained by catalytic depolymerization of isolated lignin.[4] We used chemical and biological upgrading in tandem to extract greater value from the lignin fraction by converting lignin-derived aromatics to PDC. In this previous work, we first isolated lignin from lignocellulosic biomass under mild reaction conditions using gamma-valerolactone (GVL) and water as the solvent system and dilute sulfuric acid as a catalyst. In this scheme, lignin degradation was minimized during the biomass fractionation by the solvent system, as well as low process temperatures (<120 °C). The technoeconomic analysis of the processing chain showed the cost of lignin isolation to have a large impact on the minimum-selling-price (MSP) of PDC.

In this work, we modified the "lignin-to-bioproduct" processing chain by eliminating the costly biomass fractionation step. In the new scheme, a "biomass-to-bioproduct" processing chain, we combine biomass fractionation and catalytic depolymerization of lignin into a single step, while simultaneously preserving the structure of the polysaccharide fraction. We demonstrate that the lignin fraction of poplar wood can be successfully depolymerized by hydrogenolysis over a Pd/C catalyst into a mixture of monomeric and oligomeric phenolic compounds. Using this strategy, we were able to increase the monomer/oligomer product yield on a per kg of biomass basis. We further show that, using an engineered strain of *Novosphingobium aromaticivorans* DSM12444, this complex mixture of aromatic compounds containing syringyl, guaiacyl, and *p*-hydroxyphenyl aromatics can be upgraded to PDC. Moreover, we show that PDC can be extracted from the culture broth with a simple separation and purification step (*e.g.*, precipitation with sodium chloride). Furthermore, the sugar stream was subjected to enzymatic and microbial digestion to liquid fuels to produce value-added products from both the lignin (phenolics) and polysaccharide fractions.

Combining these improvements over the previous process we demonstrate that the most expensive parts of the lignin-to-bioproduct (PDC) processing chain can be improved by using a lignin-first biomass-to-bioproduct processing chain. Most importantly, we show that tandem processes utilizing both chemical and biological upgrading can significantly improve the upgrading of a complex feedstock such as lignocellulosic biomass.

**References/Publications**


1. Abu-Omar, M.M.; Barta, K.; Beckham, G.T.; Luterbacher, J.S.; Ralph, J.; Rinaldi, R.; Román-Leshkov, Y.; Samec, J.S.M.; Sels, B.F.; Wang, F., Guidelines for performing lignin-first biorefining. *Energy & Environmental Science* (2021), **14**(1), 262-292.
2. Luterbacher, J.S.; Azarpira, A.; Motagamwala, A.H.; Lu, F.; Ralph, J.; Dumesic, J.A., Lignin monomer production integrated into the γ-valerolactone sugar platform. *Energ Environ Sci* (2015), **8**(9), 2657-2663.
3. Shuai, L.; Questell-Santiago, Y.M.; Luterbacher, J.S., A mild biomass pretreatment using γ-valerolactone for concentrated sugar production. *Green Chem* (2016), **18**(4), 937-943.
4. Perez, J.M.; Sener, C.; Misra, S.; Umana, G.E.; Coplien, J.; Haak, D.; Li, Y.; Maravelias, C.T.; Karlen, S.D.; Ralph, J.; Donohue, T.J.; Noguera, D.R., Integrating lignin depolymerization with microbial funneling processes using agronomically relevant feedstocks. (Submitted).



**Funding Statement:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409.

**Title:** Analysis of SI Engine Alternative Fuels for Mixture Formation Behavior Related to Particulate Matter Formation with a Representative Gasoline Surrogate: Application to 2-Methyl-3-Buten-2-ol (Methyl Butenol, MBO)

**Authors:** Andrea Shen,[1]* (ashen3@wisc.edu) and **David Rothamer**[1]

**Institutions:** [1]Department of Mechanical Engineering and DOE Great Lakes Bioenergy Research Center, University of Wisconsin – Madison, Madison

**Project Goals:**

The goal of this project is to develop a screening process based on component properties to determine whether the performance of new potential biofuels, such as 2-methyl-3-buten-2-ol (methyl butenol, MBO), when blended with gasoline, meet or exceed the performance of biofuels already established in the market, such as ethanol and isobutanol (IBA). The screening process will include properties important to spark-ignition direct-injection (SIDI) engines related to mixture formation, combustion, and particulate matter (soot) emissions. To assess biofuels that are available in limited quantities the screening process must be done computationally to minimize time and resources. To computationally screen biofuels for properties when blended with a gasoline, a representative gasoline surrogate was developed to accurately model and estimate properties of gasoline-biofuel blends.

**Abstract Text:**

Ethanol and isobutanol (IBA) have seen significant study of their engine performance when blended with gasoline. Recently, the terpene compound 2-methyl-3-buten-2-ol (methyl butenol, MBO) was identified as a potential gasoline blending component that may have favorable properties. To assess this, we consider biofuels blended with a gasoline blendstock for oxygenated blending (BOB), a gasoline with lower octane number that is used specifically for blending with oxygenated biofuels. Early efforts focus on estimation of gasoline-biofuel blend properties determined that the computational gasoline surrogate composition had a significant impact on these estimates. To address this, a new surrogate formulation methodology was developed that targeted matching of the chemical composition of the gasoline as well as the distillation curve (volatility), research octane number (RON) and motored octane number (MON) (autoignition), and hydrogen-to-carbon ratio (H/C) (flame speed and adiabatic flame temperature). With the new methodology, the surrogate composition matches the reference gasoline with less than 1.5% error in n-alkanes, iso-alkanes, and aromatics mole fractions. The error for other quantities is also low with average root mean-squared error for the distillation curve of 3.1% and errors for RON, MON, and H/C all less than 1.5%. The new surrogate formulation method also allows for identification of additional surrogate compositions that may have similar overall error but match particular properties better that were not included directly in the optimization process.

The reminder of this work focuses on screening properties of MBO related to in-cylinder mixture preparation in spark-ignition direct-injection (SIDI) engines that may impact particulate matter (PM) formation. The proposed screening parameters include the equilibrium distillation curve (EDC) and standard enthalpy requirement (SER). These parameters were calculated for MBO-, IBA-, and ethanol-gasoline blends at fuel matched oxygen weight percentages of 4, 7.5, and 11%. The results of the distillation curve calculation show an increase in aromatic component concentration at the 90% distillation point for ethanol at all blend percentages. For MBO and IBA the mole fraction of oxygenate increases at the 90% distillation point with increasing oxygenate weight percent, whereas for ethanol it remains almost constant. A redefined SER calculation showed that the energy needed to vaporize the fuel increases significantly with oxygen weight percent for addition of IBA and ethanol, while it remains almost unchanged for MBO. Combined, the results help explain the increased PM emissions seen from SIDI engines operating on ethanol-gasoline blends under certain operating conditions. The results also indicate that MBO will likely not suffer from this issue and may reduce PM emissions in general.

**References/Publications**


1. Shen, A., Sakai, S. & Rothamer, D. "Parameters for screening SI engine alternative fuels for mixture formation behavior related to particulate matter formation: application to 2-methyl-3-buten-2-ol (methyl butenol, MBO)." in Spring Technical Meeting of the Central States Section of the Combustion Institute. (2020).
2. Shen, A. “Parameters for screening SI engine alternative fuels for mixture formation behavior related to particulate matter formation: Application to 2-methyl-3-buten-2-ol (methyl butenol, MBO).” ACS Fall 2020 Virtual Meeting & Expo, 17-20 Aug. 2020, online.


**Funding Statement:**


This material is based upon work supported in part by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409.

**Title:** Utilization of Lignocellulosic Biofuel Conversion Residue by Diverse Microorganisms

**Authors:** Caryn S. Wadler[1,5*] (cwadler@wisc.edu), John F. Wolters[2,3,5], Nathaniel W. Fortney[2,5], Kurt O. Throckmorton[1,5], Yaoping Zhang[2,5], Caroline R. Miller[2,3,5], Rachel M. Schneider[2,3,5], Evelyn Wendt-Pienkowski[1,5], Cameron R. Currie[1,5], Timothy J. Donohue[1,2,5], Daniel R. Noguera[2,4,5], Chris Todd Hittinger[2,3.5], and **Michael G. Thomas**[1,5]

**Institutions:** [1]Department of Bacteriology, University of Wisconsin-Madison, Madison, [2]Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, [3]Laboratory of Genetics, Center for Genomic Science Innovation, J. F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, [4]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, [5]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison

**Project Goals:** To offset the costs of biofuel production, we characterized components of lignocellulosic biofuel production waste (Conversion Residue) available for microbes to metabolize into valuable bioproducts and identified *Streptomyces* and yeast species to serve as chassis for future genetic engineering towards this goal.

**Abstract:** Conversion Residue (CR) is the material remaining after deconstructed lignocellulosic biomass is subjected to microbial fermentation and distillation to remove the biofuel. Current methods of switchgrass hydrolysate fermentation to bioethanol by *Zymomonas mobilis* leave behind about 60% of the organic material from the hydrolysate as a combination of raw plant material, treatment residues, unfermented sugars, bacterial waste products, and cell debris. We analysed several batches of CR generated from AFEX- and enzyme-treated switchgrass hydrolysate fermented by *Z. mobilis* and distilled to remove bioethanol to determine what components could be utilized for a second round of microbial processing to alternate bioproducts. Through a combination of chemical oxygen demand (COD), HPLC, and GC-MS-based assays, we determined that the major components of CR are oligomeric and monomeric sugars from cellulose and hemicellulose and other carbon containing metabolites. We then tested 71 *Streptomyces* species, 163 yeast species, and an aerobic microbial consortium derived from a wastewater treatment plant for their ability to grow on CR. Many of the *Streptomyces* and yeast species were able to grow in CR, with some species capable of utilizing over 40% of the soluble COD. For comparison, the microbial community was able to metabolize about 70% of the soluble COD. HPLC analysis showed that most individual microbes and the community preferentially utilized the monomeric sugars in CR for growth, although there was evidence of further breakdown of oligomeric sugars. These analyses allowed us to identify strains that are good candidate chassis for genetic engineering towards the production of valuable bioproducts. We also developed a synthetic conversion residue (SynCR); a defined medium designed to mimic CR. SynCR allowed us to examine how individual components of CR affect microbial growth and will be a valuable tool for future production of bioproducts.

**Funding Statement:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409 and DE-FC02-07ER64494.

**Title: Elucidation of the Roles of Diazotrophic Endophyte Communities in Promoting Productivity and Resilience of *Populus* through Systems Biology Approaches**

**Authors: Sharon L. Doty[1]* (sldoty@uw.edu),** Andrew W. Sher[1], Robert Tournay[1], Darshi Banan[1], Soo-Hyung Kim[1], Adam Deutschbauer[3], Young-Mo Kim[2], Nathalie Munoz[2], Kim Hixson[2], James Moran[2], Amy Zimmerman[2], and Amir H. Ahkami[2]

**Institutions:** [1]University of Washington, Seattle; [2] Environmental Molecular Sciences Laboratory (EMSL), Pacific Northwest National Laboratory, Richland, WA; [3]Lawrence Berkeley National Lab, Berkeley, CA

**Project Goals:** The overall project goal is to move toward an understanding of the holobiont, how plants and the microbial community within them interact in ways that promote the productivity of the whole. Integration of plant physiology data with the molecular plant-microbe interactions (multi-omics) data from greenhouse and field experiments will allow us to develop a systems-level understanding of the genetic and molecular basis for diazotrophic endophytic mutualism in *Populus*. This deeper level of understanding of the plant responses will guide construction of microbial communities in order to optimize the impacts of bioinoculants for environmental sustainability of bioenergy crops.

**Abstract text:** Poplar trees are important feedstocks for bioenergy and ecosystem services, but more efficient and resilient growth is essential for sustainability. Some of the micro-organisms which make up the wild poplar microbiome can help poplar grow by providing the required nutrients of nitrogen and bioavailable phosphorus. In addition to increasing nutrient acquisition, the micro-organisms may also promote plant tolerance to other environmental stresses including drought. Previously we demonstrated that adding microorganisms from wild poplar to a variety of other plants including grasses, conifers, and cultivated poplar, increased the health and growth of these plants under nutrient limitation and drought. Since the start of this grant in autumn 2020, a suitable field site in a nutrient- and water-controlled site near Prosser, Washington was prepared for planting in spring 2021. However, after a record-breaking heat wave and physiological impacts on the plants, we had to abort the 2021 planting, re-start the experiment in August, and conduct greenhouse testing of the bio-inoculants instead in preparation for a spring 2022 field planting. Growth, physiological parameters and nitrogen levels of the inoculated and control greenhouse poplar trees are being assessed. Random barcoded TnSeq experiments to identify bacterial endophyte genes required for phosphate solubilization and nitrogen fixation were conducted. An additional set of diazotrophic strains was isolated from wild poplar roots that significantly increased growth of *in vitro* grown poplar plants in nitrogen-free media. Genomic analysis of the new endophyte strains are in progress. A new consortium of strains optimized for nitrogen fixation and poplar growth promotion will then be added to the greenhouse poplar prior to field planting. In parallel, methodology for proteomics and metabolomics of a $^{15}N$-labeled diazotrophic endophyte strain is underway. Preliminary analysis of peptides conservatively identified nearly 200 peptides with substantial $^{15}N$ enrichment. Optimization of the analytical pipeline for $^{15}N$ proteomics is on-going in preparation for experiments with the finalized consortium of strains. This will enable elucidation of the molecular microbe-microbe interactions for optimal nitrogen fixation. This first year of the grant

project is setting the stage for in-depth investigations of the mechanisms by which bacterial endophytes promote the growth and health of trees for sustainable bioenergy crop production.

**Funding statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021137. Pacific Northwest National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-76RL01830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 76183.

**Microbial regulation of soil water repellency to control soil degradation**

Emily Boak[1*] (emily_boak@lanl.gov), **John Dunbar[1]**, Marie E. Kroeger[1]

[1]Bioscience Division, Los Alamos National Lab, Los Alamos, New Mexico

https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-microbial-carbon.php

**Project Goals**

1. Determine the influence of microbial community composition on the severity and persistence of SWR
2. Identify the metabolites differentially produced from microbial activity in hydrophilic and hydrophobic soils.

**Abstract**

Soil water repellency (SWR) is a major cause of global soil degradation and a key agricultural concern with an estimated 44 billion dollars lost in the US annually to offset erosion[1-2]. This phenomenon affects soils globally in both natural and cultivated lands[3]. The effects of climate change will likely exacerbate SWR emphasizing the need to understand the mechanisms driving SWR development and persistence. The importance of the soil microbiome in the development and persistence of SWR has been postulated for decades, but very limited research has been conducted into its role. To address this gap-in-knowledge, we investigated the direct effect of microbial community composition and activity on the development and persistence of SWR in model soils using a "common garden" experiment design. We inoculated microcosms containing model soils (sterile sand, blue grama plant litter, and bentonite clay) with 15 different microbial communities. Respiration was measured throughout the experiment and a soil water drop penetration test was conducted prior to destructive sampling on days 30, 45, and 89. Six microbial communities were found to consistently produce either a hydrophobic (3) or hydrophilic (3) phenotype. We extracted DNA for metagenomics and polar and non-polar metabolites for mass spectrometry.

We found that microbial community composition does affect the development of SWR. Although preliminary, there is some data to support that there is lower microbial diversity in hydrophobic communities, though specific taxa have yet to be identified as involved in SWR development. Preliminary analysis of the metabolomic data shows a significant difference in the functionalities displayed by hydrophilic vs. hydrophobic communities. Somewhat expected, the number of metabolites associated with lipid transport and metabolism is higher in hydrophobic communities, as well as metabolites associated with carbohydrate and coenzyme transport and metabolism. Preliminary data also indicates some differences between time points by community type, with a higher number of metabolites associated with defense mechanisms and signal transduction on day 45 within hydrophobic communities, but not day 89. These initial findings suggest that microbial communities are a key missing factor in this global phenomenon. Further research is necessary to understand the mechanisms driving the development and persistence of SWR *in situ* and under changing environmental conditions and during different stages of colonization. Further research into microbial control of SWR has the

potential to lead to the development of new mitigation strategies of SWR through microbial inoculations in degraded lands.

References:


1. DeBano LF (2000) Water repellency in soils: a historical overview. J Hydrology. 231-232:4-32.
2. Telles TS, Guimarães MdF, Dechen SCF (2011) The costs of soil erosion. R. Bras. Ci. Solo, 35:287-298.
3. DeBano LF (1981) Water Repellent Soils: a state-of-the-art. In *General Technical Report PSW-46* (pp 1-17). Berkeley, CA: Pacific Southwest Forest and Range Experiment Station.



*Funding Statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (*2018SFAF255) and the Joint Genome Institute Community Science Program New Investigator Grant (506588).

**Title:** Constructing microbial networks from Genome Scale Metabolic Models.

**Authors:** James D. Brunner[1,2] (jdbrunner@lanl.gov), Michaeline Albright[1], **John Dunbar**[1]

**Institutions:** [1]Biosciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico; [2]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico

**Project Goals:**

1. Construct species-species interaction networks from genome-scale metabolic models within an environmental context.
2. Investigate the stability of species-species interactions as environmental context is changed.
3. Produce accurate predictions of stable microbial community composition outcome after perturbation of community composition and environmental context.

**Abstract:** Interactions between microbial species have become a popular basis for dynamical and predictive models that seek to understand how microbial populations organize and affect their environment. However, the poor predictive power of pairwise network models with static parameters suggests that interactions are in turn affected by their environment, and are therefore context dependent. Genome scale metabolic modeling offers techniques to understand the behavior of a microbe in a specific metabolic environment. One method for combining genome scale models into a community model, called community dynamics flux balance analysis, offers a way to create a dynamical system which couples the microbial community with its environment. However, this system suffers from complexity and non-smooth behavior, and does not directly reveal any interaction between microbes, making manipulation of the community difficult. We use the community dynamic flux balance analysis system as a basis from which to derive the implied interactions between species, and so create an environmental-context dependent networks of species interactions. These networks can be used to study how interactions between microbes change with a changing environment, and provide predictions of microbial community changes, including engraftment of invading taxa.

**Funding Statement:** This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2018SFAF255). We gratefully acknowledge the support of the U.S. Department of Energy through the LANL/LDRD Program and the Center for Non-Linear Studies for this work.

**Microbially-Drive Carbon Flow Persists During Surface Litter Decomposition**

**Authors:** Marie E. Kroeger[1*] (mkroeger@lanl.gov), Dennis Suazo[1], John Heneghan[2], La Verne Gallegos-Graves[1], Thomas Yoshida[3], Michaeline B.N. Albright[1], Sanna Sevanto[2], **John Dunbar**[1]

**Institutions:** [1] Bioscience Division, Los Alamos National Lab, Los Alamos, New Mexico, [2]Earth and Environmental Science Division, Los Alamos National Lab, Los Alamos, New Mexico, [3]Chemical Diagnostics and Engineering, Los Alamos National Lab, Los Alamos, New Mexico

**Website URL:** https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-microbial-carbon.php

**Project Goals:** The two main goals of this project were (1) to determine if microbially-driven carbon cycling variation observed in short-term litter decomposition persists over longer timescales and (2) identify the temporal dynamics of microbial traits associated with disparate carbon flow.

**Abstract**

In terrestrial ecosystems, products of microbially-driven plant litter decomposition are major inputs to the soil organic carbon pool, a key carbon sink. While climate and litter quality impact litter decomposition rates across sites, large intra-site variation in litter mass loss appears to be microbially-driven. During litter decomposition, microorganisms can either mineralize the carbon compounds from plant litter to produce $CO_2$ through respiration or transform the carbon into other organic molecules. Thus, manipulation of microbial community composition has the potential to alter the ratio of carbon directed toward the atmosphere as $CO_2$ or funneled to the soil as dissolved organic carbon (DOC), which has the possibility to be stored for longer periods of time. It is clear that microbial composition has a significant effect on litter decomposition; however, it is not understood the underlying mechanisms by which microbial community composition alters carbon cycling at any given point in time, much less over the full cycle of litter decomposition.

To address the gap-in-knowledge we conducted two yearlong experiments, one in the laboratory under controlled environmental conditions and one in the field, where we inoculate non-sterile blue grama plant litter with a 1:40 serial dilution of different soil microbial communities. We measured $CO_2$flux and DOC to assess the effect of microbial community composition on carbon flow. We collected blue grama plant litter over the year for microbial community analyses (amplicon, metagenomic, metatranscriptomic sequencing). Overall, we found that microbially-driven DOC variation persists over a year and changing environmental conditions. Furthermore, we observed no significant difference in respiration between low and high DOC microbial communities. This research is still on-going as we wait for microbial sequencing data, DOC characterization data using GC-MS and FTICR-MS, and super resolution FISH microscopy, but we anticipate that the phenotypic differences observed in both laboratory and field conditions will allow us to identify microbial traits driving carbon flow during litter decomposition.

**Funding Statement:** This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant, grant no. 2018SFAF255.

**Nitrogen Availability Strongly Affects Carbon Cycling by Sub-Surface Microbial Communities**

M. Rae DeVan[1], Rachel Hestrin[2], John M. Dunbar[1], Michaeline B. N. Albright[1], M. Francesca Cotrufo[3], Sanna Sevanto[4*] (sanna@lanl.gov)

[1] Bioscience Division, Los Alamos National Lab, Los Alamos, New Mexico
[2] Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California
[3] Department of Soil and Crop Science, Colorado State University, For Collins, Colorado
[4] Earth and Environmental Science Division, Los Alamos National Lab, Los Alamos, New Mexico

https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-microbial-carbon.php

**Project Goals**

1. Determine the microbially-driven variation in carbon cycling of sub-surface microbial communities.
2. Quantify the impacts of nitrogen addition on carbon cycling by sub-surface microbial communities

**Abstract**

Soils store more carbon than plants and the atmosphere combined, and the ultimate fate of this carbon is tied to subsurface microbial communities. Yet, our fragmented understanding of how subsurface communities influence carbon cycling prohibits accurate estimates of the fate of carbon produced on the surface and future carbon storage. Because of the tight link between nitrogen and carbon cycles, nitrogen deposition has potential to drastically alter subsurface microbial composition and function. In this study, we used a common garden microcosm experiment to assess the range of subsurface microbial driven variation in carbon and nitrogen pools during root litter decomposition under fertilized ($NH_4NO_3$) and unfertilized conditions. We demonstrate that subsurface microbial community composition can create large variation in dissolved organic carbon (DOC, ~ 6x) and total nitrogen (TN, ~7x), and two-fold variation in $CO_2$. Nitrogen addition altered the balance of DOC, TN, and $CO_2$ in ways that were specific to the origin of microbial inoculum. These communities differed drastically in their ability to use the added nitrogen, which when consumed, was related to higher levels of $CO_2$ and lower levels of DOC compared to communities from the same origin without added N. Thus, we conclude that differences in subsurface microbial community composition and their response to nitrogen deposition could have far-reaching impacts on ecosystem function and feedbacks to global nutrient cycles.

*Funding Statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (*2018SFAF255).

**Identifying Data-Driven Gene Targets to Control Bacterial Fitness**

Shara Balakrishnan[1]* (sbalakrishnan@ucsb.edu), Adam Deustchbauer[2], Enoch Yeung[3], and **Rob Egbert**[4]

**Institutions:** [1]Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, California; [2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Lab, Berkeley, California; [3]Department of Mechanical Engineering, University of California Santa Barbara, Santa Barbara, California; [4]Earth & Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington

**Website:** https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

**Project Goals:** The Pacific Northwest National Laboratory Persistence Control Scientific Focus Area aims to gain a fundamental understanding of factors governing the persistence of engineered microbial functions in rhizosphere environments. From this understanding, we will establish design principles to control the environmental niche of native rhizosphere microbes. In our first funding period, we are examining the efficacy of genome reduction and metabolic addiction to plant root exudates in environmental isolates as persistence control strategies using the bioenergy crop sorghum and defined microbial communities as a model ecosystem. Effective persistence control will lead to secure plant–microbe biosystems that promote stress-tolerant and highly productive biomass crops.

**Abstract:** An important factor in plant growth is the interdependent relationship between the plant and the soil microbiome. By genetically modifying the microbes, we can modulate the soil composition to promote robust plant growth in low-resource environments. A microbe with engineered functions, when introduced into a new environment, interacts with the native microbiome, and subjects itself to competitors and predators [1]. A first step towards introducing engineered microbial functions to promote plant growth is to control the environmental persistence of the microbe of interest in the target environment. Our objective is to develop growth harness actuators, genetic devices that can control gene expression to regulate the microbial growth. In this poster, we present how to identify "fitness genes" for a bacterium in a target environment using time-series RNAseq data along with time series growth data. We propose a novel data-driven sensor fusion technique that combines dynamical systems theory and machine learning to discover dynamic genotype-to-phenotype models that can simulate single as well as combinatorial gene knockouts to identify optimal sets of genes corresponding to the phenotype. We further validate these genes using a novel time-series transposon knockout assay.

Current approaches to identify genes for a phenotype of interest hinge on differential RNA expression of genes either across time or across media conditions; the genes that exhibit maximal differential mRNA expression across conditions are the important genes for that condition. These approaches only use the mRNA expression data and check for individual phenotypes. We represent the genotype-to-phenotype dynamics as a dynamical state-space model by assuming gene expression to represent the state of the system and phenotype to represent the output of the system. The state-space model contains the state equation which captures the gene dynamics and the output model captures the mapping of instantaneous gene activity to the growth, the phenotype of interest. This predictive modeling framework can be extended to fuse other types of data like metabolomics, proteomics, fluorescence, and microscopic data either as a

state or output depending on the availability of data and the problem being solved. We developed an algorithm called output constrained deep dynamic mode decomposition (OC-DeepDMD) algorithm which uses multilayer feedforward neural networks to identify a high dimensional linear Koopman model of a relatively lower-dimensional nonlinear system such that the state-space model becomes linear in both the state and output equation [2]. Using this model, we simulate the effect of single-gene knockouts on growth output.

To validate the model predictions, we do Random Barcoded Transposon Sequencing (RB-TnSeq) experiments which is an alternate method to identify genes that relate to fitness in the media conditions used in the experiment [3]. RB-TnSeq employs a pool of single-gene knockout mutants of a single strain via transposon insertions with unique genetic barcodes and captures the fitness of the individual strains in various media conditions; the more negative the strain fitness value in a certain media condition, the more important the gene for that condition. Typically, RB-TnSeq experiments are done by considering an initial and a final time point. We extend the experiment to include multiple intermediate time points and compute the fitness curves as a function of time for each mutant and establish a benchmark for the comparison of the fitness predictions obtained for the single-gene knockout mutants from the Koopman models.

In this work, we consider the growth of the soil bacterium *Pseudomonas putida* in R2A media with varying concentrations of two nutrients - glucose as a carbon source and casein hydrolysate as a source for amino acids. By observing the growth curves of *P. putida* under varying concentrations of the two nutrients, we selected the condition under which a maximum growth rate (MAX condition) is observed and the negative control (NC) condition in which the nutrients are absent. We performed time-series RNA sequencing and RB-TnSeq experiments for the MAX and NC conditions while obtaining optical density (OD) growth measurements. We used the RNAseq and OD data in the OC-DeepDMD algorithm to learn a Koopman operator representation of the state-space model. We simulated fitness with single-gene knockouts and validated our model predictions using the RB-TnSeq data.

In summary, we propose to learn dynamic genotype-to-phenotype models using the OC-DeepDMD algorithm, predict the fitness of single-gene knockouts and experimentally validate them using time-series RB-TnSeq experiments. In the future, we plan to simulate combinatorial gene knockout and identify optimal gene targets to control microbial fitness.

**References/Publications**

[1] Wang, F. and Zhang, W., 2019. Synthetic biology: recent progress, biosafety and biosecurity concerns, and possible solutions. *Journal of Biosafety and Biosecurity*, pp.22-30

[2] Balakrishnan, S., Hasnain, A., Egbert, R. and Yeung, E., 2021. The Effect of Sensor Fusion on Data-Driven Learning of Koopman Operators. *arXiv preprint arXiv:2106.15091*.

[3] Wetmore, Kelly M., et al. "Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons." *MBio* 6.3 (2015): e00306-15.

**Funding statement:** *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design Science Focus Area "Persistence Control of Engineered Functions in Complex Soil Microbiomes" through a subcontract to the University of California Santa Barbara. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.*

**Title:** Catabolism of Methyl-3-(4-hydroxyphenyl)propionate (MHPP), a Model Substrate for Metabolic Addiction with Unexpected Implications for *p*-Coumaric Acid Catabolism in *Pseudomonas flourescens*

**Authors:** Joshua Elmore[1*] (Joshua.Elmore@pnnl.gov), Ritu Shrestha,[1] Andrew Wilson,[1] Andrew Frank,[1] **Robert Egbert**[1]

**Institutions:** [1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354.

**Website:** https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

**Project Goals:** The Pacific Northwest National Laboratory Persistence Control Scientific Focus Area is focused on developing fundamental understanding of factors governing the persistence of engineered microbial functions in rhizosphere environments. From this understanding, we will establish design principles to control the environmental niche of native rhizosphere microbes. In our first funding period, we are examining the efficacy of genome reduction and metabolic addiction to plant root exudates in environmental isolates as persistence control strategies using the bioenergy crop sorghum and defined microbial communities as a model ecosystem. Effective persistence control will lead to secure plant–microbe biosystems that promote stress-tolerant and highly productive biomass crops.

**Abstract:** Metabolic addiction is a tool in which survival of engineered microorganisms is restricted in the absence of an essential metabolite. Engineered addiction to bioenergy crop-specific root exudate compounds has great potential to enable the responsible use of engineered plant growth-promoting rhizobacteria in the environment, while preventing uncontrolled spread of such organisms in the environment. To date, methyl-3-(4-hydroxyphenyl)propionate (MHPP) has only been reported to be present in root exudates produced by *Sorghum bicolor* and *Brachiara pasturis*, where they are produced as biological nitrification inhibitors, and thus is an excellent model carbon source for demonstrating metabolic addiction. *Pseudomonas fluorescens* SBW25 can utilize MHPP as a sole carbon source, but the pathway for its catabolism is currently unknown. We performed a combination of RB-TnSeq, RNA-Seq, and individual gene deletions to elucidate a catabolic pathway for MHPP. We identified enzymes involved in MHPP catabolism, and now propose a catabolic pathway that funnels MHPP into the pathway responsible catabolism of *p*-coumaric acid. The *p*-coumaric acid catabolic pathway in *Pseudomonas putida* KT2440 is exceptionally similar to its analog in SBW25 at the levels of protein sequence, gene sequence, and even gene synteny. Thus, it was highly surprising that we found that several enzymes that appear to be essential for phenylpropanoid catabolism in *Pseudomonas putida* KT2440 are non-essential in SBW25. In fact, at least one appears to equally share its function with a second enzyme – with deletion of the genes encoding either enzyme resulting in a similar moderate growth defect when *p*-coumaric acid is the sole carbon source. This underscores the challenges with using genomic and phenotypic information from even

closely related organisms alone to predict the roles of orthologs in other organisms. Ultimately, the approach we piloted to identify the MHPP pathway will be used to discover pathways for additional root exudates. This will lead to identification of catabolic pathways for compounds that are fully unique to sorghum (e.g. *s*orgoleone), and thus will enable the assessment of metabolic addiction as a persistence control strategy in soil and plant ecosystems.

**Funding statement:** *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design Science Focus Area "Persistence Control of Engineered Functions in Complex Soil Microbiomes". A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by BER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.*

**Title:** Simplified Microbial Communities Provide Diverse Complementation Potential for Genome-reduced Microorganisms

**Authors:** Citlali Fonseca[1*] (fonseca-garcia@berkeley.edu), Joshua Elmore[2], Ryan McClure[2], Henri Baldino[2], Ritu Shresthra[2], Andrew Wilson[2], Andrew Frank[2], Pubudu Handakumbura[3], Devin Coleman-Derr[1], **Robert Egbert[2]**

**Institutions:** [1]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, United States. [2]Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA, 99352. [3]Environmental Molecular Sciences Laboratory, Richland WA, 99352.

**Website:** https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

**Project Goals:** The Pacific Northwest National Laboratory Persistence Control Scientific Focus Area aims to gain a fundamental understanding of factors governing the persistence of engineered microbial functions in rhizosphere environments. From this understanding, we will establish design principles to control the environmental niche of native rhizosphere microbes. In our first funding period, we are examining the efficacy of genome reduction and metabolic addiction to plant root exudates in environmental isolates as persistence control strategies using the bioenergy crop sorghum and defined microbial communities as a model ecosystem. Effective persistence control will lead to secure plant–microbe biosystems that promote stress-tolerant and highly productive biomass crops.

**Abstract Text:** The beneficial interactions between plants and rhizosphere microorganisms have been well documented over the past several years. More recently, researchers have explored ways of engineering microbial species to enhance these interactions and further promote plant growth. However, the introduction of engineered organisms to soil presents questions about how they can be restricted to their intended roles and not persist outside their functional or geographic niches. Genome reduction to remove key metabolic or stress-response pathways is one strategy to achieve this goal of niche restriction but its efficacy in the presence of genetic or metabolic complementation potential from the complex soil microbiome is not known.

To investigate complementation dynamics for engineered soil microbes, we have developed several Synthetic Communities (SynComs) that were derived from a native soil microbiome harvested from sorghum fields. We established these SynComs directly from complex soil samples (top-down) or from mixtures of soil isolates (bottom-up) and collectively they represent defined microbial communities that provide a palette for genetic or metabolic complementation of genome-reduced species. Using growth on agar plates supplemented with sorghum root-exuded metabolites we reduced native soil microbiome complexity and developed SynComs comprised of tens of species with relatively even representation and containing a number of key soil microbial taxa. We show that these SynComs are initially dynamic in their development before stabilizing and falling in complexity and richness. Of the methods tested for cold storage we show that glycerol, and in some cases lyophilization, enabled reconstitution of communities with the least dissimilarity from the original SynComs. Finally, by growing SynComs across a range of single carbon sources we highlight their complementation potential. Among several C sources tested arginine, adenosine, mannitol, and hydroxycinnamic acid led to SynComs that maintained the most

diversity, suggesting that in soils with these C sources these SynComs could serve as wide ranging complementation sources. These experiments reveal how synthetic microbial communities can be generated and stored as tools for microbiome engineering and begin to show the species and carbon utilization pathways that represent the highest potential for metabolic complementation of genome-reduced species. Future experiments will focus on incubation of genome-reduced species with SynComs to characterize their complementation effects, pathways and kinetics to improve our ability to control the persistence of engineered bacterial functions in native soil systems.

**Funding statement:** *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design Science Focus Area "Persistence Control of Engineered Functions in Complex Soil Microbiomes". A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by BER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.*

**Title:** Spatially-resolved Multi-omics Analyses Reveal Key Taxa Responding to Root Depth in the Sorghum Rhizosphere

**Authors:** Pubudu Handakumbura* (Pubudu.Handakumbura@pnnl.gov), Ryan McClure, Albert Rivas-Ubach, Tamas Varga, Anil Battu, Andrew Wilson, Thomas Wietsma, Janet Jansson, Christer Jansson, **Robert Egbert**

**Institution:** Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory (PNNL), Richland, WA 99354.

**Website:** https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

**Project Goals:** The Pacific Northwest National Laboratory Persistence Control Scientific Focus Area aims to gain a fundamental understanding of factors governing the persistence of engineered microbial functions in rhizosphere environments. From this understanding, we will establish design principles to control the environmental niche of native rhizosphere microbes. In our first funding period, we are examining the efficacy of genome reduction and metabolic addiction to plant root exudates in environmental isolates as persistence control strategies using the bioenergy crop sorghum and defined microbial communities as a model ecosystem. Effective persistence control will lead to secure plant–microbe biosystems that promote stress-tolerant and highly productive biomass crops.

**Abstract:** The rhizosphere, the zone of soil influenced by plant roots, is closely associated with a complex microbial community of bacteria, archaea, viruses, and fungi[1]. Using a correlative surface imaging approach with *Brachypodium distachyon*, a genomics model for grasses[2], we have demonstrated that the root surface is metabolically heterogeneous with hot spots for bacterial attachment[3]. Recently, we developed a 3D root cartography platform to map root exudate metabolites and microbes onto a 3D image of the root, generated by X-ray computed tomography (XCT)[4]. The aims of this ongoing experiment are to 1) map the distribution of exuded metabolites and the soil microbiome in the 3D space of the sorghum root and rhizosphere, 2) examine if there are hotspots for microbial colonization, and 3) determine if these hotspots correlate with exudation of specific metabolites.

We have employed this multi-omics and imaging platform to collect spatially resolved root morphology, metabolite, and microbial species data for the bioenergy crop *Sorghum biocolor* (L.) Moensch. We collected XCT root profiles, metabolomics data, and 16S microbial profiles for three plants grown in soil collected from the Prosser, WA field site administered by the Soil Microbiome Science Focus Area at PNNL. While an integrated analysis linking spatial root morphology to metabolomic and taxonomic signatures remains in progress, we applied a network approach to the 16S taxonomic data to gain a more complete view of the role of individual taxa within the root microbial. Using amplicon data from three plants and across a profile of eight root depths we inferred a species co-abundance network and interrogated this network to identify which taxa occupied important and central positions and which taxa may shift their importance specifically as a function of depth. In a complete network of all data

Proteobacteria, Crenarchaeota, Bacteroidota and Firmicutes shows the greatest number of links to other taxa while Bdellovibrionota and Myxococcota occupied positions as bridge taxa, linking other larger groups of species. When removing data from the lowest depths and remaking the network we found that certain taxa exhibited drops in their centrality suggesting that they may be crucially important specifically at lower depths. Taxa that may be critical at lower depths included Chloroflexi and Gemmatimonadota. In contrast, Myxococcota showed an increase in its centrality when depth data was removed, suggesting it may be more important at shallow depths. This network analysis will be critical in future studies that seek to better understand both beneficial and antagonistic interactions between root microbiomes and how these interaction networks shift as a function of depth.

We anticipate this research will enable testing our hypothesis of community complementation for genome-reduced and, thus, niche-restricted biocontainment strains. Specifically, we hypothesize that species with the capacity for genetic or metabolic complementation are more likely to complement lost functions of niche-restricted strains when physically proximal in the rhizosphere. To test this hypothesis, we will assess the abundance of engineered strains directly in the sorghum rhizosphere that contains simplified microbiome communities.

**References/Publications**


1 Ahkami, A., White, I., R. A., Handakumbura, P. P. & Jansson, C. Rhizosphere Engineering: Enhancing Sustainable Plant Ecosystem Productivity in a Challenging Climate; *Rhizosphere* **3**, 233-343 (2017).

2 Jansson, C., Vogel, J., Hazen, S., Brutnell, T. & Mockler, T. Climate-smart crops with enhanced photosynthesis. *Journal of Experimental Botany* **69**, 3801-3809 (2018).

3 Liu, W. *et al.* Correlative surface imaging reveals chemical signatures for bacterial hotspots on plant roots. *Analyst* **145**, 393-401 (2020).

4 Handakumbura, P. P., Rivas Ubach, A., & Battu, A. K. Visualizing the Hidden Half: Plant-Microbe Interactions in the Rhizosphere. *Msystems*, *6*(5), e00765-21 (2021).



**Funding statement:** *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design Science Focus Area "Persistence Control of Engineered Functions in Complex Soil Microbiomes". A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by BER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.*

**Title:** Discovery of Bacterial Species and Molecular Mechanisms Driving Growth on the Plant Root Metabolite Sorgoleone

**Authors:** Ryan McClure[1*](Ryan.McClure@pnnl.gov), Ritu Shrestha[1], Aaron Ogden[1], Henri Baldino,[1] Andrew Wilson,[1] Andrew Frank[1], Jared Kroll[1], Vivian Lin[1], Yuliya Farris[1], Yasuhiro Oda[2], Bill Nelson[1], Joshua Elmore[1], Caroline Harwood[2], **Rob Egbert**[1]

**Institutions:** [1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352. [2]Department of Microbiology, University of Washington, Seattle, WA 98195

**Website:** https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

**Project Goals:** The Pacific Northwest National Laboratory Persistence Control Scientific Focus Area aims to gain a fundamental understanding of factors governing the persistence of engineered microbial functions in rhizosphere environments. From this understanding, we will establish design principles to control the environmental niche of native rhizosphere microbes. In our first funding period, we are examining the efficacy of genome reduction and metabolic addiction to plant root exudates in environmental isolates as persistence control strategies using the bioenergy crop sorghum and defined microbial communities as a model ecosystem. Effective persistence control will lead to secure plant–microbe biosystems that promote stress-tolerant and highly productive biomass crops.

**Abstract**

The safe and responsible deployment of engineered microbes to promote bioenergy crop health demands biocontainment strategies that precisely control the environmental niche of the microbe. To investigate the potential to control the persistence of an engineered microbe through metabolic addiction to a plant root exudate compound, we have focused on identifying bacteria and genes responsible for use of the plant metabolite sorgoleone for growth. Sorgoleone is a hydrophobic allelochemical unique to the bioenergy crop sorghum [1]. To date, no studies have identified a microbe or catabolic pathway capable of degrading sorgoleone, though sorgoleone is clearly mineralized by the soil microbiome [2]. Here we report three bacteria from different genera that use sorgoleone as a sole carbon source (a *Burkholderia*, *Pseudomonas*, and *Acinetobacter* species). We discovered these microbes by purifying sorgoleone from seedling roots and enriching and isolating bacteria that use sorgoleone as a sole carbon source from sorghum field soil. This was followed by genome sequencing, and growth characterization in synthetic sorgoleone growth media. Each microbe has unique growth kinetics on sorgoleone and some also use other plant-derived compounds for growth. To identify genes associated with sorgoleone utilization, we performed differential transcriptomics experiments on each bacterial strain grown on sorgoleone and acetate. The three strains that we investigated have at least seven proteins in common that we hypothesize to be required for sorgoleone degradation. These are likely involved in the degradation of the quinone ring and alkene tail of sorgoleone. Mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy analysis of spent media from each bacterial strain indicated that no sorgoleone remained in the media after growth, though further investigation is required to identify possible strain-specific byproducts. The discovery of these sorgoleone degrading bacteria

opens up new avenues of research into how plant exudates shape their bacterial microbiomes. Further, we anticipate these findings will lead to the first demonstration and characterization of metabolic addiction of a microbe to a plant root exudate for biocontainment in a field setting.

## References

[1] Dayan, F. E.; Rimando, A. M.; Pan, Z.; Baerson, S. R.; Gimsing, A. L.; Duke, S. O. Sorgoleone. Phytochemistry 2010, 71, 1032-1039, https://dx.doi.org/10.1016/j.phytochem.2010.03.011.

[2] Gimsing, A. L.; Bælum, J.; Dayan, F. E.; Locke, M. A.; Sejerø, L. H.; Jacobsen, C. S. Mineralization of the allelochemical sorgoleone in soil. Chemosphere 2009, 76, 1041-1047, https://dx.doi.org/10.1016/j.chemosphere.2009.04.048.

**Funding statement:** *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design Science Focus Area "Persistence Control of Engineered Functions in Complex Soil Microbiomes". A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by BER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.*

# Standardized Workflows for Microbiome Omics Data Analysis

**Bin Hu[1*], Shane Canon[2], Emiley A. Eloe-Fadrosh[2], Anubhav[3], Michal Babinski[1], Yuri Corilo[3], Karen Davenport[1], William D. Duncan[2], Kjiersten Fagnan[2], Mark Flynn[1], Brian Foster[2], David Hays[2], Marcel Huntemann[2], Elais K Player Jackson[1], Julia Kelliher[1], Po-E Li[1], Chien-Chi Lo[1], Douglas Mans[3], Lee Ann McCue[3], Nigel Mouncey[2], Christopher J Mungall[2], Paul D. Piehowski[3], Samuel O Purvine[3], Montana Smith[3], Neha Jacob Varghese[2], Donald Winston[4], Yan Xu[1], Patrick S G Chain[1*]**

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA,

[2]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA,

[3]Environmental Molecular Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA,

[4]Polyneme LLC, New York, NY 10038, USA.

*** Correspondence:**
Corresponding Author
B.H (bhu@lanl.gov) and P.S.C (pchain@lanl.gov)

**Abstract**

To minimize the effort required to identify reusable microbiome datasets, the National Microbiome Data Collaborative (NMDC) was established in 2019 to support microbiome data exploration and discovery through a collaborative, integrative data science ecosystem. As part of the NMDC program, we have developed a set of standardized workflows to analyze metagenomics, metatranscriptomics, metabolomics, natural organic matter, and metaproteomics data. These open-source bioinformatics workflows for processing raw multi-omics data have been developed based on production-quality workflows at the two Department of Energy User Facilities, the Joint Genome Institute (JGI) at Lawrence Berkeley National Laboratory (LBNL) and the Environmental Molecular Systems Laboratory (EMSL) at Pacific Northwest National Laboratory (PNNL). The workflows have predefined parameter settings to ensure consistency across all available microbiome data. Users can download these workflows and adjust the

parameters based on their needs. Using these workflows we have processed 637 samples with more data being added following our quarterly release plan.

All the workflows are available both as source code in Workflow Definition Language (WDL) format from https://github.com/microbiomedata/ and software containers (https://hub.docker.com/u/microbiomedata). An intuitive web application is also available to run the metagenomics, metatranscriptomics, and natural organic matter workflows from https://nmdc-edge.org/home.

**Title:** Evidence for Active, Dynamic Viral Communities in Wet Soils across Habitats

**Authors:** Anneliek M. ter Horst,[1+] Christian Santos-Medellín,[1+] Sara E. Geonczy,[1] Jane D. Fudyma,[1] and **Joanne B. Emerson**[1]* (jbemerson@ucdavis.edu)

**Institutions:** [1]University of California, Davis, [+]These authors contributed equally

**Project Goals:** The overarching goal of this project is to assess and compare the contributions of active, infectious viruses and inert viral particles to biogeochemistry across diverse terrestrial ecosystems. Using a multi-omics approach, we seek to establish spatiotemporal patterns in soil viral community composition and activity linked to host carbon and nitrogen metabolism in grasslands, shrublands, woodlands, and wetlands. Leveraging a prescribed forest fire and a peatland temperature and atmospheric $CO_2$ manipulation experiment, we are also exploring feedbacks between soil viruses and carbon dynamics in response to environmental change. Through laboratory experiments, we are investigating the chemical composition, fate, and transport of viral particles in soil. By integrating field and laboratory experiments across a variety of soil edaphic properties and spatiotemporal scales, this project is expanding our understanding of the soil virosphere and its influence on carbon and nutrient cycling.

**Abstract Text:** Viruses have been recognized as highly abundant but poorly characterized members of the soil microbiome. By infecting soil microbes, viruses likely have substantial impacts on terrestrial biogeochemical processes under their hosts' control. Viral particles (virions) may also play more direct roles in soil biogeochemical cycling as packets of carbon, nitrogen, and phosphorus, but the time scales and environmental conditions that determine virion infectivity, transport, and/or sorption to soil particles are unknown. This project uses a combination of field, laboratory, and computational approaches to distinguish between infective and inert virions and to assess their respective contributions to soil biogeochemical cycling.

Using a 'viral size-fraction' metagenomics (viromics) approach, we are exploring the conditions and temporal scales over which virions are produced, remain infective, and decay in soil. We have shown that viromes can recover 1-2 orders of magnitude more viral sequence than total metagenomes, but the temporal scales over which a soil virome might be integrated were unknown at the start of this project. In Mediterranean habitats that experience a dry season, we find that the majority of the viral community seems to decay during seasonal dry-down, such that DNA contained in viral particles is largely undetectable throughout the dry season. In moist soils, results suggest that viral communities are active and dynamic and thus likely represent recently produced virions. However, dry periods early in the rainy season can render virion DNA undetectable within three weeks. Taken together, results suggest that soil viromes tend to capture very recently active viral communities, likely integrated over days to weeks.

Our interpretation that most soil viromes capture a short window of recent activity is consistent with our repeated findings of highly divergent soil viral communities over spatial distances as short as 1 m in nearly all habitats explored thus far. In other words, if viral communities turn over rapidly in time, communities far enough apart to experience dispersal limitation would likely be decoupled in both space and time, thus appearing very dissimilar at a single time point. Centered on a highly spatiotemporally resolved viromic study of the Jepson Prairie grassland (8 locations, 28 time points since November 2020), our ongoing work seeks to unravel the relative contributions of space, time, and dispersal on patterns of soil viral community composition.

We are also performing a suite of laboratory analyses on purified viral fractions to assess the chemical composition, sorption, and transport of soil virions. In soils from three habitats, viral community composition did not differ significantly across a range of buffer pHs for removing viral particles from the soil matrix, suggesting that our viromics protocol is relatively robust to different virion chemistries. We are using a ZetaSizer to determine whether viral particles tend to have isoelectric points above, below, or near the pH of their native soils to assess their potential sorption to minerals and/or transport within soil hydrological conduits. Exploratory imaging analyses in collaboration with EMSL and SSRL, planned for early 2022, are expected to reveal the phosphorus content and chemistry in virions from low-phosphorus wetlands.

To further investigate physicochemical constraints on virion integrity and viral community composition, we are analyzing data from three burned habitats. In shrublands and woodlands that burned during the dry season in the LNU Complex Fires in August 2020, viromic DNA yields remained below detection limits throughout the dry season, increased substantially after rain, and remained high until the following dry season. As these results mirrored patterns observed in unburned Mediterranean soils, we sought to decouple the effects of fire from desiccation alone by leveraging a prescribed burn in a mixed conifer forest that occurred during the rainy season (common for management) in Spring 2021. One week post-fire, viromic DNA yields were lower in burned compared to unburned soils, but this pattern was short-lived, suggesting that the impact of fire on virion abundance may be ephemeral. These preliminary results suggest that the degree of virion inactivation and the timing of viral community recovery post-fire seem to depend on soil moisture and depth. Viromes from these datasets are currently being sequenced, and laboratory experiments are being developed to tease apart the relative effects of temperature and soil moisture on viral community composition and virion integrity.

The DOE Spruce and Peatland Responses Under Changing Environments (SPRUCE) experiment provides a platform for testing the vulnerability of boreal peat viruses and processes under their control, including host biogeochemical cycling, to elevated temperature and atmospheric $CO_2$ concentrations. Over the first two years of whole ecosystem warming and deep peat heating (2015-2016), peat viral community composition was significantly correlated with peat depth, water content, and porewater $CH_4$ and $CO_2$ concentrations, but not with temperature. In order to more thoroughly assess feedbacks between peat viruses and carbon cycling, we are tracking peat viral community composition and virus-host dynamics over longer time scales in the SPRUCE experiment (through 2022), with a focus on viral predation of methanogens and methanotrophs responsible for $CH_4$ cycling and release to the atmosphere.

Results from this project are facilitating a better understanding of viral contributions to terrestrial biogeochemical cycling, both as dynamic components of soil organic matter and through their infection of hosts responsible for carbon and nutrient cycling.

**Publications**

1. ter Horst, A.M., Santos-Medellín, C., Sorensen, J.W., Zinke, L.A., Wilson, R.M., Johnston, E.R., Trubl, G.G., Pett-Ridge, J., Blazewicz, S.J., Hanson, P.J., Chanton, J.P., Schadt, C.W., Kostka, J.E., Emerson, J.B. (2021) Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations. *Microbiome*. 9:233.

**Funding statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021198.

# PhytoOracle: Modular, Scalable Phenomic Data Processing Pipelines

Emmanuel Gonzalez*[1] (emmanuelgonzalez@email.arizona.edu), Ariyan Zarei,[2] Nathanial Hendler,[1] Michele Cosi,[1] Jeffrey Demieville,[1] Travis Simmons,[1,3] Holly Ellingson,[4] Nirav Merchant,[4] Eric Lyons,[1,4] Duke Pauli,[1,4] and **Andrea Eveland**[5]

[1]School of Plant Sciences, University of Arizona, Tucson, AZ; [2]Department of Computer Sciences, University of Arizona, Tucson, AZ; [3]College of Coastal Georgia, Brunswick, GA; [4]Data Science Institute, University of Arizona, Tucson, USA; and [5]Donald Danforth Plant Science Center, St. Louis, MO

PhytoOracle GitHub Repository, CyVerse Data Commons

## Project goals

As phenomics continues to generate larger and higher dimensional data sets, there is an urgent need to develop and implement robust data processing pipelines to extract biological insight and knowledge from these data. Current phenomics processing pipelines lack modularity and the ability to exploit distributed computational infrastructure. To address these challenges, we developed PhytoOracle (PO), a suite of modular, scalable pipelines that aim to improve data processing efficiency for plant science research. PO integrates CCTools' Makeflow and Workqueue frameworks for distributed task management on local, Cloud, or high-performance computing (HPC) systems. Each pipeline component is available as a standalone container, providing transferability, extensibility, and reproducibility. PO efficiently extracts phenotype trait values, capturing phenotypic variation for elucidation of the genetic components of complex traits.

## Abstract

Understanding how plants dynamically respond to environmental conditions has been historically difficult due to the low throughput and long-term cost of longitudinal data collection in the field setting[1]. Recent technological advances have resulted in small, cheap, and high resolution sensors that can be used to rapidly collect phenotypic trait data at regular time intervals in field or greenhouse settings[2,3]. Such platforms will continue evolving and become more widespread, likely exacerbating the phenotyping bottleneck by creating a strain on current data processing frameworks[4].

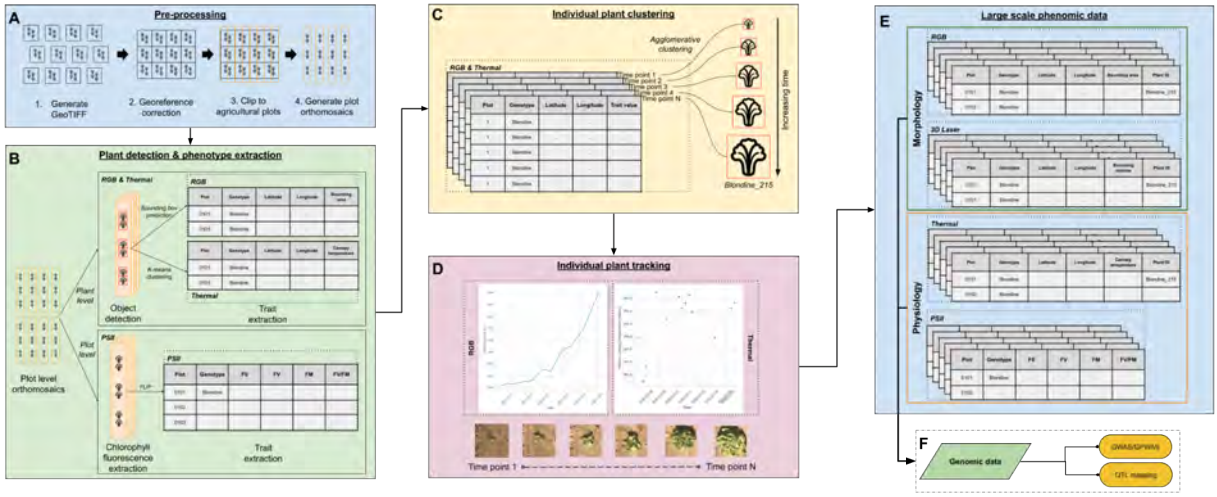



***Figure 1*** PhytoOracle pipeline workflow. (A) Image georeferencing metadata is corrected and images are clipped to agricultural plot boundaries. (B) Plants are detected using a Faster R-CNN model and (C) clustered and given a unique ID. (D) Single plants can be tracked using unique IDs. (E) Multimodal data is merged using unique IDs and plot information. (F) Large scale phenomic data and genomic data can enable GWAS and/or QTL mapping.

To address this phenotyping bottleneck, we developed PO, a suite of modular, scalable data processing pipelines for RGB and thermal cameras, Photosystem II (PSII) imager, and structured-light laser scanners (3D laser) raw data[5]. PO distributes tasks using CCTools Makeflow and Workqueue[6] to extract plant-level bounding area, canopy temperature, height, bounding volume and plot-level maximum quantum efficiency of PSII ($F_V/F_M$) using Faster R-CNN models (**Fig. 1**). PO scales with processing times of 235 minutes for 9,270 RGB images, 235 minutes for 9,270 FLIR images, and 13 minutes for 39,678 PSII

images (**Fig. 2**). The resulting processed data are associated using agglomerative clustering to enable time-series analysis of individual plant phenotypes (**Fig. 1**). Images and point clouds of each plant throughout the growing season are collected. These datasets are useful for a wide variety of future ML applications, such as the classification of plant varieties for precision agriculture, segmentation of plant disease, and modeling of plant growth and its influence on light use efficiency.

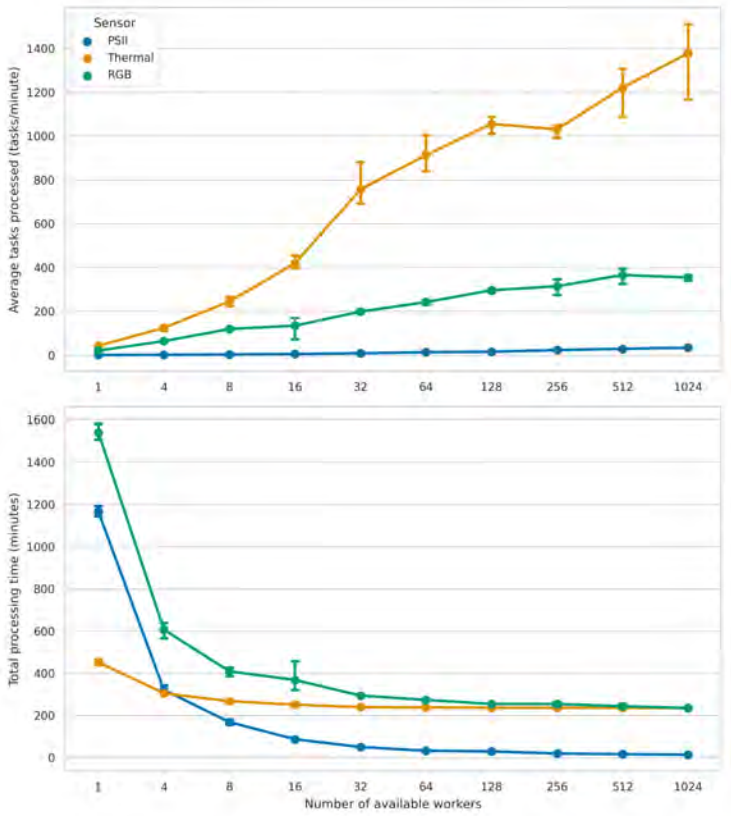



***Figure 2*** PhytoOracle benchmarking of PSII, thermal, and RGB pipelines. Increase in average tasks processed as the number of workers increases (Top). Decrease in total processing time (minutes) as the number of workers increases (Bottom). Benchmark datasets for each sensor were processed over the following range of available workers: 1, 4, 8, 16, 32, 64, 128, 256, 512, and 1024. Each configuration was run three times totaling 30 benchmark observations.

**Funding statement**

This material is based upon work supported by the U.S. Department of Energy Biological and Environmental Research program under Award Number DE-SC0020401, the U.S. Department of Energy Advanced Research Projects Agency - Energy OPEN program under Award Number DE-AR0001101, and the National Science Foundation CyVerse project under Award Number DBI-1743442.

**References**

1. Reynolds, D. *et al.* What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Sci.* **282**, 14–22 (2019).
2. Li, B. *et al.* Phenomics-based GWAS analysis reveals the genetic architecture for drought resistance in cotton. *Plant Biotechnol. J.* **18**, 2533–2544 (2020).
3. Sooriyapathirana, S. D. S. S. *et al.* Photosynthetic Phenomics of Field- and Greenhouse-Grown Amaranths vs. Sensory and Species Delimits. *Plant Phenomics* **2021**, 1–13 (2021).
4. Furbank, R. T. & Tester, M. Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **16**, 635–644 (2011).
5. Gonzalez, E. *et al. PhytoOracle: Scalable, Modular Phenomic Data Processing Pipelines*. (2021).
6. Albrecht, M., Donnelly, P., Bui, P. & Thain, D. *Makeflow: a portable abstraction for data intensive computing on clusters, clouds, and grids*. 13 (2012).

**Using Molecular Genetics and Precision Phenotyping to Map Gene Function Contributing to Drought Resilience in Sorghum**

Yuguo Xiao,[1]* (yxiao@danforthcenter.org) Indrajit Kumar[1], Rajdeep S. Khangura[2], Maxwell Braud,[1] Emmanuel Gonzalez,[3] Philip Ozersky,[1] Duke Pauli,[3] Todd C. Mockler,[1] Brian Dilkes,[2] and **Andrea L. Eveland**[1]

[1]Donald Danforth Plant Science Center, St. Louis, MO; [2]Purdue University, West Lafayette, IN; [3]University of Arizona, Tucson, AZ

**Project Goals:**

- **Overall project objective: To define and functionally characterize genes and pathways related to drought stress tolerance in sorghum and the molecular mechanisms by which these factors drive phenotypic diversity.**
- **Establish a foundation for deep explorations of gene regulatory networks in sorghum through integrative genomics analyses.**
- **Enhance understanding of how genotype drives phenotype and environmental adaptation using high-resolution, field-based phenotyping of sorghum mutant collections.**
- **Map and characterize genes contributing to drought responsive phenotypes in sorghum.**
- **Experimentally validate predictions of gene function using molecular and genetic assays and targeted gene editing.**

Development of the next generation of bioenergy feedstocks will require strategies that utilize resource-limited agricultural lands, including the introduction of novel traits into crops to increase abiotic stress tolerance. This project investigates the innate drought resilience of sorghum (*Sorghum bicolor*), a bioenergy feedstock and cereal crop. Drought is a complex trait and identifying the genes underlying sorghum's innate drought tolerance and how they are regulated in the broader context of the whole plant and its environment requires advanced approaches in genetics, genomics, and phenotyping.

This project leverages a sequence-indexed population of EMS mutagenized sorghum and a field-based phenotyping infrastructure at Maricopa, AZ, which provides an exceptional capability for managed stress trials in a hot and arid environment through controlled irrigation. An automated field scanner system collects high-resolution phenotyping data using a variety of sensors throughout the growing season, from seedling establishment to harvest. 430 EMS families in the tx623 background were phenotyped under the field scanner to compare drought-stressed and well-watered plants. Each mutant's genome has been sequenced so that sequence variants can be linked with phenotypes. Being able to assess the genotype-to-phenotype link in response to drought over the life cycle of the plant will facilitate discovery of genes and their functions. To accelerate mapping of causal loci that underlie mutants of interest, we use bulked segregant analysis-seq. So far, we've identified candidate genes underlying defects in leaf senescence, plant architecture and male fertility. Regulatory maps generated from diverse sorghum lines in response to stress are being used to nominate gene candidates and place them in the larger context of a drought response network. This work will identify control points for enhancing the

productivity of bioenergy crops in marginal environments through precision breeding or engineering, and thus accelerate the development of improved varieties that are high-yielding with limited water resources.

*This work is funded by DOE BER award #DE-SC0020401*

**Spatial Turnover of Soil Viral Communities and Genotypes Overlain by Cohesive Responses to Moisture in Grasslands**

Christian Santos-Medellin[1]* (cmsantosm@ucdavis.edu), Katerina Estera-Molina,[2] Mengting Yuan,[2] **Jennifer Pett-Ridge**,[3] **Mary K. Firestone**,[2] and **Joanne B. Emerson**[1]

[1]University of California, Davis; [2]University of California, Berkeley; [3]Lawrence Livermore National Laboratory, Livermore, CA

**Website URL:** https://nature.berkeley.edu/crosskingdominteractions

**Project Goals:** The overarching goal of our project is to understand how cross-kingdom and within-kingdom interactions (involving viruses, bacteria, archaea, fungi, protists, microfauna, and plant roots) provide a functional foundation for nutrient cycling in grassland soils. Within this framework, we seek to identify the biotic and abiotic factors that govern the structure, variation, and assembly of viral communities inhabiting these environments. By revealing the conditions under which viral community compositional patterns are and/or are not tightly coupled to their microbial host communities and biogeochemistry, we can begin to unravel the extent to which virus-host interactions impact soil carbon and nutrient cycling.

**Abstract Text:** Soil and rhizosphere microorganisms play key roles in biogeochemical cycling and plant productivity, and by infecting soil microbiota, viruses likely have substantial direct and indirect impacts on these processes. In the oceans, viruses lyse (burst and kill) an estimated 20-40% of microbial cells daily, impacting global ocean food webs, carbon and nutrient cycling, and climate. At ~$10^7$ to $10^{10}$ viruses per gram, soil viruses may play similarly important roles in terrestrial ecosystems and have been recognized as abundant but virtually unknown members of the soil microbiome.

As part of our large-scale field manipulation study on the impact of reduced precipitation on soil biotic interactions in a Mediterranean grassland, we collected surface (0-15cm) soils from 22 densely-rooted locations distributed across 15 experimental plots that have received treatments of 100% or 50% 50-year averaged precipitation since 2017. Soils were collected twice during the 2020 growing season of *Avena barbata*, the naturalized annual grass that dominates the ecosystem, for a total of 44 samples. To profile dsDNA viral diversity in these samples, we generated 44 viral size-fraction metagenomes (viromes) by separating smaller virions from larger microbes with 0.2 μm filtration prior to DNA extraction and sequencing.

Viral community beta-diversity was most significantly explained by spatial distance, a pattern driven by the steady turnover of viral populations (vOTUs) along a 16 m horizontal transect. This spatial structuring also impacted the genetic microdiversity of the more ubiquitous vOTUs, such that genome sequences from predominant allelic variants tended to diverge with increasing spatial distance. Both viral community composition and the genetic makeup of viral populations exhibited significant distance-decay relationships across the field, highlighting potential

constraints on the spatiotemporal scales of soil viral dispersal and virus-host interactions in these soils.

In addition to spatial structuring, reduced precipitation was significantly correlated with viral community composition, with evidence for more substantial viral responses to current or very recent moisture regimes than to legacy precipitation treatment. Interestingly, many of the vOTUs that were enriched in lower moisture samples were also grouped according to shared predicted protein content in a network analysis, suggesting that viral populations with higher-level genome conservation displayed a cohesive response to reduced moisture content. Furthermore, many of these vOTUs shared genomic features with known actinophages, suggesting that their responses to desiccation could result from interactions with host Actinobacteria, which have been established as drought responsive in prior studies.

In a companion study, we are exploring virus-host dynamics in response to laboratory rewetting of dry soils from four distinct grasslands, including those adjacent to the Hopland field experiment described above. Tens of thousands of vOTUs were recovered from these 10-day laboratory experiments, in which viral communities separated most significantly by soil source and then by dry versus wet soils. For all wet soils analyzed so far, highly diverse viral communities exhibited rapid temporal succession that was reproducible across replicate microcosms, with viromes from 24 hours and 10 days post-rewetting sharing on average fewer than 25% of their vOTUs.

Together, results suggest active and highly dynamic grassland viral communities that seem to respond rapidly to both increases and decreases in soil moisture. The high degree of spatial turnover within one field during one growing season suggests dispersal limitation for most viral populations on scales of meters and months. During the final year of this project, we will focus on further incorporation of host community composition and activity in our analyses to better understand how virus-host interactions contribute to carbon and nitrogen cycling in grassland soils.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0020163. Lawrence Livermore National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. xx. This program is supported by the U.S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP xx.

**Additions to the BAHD Acyltransferase Toolbox**

Rebecca A. Smith,[2] Lisanne de Vries,[3] Heather A. MacKay,[3] Eliana Gonzales-Vigil,[3] Justin F. Acheson,[1] Emily T. Beebe,[1,2] Craig A. Bingman,[1,2] Debayan Chaudhury,[1] Jeff Y. Chen,[3] Alexis Eugene,[3] Tegan M. Haslam,[3] Steven D. Karlen,[2] Ljerka Kunst,[3] Leta Landucci,[1,2] Sitong Li,[3] Yaseen Mottiar,[2,3] Emilia Muirragui,[1] Alexander J. Steiner,[3] Lewis Tanner,[1,2] Faride Unda,[3] Kirk Vander Meulen,[1,2] Michelle E. vonLoessl,[3] Shawn D. Mansfield,[2,3] John Ralph,[1,2] **Brian G. Fox**[1,2*] (bgfox@wisc.edu)

[1]Department of Biochemistry and [2]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA; [3]Department of Wood Science, University of British Columbia, Vancouver, BC, Canada

**Project Goal:** Use ligases and BAHD acyltransferases to modify lignin composition in bioenergy crop species.

**Abstract:** Plants create a myriad of primary and secondary metabolites through the action of acyl-CoA ligases and acyl-CoA transferases. Acyl-CoA ligases use ATP and an organic acid to form an acyl-CoA donor, whereas the BAHD transferases combine the acyl-CoA donor with an acceptor, typically an alcohol or phenol. The resulting ester-containing metabolites can be used as 'monomers' for lignification, and function as donors to further modify lignin with clip-offs such as *p*-hydroxybenzoate, in biosynthesis of cuticular wax esters, or in many other vital biological reactions.

To expand knowledge of these reactions, we combine phylogenetics, gene synthesis, cell-free protein synthesis, biochemical assays, structure determination, state-of-the art analytical methods, and protein expression *in planta,* microbes or yeast. Wheatgerm cell-free protein synthesis was used to produce previously known *p*-coumaroyl-CoA and feruloyl-CoA monolignol transferases (PMT and FMT) from rice as controls (1). Comparable synthesis of putative PMT and FMT genes from sorghum and switchgrass revealed new enzymes that produced monolignol *p*-coumarate and ferulate conjugates with a breadth of substrate selectivities and relative rates (1). The new FMTs and PMTs were transformed into *Arabidopsis thaliana* (which does not naturally make monolignol conjugates) and the presence of ferulates and *p*-coumarates on the lignin demonstrated enzymatic function within plants. Additional screening of the poplar BAHD acyltransferase superfamily (116 genes) revealed five enzymes capable of producing varying levels of monolignol *p*-hydroxybenzoates (2). Analysis of transcript abundance and *p*-hydroxybenzoate levels in unrelated genotypes of poplar indicated that *pHBMT1* would be the best candidate for studies in transgenics. Three analytical methods showed that an increased level of *p*-hydroxybenzoate, a precursor to parabens used in pharmaceuticals and cosmetics, could be released by saponification from the transgenic poplar. In addition, five uncharacterized paralogous BAHD transferases from clade II genes (*PtCER2-like1* through *7*) were expressed in yeast and PtCER2-like5 produced the highest levels of C28 fatty acids when expressed in the presence of the condensing enzyme AtCER6. Its expression

was localized to the epidermis in GUS-reporter poplar lines, consistent with a role in cuticular wax biosynthesis. We have also found new examples of aromatic acid CoA ligases in bacteria, and these have unique domain architectures, reaction mechanisms, substrate specificities, and rates of reaction relative to ligases from plants. These newly identified genes and enzymes are being further evaluated for potential use in engineering lignocellulosic biomass with increased value for emerging biorefinery strategies.

**References:**


1. Rebecca A. Smith, Emily T. Beebe, Craig A. Bingman, Kirk Vander Meulen, Alexis Eugene, Steven D. Karlen, John Ralph, and Brian G. Fox. (2022). Identification and characterization of a new set of monocot BAHD monolignol transferases. Plant Physiol, in press.
2. Lisanne de Vries, Heather A. MacKay, Rebecca A. Smith, Yaseen Mottiar, Steven D Karlen, Faride Unda, Emilia Muirragui, Craig A. Bingman, Kirk Vander Meulen, Emily T. Beebe, Brian G. Fox, John Ralph, and Shawn D. Mansfield. (2021). *p*HBMT1, a BAHD-family monolignol acyltransferase, mediates lignin acylation in poplar. Plant Physiol. https://doi.org/10.1093/plphys/kiab546.
3. Eliana Gonzales-Vigil, Michelle E. vonLoessl, Jeff Y. Chen, Sitong Li, Tegan M. Haslam, Ljerka Kunst, Shawn D. Mansfield. (2021). Understanding the role of Populus ECERIFERUM2-likes in the biosynthesis of very-long-chain fatty acids for cuticular waxes. Plant Cell Physiol. 62(5):827-838. doi: 10.1093/pcp/pcab040. PMID: 33749753.



*This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0020349 (BG Fox, PI), DE-SC0018409 and DE-FC02-07ER64494 (Great Lakes Bioenergy Research Center, TJ Donohue, PI). Gene synthesis was carried out by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, under Contract No. DE-AC02-05CH11231.*

**Title: Optimizing enzymes for plastic upcycling using machine learning design and high throughput experiments**

**Authors:** Benjamin Fram[1,2] (fram@g.harvard.edu), Chris Bahl[3], Gregg Beckham[4], Richard Brizendine[4], Karla Ilic Durdic[5], Perry Ellis[5], Japheth Gado[4], Erika Erickson[4], Samuel Lim[2], Maria Llases[3], Debora Marks[2], Nathan Rollins[2], Chris Sander[1,2], Nicole Thadani[2], Ian Truebridge[3], David Weitz[5], Allison Werner[4], Yi Xiao[5], Xinge Zhang[5], **Nicholas Gauthier**[1,2]

**Institutions:** [1]Dana-Farber Cancer Institute, Boston; [2]Harvard Medical School, Boston; [3]Institute for Protein Innovation; [4]National Renewable Energy Laboratory; [5]Harvard University;

*Authors listed alphabetically except presenter (BF) and PI (NG).

**Project Goals:** We aim to create new and optimized PET-depolymerizing enzymes (PETases) useful for industrial application. [**Aim 1]** Design novel PETases that are significantly different (25-65+ mutations) from known PET-depolymerizing enzymes and contain unique properties useful for performant enzymatic PET recycling and upcycling. Introducing many simultaneous mutations, while maintaining function, will enable us to more efficiently search for altered properties that depend on primary amino acid sequence. **[Aim 2]** Optimize previously described PETases by testing millions of mutagenized variants using directed evolution. Starting with existing functional PETases and exploring small changes in many distinct sequences using a novel ultra-HTP functional assay, we will optimize enzymes with improved properties by varying experimental conditions. **[Aim 3]** Characterize performance metrics of new and optimized PETases in detail including solvent tolerance, stability, catalytic rate, and substrate promiscuity.

**Abstract Text**:

Plastic use is ubiquitous in the modern world, and polyethylene terephthalate (PET) is one of the most abundantly produced plastics (and the most highly produced polyester), with ~65 million metric tons manufactured annually. To the consumer, PET is likely most recognizable as the plastic used to make beverage bottles. Like many plastics, traditional mechanical or chemical means of PET deconstruction and upcycling are costly and inefficient.

Recently, biological enzymes capable of breaking down PET into its basic building blocks (terephthalic acid and ethylene glycol) have garnered significant attention as an attractive means of dealing with the plastic problem. These enzymes are currently undergoing pilot studies for implementation in enzyme-based recycling. However, there are significant limitations to current enzymes, including the need to perform costly pre-processing of the plastic waste before the enzymes are able to work. Further optimization of these enzymes is necessary to make the

process profitable and thereby incentivize commercialization of this biology-based green recycling technology.

We aim to apply recent advances in artificial intelligence and machine learning to design new versions of enzymes capable of breaking down PET. Based on preliminary experiments using this evolution-informed computational design strategy, we believe we can create a highly diverse set of enzymes that have exceptional properties useful for industrial recycling. Testing these enzymes is typically labor intensive, but using a new robotically-enabled platform we will be able to experimentally characterize key enzymatic properties of thousands of these designed enzymes.

In addition to applying machine learning approaches to design new enzymes, we have developed a novel method that, by experimentally testing millions of small changes to enzyme structure, enables us to optimize existing enzymes that are known to break down plastic. The key to this approach is that we encapsulate individual variations of each enzyme in single droplets together with plastic nanoparticles creating a “mini reaction”, and then we select those droplets which successfully break down PET to isolate the ‘winning’ enzyme variants.

Ultimately, we believe that the result of these studies will be the discovery of highly-optimized enzymes capable of breaking down PET plastics in an industrial recycling setting, enabling a powerful and green solution to the plastic problem.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022024

**Title:** Does Predation by Protists Mediate the Effects of Temperature and Nutrient Additions on Microbial Food Webs?

**Authors:** Katrina DeWitt[1]* (Katrina.dewitt@duke.edu), Alyssa A. Carrell[2], Jennifer D. Rocca[1], Samantha Votzke[1], Andrea Yammine[1], Dale A. Pelletier[2], **Jean P. Gibert[1]**

**Institutions:** [1]Department of Biology, Duke University, Durham, NC; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831, USA

**Project Goals:** To assess whether predation by protists can mediate the responses of microbial communities to temperature and nutrient increases.

**Abstract Text:** Rising global temperatures and increases in nutrient deposition through severe air pollution may have catastrophic effects on natural ecosystems. Warmer temperatures increase metabolic rates, leading to increased energetic demands among endotherms. The burning of fossil fuels, emission of ammonia by industrial regions, and the indiscriminate use of fertilizer have contributed to ever-increasing nutrient deposition rates. Worse, anthropogenic nutrient additions interact with warming temperatures in often unpredictable ways. While we have previously shown that the temperature response of microbial communities is mediated by ecological interactions, how these communities may respond to both temperature increase and nutrient additions in the presence of predators is unknown. To test how predation mediates the combined effects of nutrient additions and temperature, we set up microbial microcosms where we manipulated temperature, nutrients, and the presence and absence of protists. Specifically, microbial microcosms were assembled using moss-associated bacteria from the DOE-supported SPRUCE site in the Marcell experimental forest in Minnesota. We manipulated microcosm temperatures (22℃ and 25℃), nutrient conditions (standard and half concentration Carolina protist media), and the presence of protist predators on 200mL glass jars over three weeks. To assess the response of the bacterial communities, we quantified total microbial biomass using OD600, total respiration rate using real-time respirometry, and changes in microbial community structure using amplicon sequencing.

Our results show that temperature, nutrient concentration, and the presence of protists interactively influence microbial growth and community structure. Increased temperature and low nutrient conditions negatively affect total microbial biomass, while the presence of protists has a positive effect on biomass independent of temperature and nutrient conditions. The microbiome functional changes are accompanied by distinct shifts in community structure that show a stronger response to protist predation than temperature. These results suggest that while temperature and nutrients may indeed interactively influence the structure and function of microbial communities, their response is ultimately determined by top-down controls exerted by bacterivore organisms that may themselves be responding to increased temperature). Moreover, we show that while protist presence may certainly provide top-down control, it can also fertilize microbial growth by recycling nutrients stored in bacterial biomass. Overall, these results can help improve predictions on how the combined effects of nutrient additions and warming will affect ecological communities and food web structure.

**Funding Statement:** This work was supported by a U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program Grant to JPG, under Award Number DE-SC0020362.

**Title:** Predicting complex microbial temperature responses across scales

**Authors:** Daniel J. Wieczynski[1]* (daniel.wieczynski@duke.edu), Pranav Singla[1], Adrian Doan[1], Alexandra Singleton[1], Zeyi Han[1], Samantha Votzke[1], Andrea Yammine[1], **Jean P. Gibert[1]**

**Institutions:** [1]Department of Biology, Duke University, Durham, NC

**Website URL:** https://www.pnas.org/content/118/42/e2104863118.short

**Project Goals:** This project scales up biological responses to warming from the individual level to the ecosystem level by using microbial traits to link complex population, community, and ecosystem processes.

**Abstract Text:** Microbial communities regulate ecosystem responses to climate change. But predicting these responses is challenging due to complex interactions among processes at multiple levels of organization. Organismal traits that determine individual performance and ecological interactions are essential for scaling up environmental responses from individuals to ecosystems. We combine protist microcosm experiments and mathematical models to show that key traits—cell size, shape, and contents—each explain different aspects of species' demographic responses to changes in temperature. These differences in species' temperature responses have complex, cascading effects across levels of organization—causing nonlinear shifts in total community respiration rates across temperatures via coordinated changes in community composition, equilibrium densities, and community-mean species mass in experimental protist communities that tightly match theoretical predictions. Our results suggest that traits explain variation in population growth and, together, these two factors scale up to influence community- and ecosystem-level processes across temperatures. Connecting the multi-level microbial processes that ultimately influence climate in this way will help refine predictions about complex ecosystem-climate feedbacks and the pace of climate change itself.

**Funding Statement:** This work was supported by a U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program Grant to JPG, under Award Number DE-SC0020362.

**Enhancement of the Random Barcode Transposon Sequencing Applications within KBase**

[2] Omree Gal-Oz (**ogaloz@lbl.gov**), [2] Adam Arkin, and [1,3] **Ryan Gill**

[1] Renewable and Sustainable Energy Institute, University of Colorado, Boulder CO;
[2] Lawrence Berkeley National Laboratory, Berkeley, CA; [3] Danish Technical Institute, Copenhagen, Denmark.

**http://kbase.us/**

**Project Goals:**

**One of the central challenges in the synthetic design of microbes for operation in variable environments is that productivity and fitness can be negatively affected as environments vary. A central hypothesis of this project is that it is possible to build synthetic circuits in cells that deploy certain genes at the right times to respond to environmental change and optimize fitness and productivity across variable environments. We have been adding computational tooling into the DOE Systems Biology Knowledgebase to support this design workflow[1]. Previously, we have prototyped the use of a circuit design system developed by the Voigt lab at MIT called Cello that, if you know which genes you wish to express when, will design a minimal genetic circuit to meet performance goals. To support such designs, it is necessary to be able to assess which genes of an organism are likely to tune both fitness and productivity in each environment. Recently, a number of different high-throughput genetic techniques have emerged to efficiently enable this assessment. The Arkin lab has developed RB-TNSEQ as an approach to this problem[2]. This version of tagged transposon mutagenesis has been used across diverse bacteria (and some fungi) and has been used to identify critical genes that could be engineered to improve production goals. The experimental protocols and reagents have been published and distributed widely to the community. However, the software to quality assess, analyze and produce the final prioritized list of genes important for a given phenotype is complex and difficult to use. Our goal this period has been to lower the barrier to its use by simplifying, enhancing, and documenting the workflow within KBase.**

The current implementation of RB-TNSEQ on KBase allows users to step through mapping their RB-TNSEQ library, assessing it for bias, and analyzing fitness/phenotype differences across conditions. This involves nine applications listed in tables below which are partially in beta and nearly ready for release. We have created a comprehensive user manual, a well-annotated demonstrative public Narrative, and an instructional video to help users learn the system. We expect full release of this system by the end of Q2, 2022.

The following page consists of three tables which group the applications by function. “Mapping Applications” refers to the applications that should be run using the sequencing reads that are acquired after running transposon insertion. “Fitness Assessment Applications” refers to applications that should be run using the sequencing reads that are acquired when counting the barcodes which have been previously decided upon. “Complimentary Applications” refers to applications that assist with the workflow. Within all categories, the apps that are not necessary for the entire workflow to be completed are labeled with a star.

## Mapping Applications

| | |
|---|---|
| **RBTnSeq MapTables*** | Map DNA reads to insertion locations within the genome |
| **RBTnSeq Maps to Pool** | Create insertion locations (as above) and create a pool of curated insertions. Returns a list of barcodes to be used downstream. |
| **RBTnSeq Mutant Pool Visualizer*** | Run the data visualization component of 'RBTnSeq Maps to Pool' - allows for users to visualize **completed** data after edits |

## Fitness Assessment Applications

| | |
|---|---|
| **RBTnSeq Reads to Pool Counts** | Count barcodes from reads representing growth in conditions |
| **RBTnSeq BarSeqR** | Compute Fitness Results from counted barcodes files |
| **RBTnSeq Fitness Visualizer*** | Run the data visualization component of 'RBTnSeq BarSeqR' |

## Complimentary Applications

| | |
|---|---|
| **RBTnSeq Find TnSeq Model*** | To be run before mapping applications. This finds the most likely TnSeq Transposon Model used in a previous sequencing run. |
| **RBTnSeq Genome to Gene Table** | Convert KBase Genome Objects to Gene Tables (which are similar to .gff files) |
| **RBTnSeq Download Tables*** | Download RBTnSeq data objects from KBase as tables (TSV files) |

* - *auxiliary application (i.e. not a necessary application to produce fitness results)*

**References**


1. Wetmore, K.M. *et al*. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6**, e00306–e00315 (2015).
2. Arkin, A., Cottingham, R., Henry, C. *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36, 566–569 (2018).



*This research on Design and Engineering of Synthetic Control Architectures is supported by the U.S. Department of Energy, Office of*
*Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231 and DE‑SC0018368*

**Genome-Scale CRISPRi in the Rapidly Growing Cyanobacterium *Synechococcus* sp. PCC 7002**

Andrew Hren[1*] (andrew.hren@colorado.edu), Emily Freed[1], Carrie Eckert,[1,2] and **Ryan Gill[1,3]**.

[1]Renewable and Sustainable Energy Institute, University of Colorado, Boulder; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Danish Technical Institute, Copenhagen, Denmark

http://gillgroup.org/research/

**Project Goals:** Our objective is to accelerate the engineering of designer organisms with traits amenable to sustainable biochemical production. By devising functional genetic screens using a genome-scale CRISPRi library in the photoautotrophic cyanobacterium *Synechococcus* sp. PCC 7002, we seek to: i) identify gene targets for maximizing short chain alcohol synthesis, ii) enable CRISPR-based trackable genome engineering (CREATE) experiments[1], and iii) uncover light-responsive signaling pathways in a host capable of direct $CO_2$ capture.

**Abstract text:** Sustainable bioenergy production in microbes requires deep understanding of a host's native biological functions and the development of tools and design principles necessary to engineer production at sufficient titers. While these objectives have been heavily investigated in model organisms (e.g., E. coli, S. cerevisiae), similar development in chassis organisms capable of efficient conversion of sustainable feedstocks has lagged behind. Synechococcus sp. PCC 7002 (PCC 7002) has emerged as a choice host for biochemical production thanks to its rapid growth rate and direct conversion of CO2 into products through photosynthesis. This work seeks to establish genome-scale CRISPR interference (CRISPRi) in PCC 7002, revealing uncharacterized gene functionalities, validating gRNA activities, and enabling downstream design and engineering of synthetic regulatory networks and hybrid control structures for enhanced production of industrially relevant short chain alcohols. The library will initially be screened across light wavelengths as a means of broadening understanding of light and color perception in PCC 7002, facilitating development of tunable production phenotypes using novel optogenetic control systems active across the light spectrum.

**References:**

1. Garst, A., Bassalo, M., Pines, G. *et al.* Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotechnol* **35,** 48–55 (2017). https://doi.org/10.1038/nbt.3718

**Funding Statement:** This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science under contract number DE-SC0018368.

**Title:** A *S. cerevisiae* "Marionette" strain to control metabolic pathways

**Authors:** Jong Hyun Park[1*] (jonghyun@mit.edu), Marcelo Bassalo[1], Chen Ye,[1] Joep Schmitz,[2] Johannes A Roubos[2], Ryan Gill[3,4], and **Christopher A Voigt[1]**

**Institutions:** [1] Massachusetts Institute of Technology, Cambridge, MA; [2] DSM Biotechnology Center, Delft, The Netherlands; [3] Renewable and Sustainable Energy Institute, University of Colorado, Boulder CO; [4]Danish Technical Institute, Copenhagen, Denmark.

**Project Goals: Our objective is to develop smart yeast strain that can efficiently regulate gene expressions by using multiple biosensors. To do so, we repurposed prokaryotic small molecule sensors and introduced them into *Saccharomyces cerevisiae*. As a result, we can construct a yeast "Marionette" strain containing 4 optimized sensors that enable independent control of metabolic pathway genes. To demonstrate the yeast "Marionette" strain, we optimized production of the monoterpene Linalool with precise controls of both expression levels and timing to get high titer.**

Microbial fermentation with the budding yeast *Saccharomyces cerevisiae* has been used for over 8000 years, and recent engineering advances have expanded its application to a wide panel of valuable chemicals. In order to engineer more robust microbial cell factories, smarter control strategies need to be implemented, enabling cells to redirect flux at the right levels and at the right time to optimally balance fitness and production. However, in yeast, a lack of multiple biosensors often leads to uncontrolled and 'constitutively ON' pathway designs.

Introduction of multiple sensors into the yeast strain require that sensors meet certain performance criteria. First, lower basal expression, as leakiness can restrict optimization of genes whose expression need to be low (e.g., toxic genes). Second, high fold change between the on and off states (dynamic range), providing a wide range of expression levels to explore in the controlled target gene. Third, a high level of orthogonality, as sensors should not crosstalk with other promoters and small molecules in the system. Finally, each sensor incurs a tax on cellular resources, which introduces burden to the strain and can influence the response of the other sensors. Thus, combining sensors should produce minimal impact to the engineered strain.

Here, we developed a yeast Marionette strain containing 4 optimized sensors that enable independent control of metabolic pathway genes. We demonstrated the Marionette strain by optimizing production of the monoterpene Linalool, exploring both expression level profiles as well as temporal dynamics in the biosynthetic pathway. "Marionette" comprises a flexible platform that should accelerate the design-build-test strain engineering cycles, uncovering design principles that can guide implementation of more complex synthetic regulation in metabolic engineering applications.

*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018368.*

**Title:** Design and Engineering of Synthetic Control Architectures

**Authors:** David Romero-Suarez,[2] (darosu@biosustain.dtu.dk), Linas Tamošaitis[2], Sergi Muyo[2] , Giovanni Schiesaro[2], Marina Mohr[2], and Ryan Gill[1,2]

**Institutions:** [1]Renewable and Sustainable Energy Institute, University of Colorado, Boulder; [2] Technical University of Denmark

**Project Goals:** Our objective is to develop a new standard for the engineering of microbial systems based on rational design, engineering, and optimization of hybrid regulatory networks. We envision a future biorefinery that is based on the development of designer organisms that have exquisite and predictable control architectures governing the expression of a range of valuable traits. Computer aided design platforms will guide the assembly of synthetic constructs containing orthogonal heterologous circuits to recode native regulatory networks. Together, these will enable predictable and dynamic control of multiple designer phenotypes such as: i) growth on various feedstocks in consolidated bioprocesses, ii) feedback control to mitigate accumulation of toxic metabolites, iii) production of target molecules (C3-C4 alcohols), and/or iv) robustness to process upsets (e.g. temp., phage). The focus of this proposal is to develop the technical and computational infrastructure to enable this vision. We will develop this platform first in the model organisms E. coli and S. cerevisiae and then in DOE relevant non-model organisms. (Limit to 1,000 characters)

**Abstract Text:** Please limit such that entire document does not exceed 2 pages.

**Genome engineering for improved protein folding and metabolite production in yeast**

Climate change requires sustainable solutions that permit renewable synthesis of chemicals and fuels toward decreased consumption of fossil fuels. Advances in metabolic engineering, synthetic biology and systems biology have permitted to increase the turnover of the design-build-test cycle of cell factories with high productivity and titers. Although rational design has gained great momentum, the engineering of complex phenotypes has not been possible until recently due to the lack of mechanistic knowledge of these phenotypes and the limitations of random mutagenesis and rational approaches for engineering such phenotypes.

In recent years our group developed CRISPR Enabled Trackable Genome Engineering (CREATE) and applied it in parallel engineering of regulatory networks in *E. coli* and *S. cerevisiae* regulatory networks to gain access targeted complex phenotypes. With such it was possible to build and screen libraries 100,000 and 80,000 variants in *E. coli* and *S. cerevisiae* respectively. Such libraries were screened for improved tolerances to toxic chemicals from renewable feedstocks and short chain alcohols, molecules of interesting properties for use as biofuels.

We are currently leveraging the CREATE approach in yeast to engineer proteins towards improved folding and functionality focusing on enzymes and transcription factor-based biosensors. We are focusing on enzymes involved in the synthesis of natural products and bulk

chemicals that have low activity because of improper folding. We aim to mitigate misfolding by tuning substrate specificity of molecular chaperones. To that end we have mapped substrate binding sites of two major yeast chaperones and designed libraries towards enhanced folding of target enzymes. We will screen such libraries with a misfolding biosensor based on yeast native response to stress. In parallel, we are working in the development of transcription factor-based biosensors for biofuels. As our target transcription factors are of bacterial origin, transplantation into yeast requires engineering of both the transcription factor protein and the transcription factor-responsive promoter. We will engineer both transcription factor and responsive promoter using the CREATE method with the aim of achieve improved dynamic range, operational range, specificity and sensitivity.

Additionally, we are engineering phosphorylation regulatory networks. Phosphosites are one of the main ways that signaling is regulated within the cell. It plays key roles in almost all cellular pathways including cell division and replication, cell/environment interactions and metabolic function. Naturally, a lot of effort has been done in order to map the phosphoproteomes in various organisms, however studies that experimentally validate this data are rare. Here, we attempt to utilize existing phosphoproteomic data in yeast in order to increase cis-cis-muconic acid and isopentanol production. To this end, we designed and implemented metabolic and transcription factor phosphosite mutation libraries that leverage the CREATE approach in order to validate high-throughput phosphosite editing as host engineering strategy for microbial production.

We expect that leveraging the power of designer genome engineering from CREATE together with biosensor-mediated high-throughput screenings will permit as to access newly target phenotypes as optimized protein folding, enhanced biosensors properties and controls of regulatory networks through phophosite engineering.

**Funding Statement:** *Example for grants* - This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012345. *Example for DOE National Laboratories* - Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

**Notes on abstract:**

- Note the placement of superscripts in the authors and affiliations.

- URL above should be specific to the project. More than one URL is permitted. Not all projects will have devoted websites.
- For **References** / **Publications,** use any common style for these citations.
- Use Times New Roman with font size of 12pt

**Title:** IMAGINE BioSecurity: Mesocosm based methods to evaluate biocontainment strategies and impact of industrial microbes upon native ecosystems.

**Authors:** Katie Arnolds[1]*, (katie.arnolds@nrel.gov), Gabriella Li[1], Jeffrey G Linger[2], and **Michael T. Guarnieri**[1]

**Institutions:** [1]Biosciences Center, National Renewable Energy Laboratory, Golden, CO [2]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO

**Website URL:** https://genomicscience-qa.ornl.gov/research/sfas/nrelimagine.shtml

**Project Goals:** This SFA project task seeks to develop a methodological pipeline that utilizes lab mesocosms to evaluate the integrity of novel biocontainment designs and assess the impact of industrial microbes on native environments. Additionally, we are designing the mesocosms with the aim of simulating the complexity of an environmental escape while still being streamlined and reproducible enough that they can be used to broadly evaluate biocontainment strategies in diverse microbes from different trophic regimes. Findings from these systems will be used to elucidate fundamental principles that drive engineered biosystems in their natural and non-native environments.

**Abstract Text:** Industrial production microbes and their associated bioproducts have emerged as an integral component of a sustainable bioeconomy. However, the rapid development of these innovative technologies raises biosecurity concerns, namely, the risk of environmental escape. Thus, the realization of a bioeconomy hinges not only on the development and deployment of microbial production hosts, but also on novel biocontainment designs to reduce risk associated with environmental escape. To date, most biocontainment strategies have been largely evaluated in the lab, often in monocultures and under tightly regulated conditions, thus their escape potential has never been thoroughly evaluated under conditions that even begin to recapitulate an environmental escape, where they would face complex microbial communities, a diverse array of resources and metabolites that may influence escape frequency, as well as increased probability of horizonal gene transfer[1]. Here, we are developing methods that utilize soil mesocosms to evaluate the efficacy of novel biocontainment strategies and to assess the impact of production systems upon terrestrial soil microbiome dynamics. This methodological pipeline will allow us to screen a broad range of biocontainment modules in diverse microbes from different trophic regimes via a down-selection strategy that allows for detection of rare escapees, effect of associated bio-products, and impact on native ecologies. To our knowledge these efforts represent a first in-kind testing pipeline of this rigor and will allow for the establishment of a national standard for the assessment of the safety of industrial microbes and their associated bioproducts, accelerating the realization of a secure bioeconomy.

**References/Publications**

1. Arnolds, K. L. *et al.* Biotechnology for secure biocontainment designs in an emerging bioeconomy. *Curr. Opin. Biotechnol.* **71**, 25–31 (2021).

**Funding Statement:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Secure Biosystems Design Science Focus Area IMAGINE BioSecurity: Integrative Modeling and Genome-scale Engineering for Biosystems Security, under contract number DE-AC36-08GO28308.

**IMAGINE BioSecurity: Biocontainment Efficacy of Toxin-Antitoxin Cassettes in Laboratory and Industrial *S. Cerevisiae***

Gabriella A. Li[1*](gabriella.li@nrel.gov), Katie L. Arnolds[1], Jeffrey G. Linger[2], and **Michael T. Guarnieri[1]**

[1]Biosciences Center, National Renewable Energy Laboratory, Golden, CO [2]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO

https://genomicscience-qa.ornl.gov/research/sfas/nrelimagine.shtml

**Project Goals: The IMAGINE BioSecurity initiative aims to develop genome-scale engineering tools to test fundamental principles that drive biological systems, with the specific goal of conferring enhanced stability, resilience, and controlled performance in DOE-relevant plant and microbial systems. Specifically, this task will design and develop a library of biocontainment modules in *Saccharomyces cerevisiae* and experimentally analyze the growth, escape frequency, and bioproductivity of the engineered strains using high-throughput screening analyses in laboratory and environmental settings.**

**Abstract:** Genetically modified organisms are widely used to produce a variety of bioproducts and fuels. With the increasingly sophisticated genetic engineering used to produce these genetically modified organisms comes the elevated risk of environmental escape. To establish a secure bioeconomy, new biocontainment strategies must be developed and deployed to maintain optimal microbial fitness and production while minimizing the risk of escape. *Saccharomyces cerevisiae,* often referred to as budding yeast, are a highly studied model organism commonly used in the industrial production of various fuels and bioproducts. In order to establish secure biocontainment designs in *S. cerevisiae*, the IMAGINE BioSecurity SFA is pursuing the high-throughput design and screening of a library of toxin-antitoxin biocontainment modules to determine how they affect fitness, productivity, and escape frequency in laboratory (BY4171) and Brazilian (PE2) industrial *S. cerevisiae* strains to inform and improve future biocontainment designs. We hypothesize stacking biocontainment modules will increase efficacy while providing insights into the mechanisms governing biocontainment. Here we report the effects of biocontainment copy number on efficacy and bioproduction finding a two layered system provides longer term biocontainment while not altering ethanol production.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Secure Biosystems Design Science Focus Area IMAGINE BioSecurity: Integrative Modeling and Genome-scale Engineering for Biosystems Security, under contract number DE-AC36-08GO28308.

**Title:** IMAGINE BioSecurity: Biocontainment of Genetically Engineered Cyanobacteria.

**Authors:** Jacob Sebesta[1]* (Jacob.sebesta@nrel.gov), Jianping Yu[1], and **Michael T. Guarnieri**[1]

**Institutions:** [1]Biosciences Center, National Renewable Energy Laboratory, Golden, CO
**Website URL:** https://genomicscience-qa.ornl.gov/research/sfas/nrelimagine.shtml

**Project Goals:** This SFA project task will develop active biocontainment modules that prevent growth of cyanobacteria in the natural environment should they escape the lab or industrial setting. We will also assess the efficacy of a second strategy for biocontainment in which genes responsible for synthesis of storage molecules are knocked out to reduce fitness in the natural environment where fluctuations in nutrient concentrations are common.

**Abstract Text:** Algae (including eukaryotic microalgae and cyanobacteria) have been genetically engineered to convert light and carbon dioxide to many industrially and commercially relevant chemicals including biofuels, materials, and nutritional products. At industrial scale, genetically engineered algae may be cultivated outdoors in open ponds or in closed photobioreactors. In either case, industry must address the potential risk that accidental release of the engineered algae into the natural environment may result in a variety of negative impacts to the environment. Genetic biocontainment strategies are therefore under development to reduce the probability that these engineered bacteria can survive outside of the laboratory or industrial setting. As part of the IMAGINE SFA project, we are testing strategies for biocontainment of strains of *Synechocystis* sp. PCC6803 that have been engineered to produce ethylene. We are combining two strategies: 1) knockout of storage molecule synthesis genes, and 2) active lethal genes induced by an environmental signal.

Knockout of storage molecule synthesis genes is expected to reduce fitness of engineered cyanobacteria if they escape to the natural environment where fluctuations in nutrient concentrations are common and elemental storage is important for surviving periods of low nutrient concentrations. We have initially targeted the genes responsible for synthesizing storage molecules of carbon (glycogen), phosphorous (polyphosphate), nitrogen (cyanophycin), and iron (bacterioferritin) as well as a gene essential for salt acclimation. In addition to these knockouts, we have designed active biocontainment modules which can sense that the cells are no longer in the lab setting and induce the expression of lethal genes. Initially, we are testing a biocontainment module in which expression of the RNase, *mazF*, is controlled by the promoter, PisiAB, which is induced by iron limitation. Biocontainment efficacy will be assessed in conditions that are representative of what the cells are subject to in the natural environment. These strains engineered for biocontainment will be assessed for fitness and bioproductivity, including ethylene production, as a function of biocontainment constraints. The resultant data will be integrated into genome-scale models to establish a predictive framework for optimal biodesigns.

**Funding Statement:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Secure Biosystems Design Science Focus Area IMAGINE BioSecurity: Integrative Modeling and Genome-scale Engineering for Biosystems Security, under contract number DE-AC36-08GO28308.

**Title:** Combinatorial Biocontainment Design and DNA Barcode Genotyping

**Authors:** Ayako Murao,[1] Diána Hernández Hernández,[1] Melissa Amezola,[1] Lin Ding,[1] Jacob Sebesta,[2] Bin Yang,[2] Gabriella Li,[2] Katie Arnolds,[2] Cristal Zuniga,[3] Rodrigo Santibanez Palominos,[3] Marcus Bray,[3] Christopher Johnson,[2] Karsten Zengler,[3] Wei Xiong,[2] Jeffrey Linger,[2] Jianping Yu,[2] **Michael Guarnieri**,[2] and Yo Suzuki[1*] (ysuzuki@jcvi.org)

**Institutions:** [1]J. Craig Venter Institute, La Jolla, California; [2]National Renewable Energy Laboratory, Golden, Colorado; and [3]University of California San Diego, La Jolla, California

**Project Goals:** Synthetic biology promises a transformative bioeconomy, but much work remains in ensuring biocontainment for engineered organisms. A unique challenge for biocontainment in an industrial setting is that massive culture-scale necessitates ultra-low escape frequencies, while it is undesirable if genetic mechanisms for control interfere with organism fitness and biosynthesis capabilities. The IMAGINE team aims to solve this challenge with combinatorics and synergy, where multiple mechanisms that control viability in different ways with little adverse effect on the valuable properties of the microbes are combined to generate robust biocontainment for ensuring biosafety. The overall goal of our project is to examine the incorporation of synergy among mechanisms for biocontainment as a widely applicable strategy for establishing safe industrial organisms without sacrificing efficiencies for bioproduction.

**Abstract Text:** To facilitate the analysis of combinatorial constructs in our target organisms, a method termed combinatorial genetics *en masse* (CombiGEM; Wong et al., 2016) was implemented. In this method, the recursive cloning of DNA-barcoded modules via designated restriction sites within the modules results in the accumulation of constructs of interest on one side and the DNA barcodes on the other side of the restriction sites. The concatenated barcodes reveal the identity of the multiple modules in each strain. To conduct the CombiGEM process for biocontainment modules, we generated the base plasmid pCombi-CC (copy control) that can be maintained in *E. coli* as a single-copy entity to cope with any toxicity from biocontainment genes and can also be induced to increase its copy number (Lucigen). pCombi-CC enables the recursive assembly of modules using the BamHI, BglII, EcoRI, and MfeI sites. We demonstrated this in an experiment with four mock biocontainment modules. To integrate next-generation sequencing (NGS) into the analysis of concatenated DNA barcodes that mark our combinatorial constructs, we incorporated both a widely used amplicon sequencing approach (Caporaso et al., 2011) and the standard Illumina TruSeq process. The compatible pCombi-CC-Next vectors contain part of the TruSeq Index Adapter sequence, so that a single primer without any strong secondary structure can be used to introduce an index for multiplexing and complete the construction of a library to be sequenced with universal primers. With this approach, we can obtain 1 million reads for a concatenated DNA barcode library for $7 using the Illumina NovaSeq platform. With PCR optimization, artifactual reads were reduced to 0.01% of the whole. A computational analysis pipeline for rapid data analysis was also established. Toward applying the CombiGEM strategy to the analysis of biocontainment mechanisms, we completed

the initial design of the vector, regulatory sequences, and six toxin systems for our five target industrially relevant organisms, *Saccharomyces cerevisiae*, *Pseudomonas putida*, *Synechocystis* sp. PCC 6803, *Clostridium ljungdahlii*, and *Mycoplasma mycoides*. This design reflects the current organism-specific strategies the IMAGINE team is taking, and some of the molecular tools are being developed. In *P. putida*, there are effective tools for inducible gene expression. We are using the AraC-$P_{BAD}$ system (Gauttam et al., 2021) for driving the expression of antitoxin genes. However, there is a shortage of tools for inducible repression needed for inactivating toxin genes when the organism is in a growth-permissible space. Our group is developing a system with nickel-activated Nik repressor for this purpose. Transitioning from the designed sequences to the actual DNA constructs and strains will be our next challenge. With the combinatorial approach, we expect to generate hundreds of genotypes, each corresponding to a particular combination of biocontainment modules, in each organism and use the abundances of concatenated DNA barcodes in mixed populations to elucidate the synergistic interactions among the biocontainment modules tested.

**References/Publications**

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences, USA *108*, 4516–4522.

Gauttam, R., Mukhopadhyay, A., Simmons, B.A., and Singer, S.W. (2021). Development of dual-inducible duet-expression vectors for tunable gene expression control and CRISPR interference-based gene repression in *Pseudomonas putida* KT2440. Microbial Biotechnology *14*, 2659–2678.

Wong, A.S.L., Choi, G.C.G., Cui, C.H., Pregernig, G., Milani, P., Adam, M., Perli, S.D., Kazer, S.W., Gaillard, A., Hermann, M., et al. (2016). Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proceedings of the National Academy of Sciences, USA *113*, 2544–2549.

**Funding Statement:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Secure Biosystems Design Science Focus Area IMAGINE BioSecurity: Integrative Modeling and Genome-scale Engineering for Biosystems Security, under contract number DE-AC36-08GO28308.

**Title:** Modeling Bacterial Metabolism and Expression to Develop Biocontainment Strategies

**Authors:** Juan D. Tibocha-Bonilla[1]* (j1tiboch@eng.ucsd.edu), Rodrigo Santibanez[2]* (rsantibanezpalominos@health.ucsd.edu), Blake Estefan,[3] Brian Arianpour,[3] Kanksha Patel,[3] Cristal Zuniga,[2] **Karsten Zengler**[2,3,4], and **Michael Guarnieri**[5]

**Institutions:** [1]Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0760, USA; [2]Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0760, USA; [3]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA; [4]Center for Microbiome Innovation, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0436, USA; [5]National Renewable Energy Laboratory, Golden, CO, USA.

**Website URL:** https://genomicscience.energy.gov/research/sfas/nrelimagine.shtml

**Project Goals:**

To develop computational tools for the predictive design of biocontainment strategies with enhanced stability and resilience in diverse microbial hosts while maintaining maximal fitness and bioproductivity of the engineered microbial strains. To do so, we are developing genome-scale metabolic and expression models (ME-models) for a selected group of industrially relevant bacteria and exhaustively reviewing developed metabolic models (M-models) to employ them. We will 1) predict the impact of biocontainment strategies in the fitness and productivity of industrial strains, 2) develop community metabolic models (CM-models) to determine possible microbe-microbe interactions and the subsequent capability of a native ecosystem to support the growth of an “escaped” industrial strain, and 3) develop novel biocontaiment strategies based on a conditional metabolism that could not be rescued on natural ecosystems.

**Abstract Text:**

Genome-scale metabolic models (GEMs) of industrial relevant bacteria are largely employed to determine genetic interventions for optimizing the production of a target metabolite. However, strategies for the biocontainment and eliminating the impact on native ecosystems if they escape from an industrial environment are lacking behind. Here, we present manually validated and curated GEMs of four industrially relevant bacteria and the first ME-model for *Pseudomonas putida* KT2440 (iJT1667-ME). To do so, we developed a bioinformatics pipeline to validate and curate those and other 25 published bacterial GEMs and obtain M-models having identical identifiers to be used in CM-models, and we established a methodology around the COBRAme software to develop bacterial ME-models.

The validation and curation pipeline obtains and combines the annotations derived from the KOFAM and the InterProScan software with the annotation of the best protein sequence

alignment against the UniProt, KEGG, Prokka, and TCDB databases. Our UniProt database contains all protein sequences from the SwissProt database and a filtered TrEMBL database, keeping only protein sequences with a "catalytic activity" annotation and removing sequences with a "fragment" annotation. Similarly, we keep only protein sequences larger than 11 amino acids and that are annotated with an ontology number from the KEGG database. The resulting UniProt and KEGG databases contain 24.1 and 18.2 million sequences, respectively, covering 10.7% and 52.1% of all proteins in the original databases. After bioinformatics analyses, we manually reviewed the annotations and assigned a "high-confidence" validation score when the annotation agreed and coincide with the reaction consulting BRENDA, KEGG, IntEnz, and RHEA databases. Finally, the GEM is curated and updated to accurately reflect information from databases and publications. In addition, genes that were not validated were discarded from the model.

We successfully designed a methodology to process data available in diverse databases, such as BioCyc, to COBRAme-ready files to aid ME-model reconstruction. This approach allowed for the rapid development of the first ME-model for *P. putida* KT2440. The *i*JT1667-ME model consists of 7,110 metabolites, 10,955 reactions, and 1667 genes, covering 30.8% of the 5,419 proteins of *P. putida* KT2440 proteome. *i*JT1667-ME includes a curated metabolic network of the *i*JN1462 model with an additional manual curation to incorporate ion and cofactor transport reactions, as well as resolution on the strain-specific gene expression machinery that resulted in the incorporation of 205 additional genes into the reconstruction. Simulations showed the correct minimization of metabolic loops, such as the ATP:polyphosphate phosphotransferase and L-lysine:pyruvate aminotransferase reactions, and significant differences in the prediction of metal ion transport, NAD synthesis, and pentose phosphate pathway fluxes compared to *i*JN1462. Next steps will involve the modeling of biocontainment strategies to determine *in silico* fitness and productivity.

**Funding Statement:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Secure Biosystems Design Science Focus Area (SFA) contract number DE-AC36-08GO28308.

**IMAGINE BioSecurity: Metabolic Modeling-Enabled Biocontainment Redesign in Microbial Chasses**

**Authors:** Bin Yang[1], Chao Wu[1], **Michael T. Guarnieri**[1], Wei Xiong[1]*(wei.xiong@nrel.gov)

**Institutions:** [1]Biosciences Center, National Renewable Energy Laboratory, Golden CO 80401

**Project Goals:** Genetically modified organisms (GMOs) are widely used in agriculture and bioenergy industries, raising biosafety concerns about unintentional cellular proliferation or accidental environmental spread of synthetic genes. Supported by DOE BER Scientific Focus Area (SFA) program, we aim to design predictable and generalizable biocontainment strategies to prevent the potential hazards caused by GMO microbes.

**Abstract Text:** Our research goals in this SFA subtask are twofold. *First*, we focus on the development of new tools that can characterize metabolic responsiveness of microbial hosts to genetic safeguards as well as biocontainment constraints in controlled and simulated ecosystems. In this respect, we have successfully developed machine learning (ML)-$^{13}$C-fluxomics in which ML algorithm "learns" intrinsic relationship between $^{13}$C-labeled metabolites and metabolic flux. Thus, the tool can read out the fluxomic phenotypes "directly and immediately" from isotopically metabolomic dataset [1]. One key merit of this tool is that the training dataset for ML can be generated by algorithm, such that it does not have to rely on experimental data, which is not always sufficiently available nor covering all realistic scenarios. The solvability of metabolic flux depends on the linear dependence of the adjoint labeling patterns of precursor metabolites, which is determined by the topology of the network and the labeling strategy of substrates. In our tool, screening for solvability is automatically performed. It eliminates the invalid variables and therefore increases the accuracy of prediction. The presented ML approach greatly reduced computation time for metabolic flux estimation down to <1 s in an exemplary *E. coli* model. Employing it for large scale flux analysis, high-throughput metabolic phenotyping and strain screening are enabled. Development of this efficient fluxomics approach for biocontainment hosts will promote iterative Design-Build-Test-Learn cycle and help biosystems redesign for safe production of next-generation biofuels.

*Second,* we are developing novel biosecurity strategies specifically in industrial clostridia. One strategy is based on metabolic robustness control, a generalizable design principle from inherent properties of metabolism [2]. This approach offers a unique capability to predict and govern the fitness of a dynamic biological system, especially when the system exhibits differential behavior in laboratory and natural conditions. In a modeled biocontainment circuit (Figure 1), for example, the GMO microbe will exhibit steady-state metabolism when a laboratory effector stabilizes the function of a key enzyme. In an uncontrolled environment, where the effector is absent, the targeted enzyme will be down regulated, thus depleting the pool of the vital metabolite(s) and resulting in instability of the GMO during escape. Metabolic robustness can thus be analyzed in the context of biocontainment modules by ensemble modeling, where a set of models with different kinetic data are parameterized and perturbed by varying maximum rate

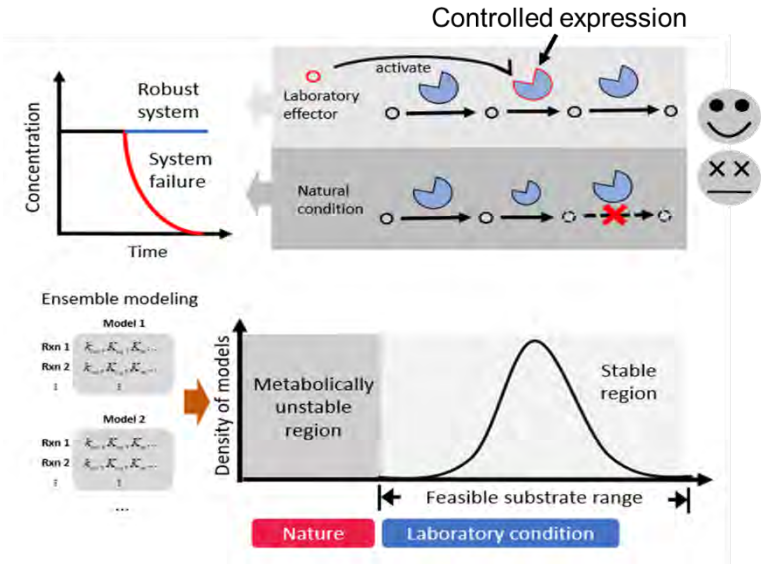



**Figure 1** Predictive biosecure design based on ensemble modeling and robustness control [2].

($V_{max}$), which is largely proportional to the control level of the enzyme. This approach counts the probability of system failure per perturbation. Such approaches can ultimately pinpoint new metabolic targets for optimal biosecure design, as well as assess viability of a modified laboratory organism in response to environmental changes. Currently, we have developed Ensemble Modeling for Robustness Analysis (EMRA) and are identifying in *Clostridium ljungdahlii* the most likely gene targets, modulation of which may predictably affect cell growth. Experimentally, we are developing a state-of-the-art multiplexed genome editing approach as well as inducible switch-offs for robustness control of clostridia. The aim is to enable simultaneous targeting of multi-sites for growth arrest, and to minimize escape efficacy.

**References/Publications**


1. Chao Wu, Jianping Yu, Michael Guarneri, Wei Xiong. A Computational Framework for Machine Learning-Enabled $^{13}$C-Fluxomics. ***ACS Synthetic Biology.*** 2021 10.1021/acssynbio.1c00189 (Cover Art)
2. Kathleen L. Arnolds, Lukas R. Dahlin, Lin Ding, Chao Wu, Jianping Yu, Wei Xiong, Cristal Zuniga, Yo Suzuki, Karsten Zengler, Jeffrey G. Linger, and Michael T. Guarnieri. Biotechnology for Secure Biocontainment Designs in An Emerging Bioeconomy. ***Current Opinion in Biotechnology***. 2021,71(10):25-31




**Funding Statement:** *This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Secure Biosystems Design Science Focus Area IMAGINE BioSecurity: Integrative Modeling and Genome-scale Engineering for Biosystems Security, under contract number DE-AC36-08GO28308.*

**Title:** The Soil Lipidome is a Robust Indicator of the Microbial Community Response to Rewetting Following a Summer Drought

**Authors:** Sneha P. Couvillion[1*](sneha.couvillion@pnnl.gov), Robert E. Danczak,[1] Dan Naylor,[1] Montana L. Smith,[1] Kelly G. Stratton,[1] Vanessa L. Paurus,[1] Kent J. Bloodsworth,[1]Yuliya Farris,[1] Darren J. Schmidt,[1] Rachel E. Richardson,[2] Lisa M. Bramer,[1] Sarah J. Fansler,[1] Ernesto S. Nakayasu,[1] Jason E. McDermott,[1,3] Thomas O. Metz,[1] Mary S. Lipton,[1] Janet K. Jansson,[1] and **Kirsten S. Hofmockel**[1]

**Institutions:** [1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA; [2]National Security Directorate, Pacific Northwest National Laboratory, Richland, WA; [3]Department of Molecular Microbiology and Immunology, Oregon Health & Science University, Portland, OR

**Website URL:** https://www.pnnl.gov/projects/soil-microbiome/research

**Project Goals:** PNNL's Phenotypic Response of Soil Microbiomes SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions throughout the soil profile. Data are used to parametrize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in lab and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

**Abstract Text:**

Extreme environmental change such as a severe drought followed by a precipitation event can trigger a complex cascade of microbial physiological responses, impacting soil microbial community structure and functions. Lipids are vital to microbial structural and signaling functions, yet the effect of environmental change on soil microbial lipids is only beginning to be explored. To reveal the importance of lipids in regulating microbial physiological responses to environmental stress, we used untargeted lipidomics, 16S rRNA gene and ITS region sequencing, to monitor an arid grassland soil microbiome for 3 hours following rewetting in a summer drought incubation experiment.

We show that the soil lipidome is rapidly remodeled following rewetting and the distinct changes in lipids are important for stress adaptation, substrate use, and recovery during drought and subsequent rewetting. Comprehensive coverage of a broad range of lipid classes combined with our ability to characterize the fatty acid compositions of the measured lipids is an important

advantage of our work, enabling the discovery of previously unknown responses at the lipid subclass and fatty acid level. Our findings suggest that lipids may be critical in orchestrating the broad differences in stress response strategies used by bacteria and fungi to survive environmental stress. Drought resulted in an increase in lipids implicated in mediating heat, osmotic and oxidative stress and nutrient deprivation. Drought also induced elevated levels of lipids containing fatty acid moieties that were characteristic of fungal metabolism. The increase in lipids with fatty acids typical of bacteria following wetting suggested rapid metabolic reactivation in the bacterial community as nutrient diffusion increased and conditions became more favorable for growth.

These results underscore the importance of the soil lipidome as a robust indicator of microbial community responses, even at short time scales. We demonstrate the immense potential for harnessing information from the soil lipidome to understand how soil microbiomes adapt and respond to stress.

**Funding Statement:** PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 70880. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL.

# Distribution of soil microbial necromass accumulation controlled by microbe-mineral interactions

**Kirsten S. Hofmockel**[1*] (kirsten.hofmockel@pnnl.gov), Qian Zhao[1], Sheryl L. Bell[1], Sarah Leichty[1]
**Institutions:** [1]Environmental and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

**Project Goals: The overall goal of this project is to test if plant-microbe interactions are limited to influencing the rate of C accrual, while mineralogy regulates the sink capacity of biofuel cropping systems. To accomplish this goal, we are identifying the microbial functions and biopolymers of microbial necromass that contribute to soil C accumulation under controlled conditions and characterizing microbial necromass accumulation in response to crop selection and edaphic factors in situ.**

Abstract:
Soil organic matter (SOM) is a reservoir for global carbon (C) that maintains soil fertility and prevents the release of greenhouse gases. A major contributor to SOM is necromass, consisting of dead microbial cells and cellular components. Yet methods for enhancing the accumulation of microbial residues and necromass in soil are still unknown. We hypothesized that the quantity of microbial residues that accrue and persist is positively correlated with the abundance of poorly crystalline iron minerals in soil. To test this hypothesis, we incubated soils from two agricultural research sites varying in soil texture and background C content, derived from the Great Lakes Bioenergy Research Centers. Sandy and silty loam soils were incubated with $^{13}$C-glucose for 12 months to trace the fate of microbial-derived residues. We found an average of 30% of added $^{13}$C was recovered as microbial necromass after the course of the incubation. Density fractionation was used to separate light and heavy fractions of the mineral-associated organic matter (MAOM). Approximately 88% of the total recovered $^{13}$C labeled necromass was found in the fine (<53 μm), MAOM fraction. The light MAOM fraction, which is typically dominated by organo-mineral complexes, accumulated three times more $^{13}$C necromass than the heavy MAOM fraction. Further investigation via Mössbauer spectroscopy revealed that the light MAOM contained more amorphous iron-bearing minerals, whereas the heavy MAOM fraction had more phyllosilicates. Specifically, light MAOM had twice the amount of nano-size goethite and organo-Fe complexes compared with heavy MAOM. Most interestingly, light MAOM of the sandy loam soil had up to 11 times the ferrihydrite content as the heavy MAOM. This high concentration of amorphous iron-bearing minerals likely contributed to the accumulation of microbial-derived necromass in light MAOM via sorption or complexation. In addition, nanoscale secondary ion mass spectrometry (NanoSIMS) images revealed more $^{13}$C enriched hotspots in light MAOM than that in heavy MAOM, supporting the abundance of microbial-derived necromass in light MAOM after the long-term incubation. Surface C chemistry analysis by X-ray photoelectron spectroscopy (XPS) revealed that light MAOM fraction had a lower abundance of alkyl or aryl compounds commonly derived from plant detritus (i.e., carbohydrates or lignin) and a higher abundance of carbonyl or

carboxylic compounds (i.e., proteins) compared to heavy MAOM. By labeling and tracing microbial residues, we demonstrate the importance of fine amorphous mineral surfaces in accumulating microbial-derived necromass in both sandy and silty loam soils.

*This research was supported by an Early Career Research Program award to KS Hofmockel, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP 68292. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.*

**Profiling of the Microbiome Metabolic Response to Soil Rewetting**

Mary Lipton[1*](mary.lipton@pnnl.gov), Montana Smith[1], Lisa Bramer[1], Sheryl Bell[1], Nikola Tolic[1], Sneha P. Couvillion[1], Vanessa Paurus[1], Emily Graham[1], Janet Jansson[1], **Kirsten Hofmockel**[1]

[1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland WA, 99352

**Website URL:** https://www.pnnl.gov/projects/soil-microbiome/research

**Project Goals:** PNNL's Phenotypic Response of Soil Microbiomes SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions throughout the soil profile. Data are used to parametrize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in lab and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

**Abstract:** Soil microbes are highly sensitive to changes in their environment, making rapid measurements essential for a comprehensive understanding of microbiome functions. For example, drought is one of the most common environmental stresses that soil microbiomes experience, with important implications for larger scale biogeochemical fluxes. Water drives microbial interactions because it influences everything from cell function to substrate fate within the soil system. Changing levels of soil moisture can affect the retention or loss of soil carbon. When soil organic carbon is rapidly mineralized from desiccated soils upon rewetting there is a rapid increase in the release of $CO_2$. It has also been observed that the release of $CO_2$ increases with the length of time that the soil has been exposed to the drought conditions. Understanding the molecular mechanisms behind this phenomenon is important for managing soil carbon, especially with the increased frequency of drought-rewet events due to climate change.

Here, we aimed to understand how the physiology, metabolism, and interactions between soil microorganisms change in response to desiccation and rewetting, and to use this understanding as a basis for predicting the soil metaphenome. Soil samples were collected from our irrigated tall wheatgrass field experiment, in Prosser, WA. This soil was used to in an incubation experiment where soils were exposed to dehydration, allowed to incubate in the desiccated state for 1 or 2 weeks and then rehydrated for 90 min. The metaphenome was evaluated by measuring real-time respiration, carbon and nitrogen contents, as well as metabolite, protein, and transcript profiles throughout the time course.

Our results revealed that microbial biomass carbon decreased through the dehydration period and additionally during the rewetting process, with similar results for both the 1 week and 2 week drought exposures. The microbial biomass nitrogen also decreased during the dehydration event. However, no significant changes were observed during rehydration suggesting that the microbial biomass C:N ratio decreased over the course of the experiment. Additionally, the reduction of microbial carbon during rewetting suggests that the microbes are not reassimilating the carbon immediately after rewetting (90 min). Instead, this carbon is being directed toward respiration. Real time mass spectrometry methods (Weitz et al., 2020) detected an immediate burst of $CO_2$ production within the first 10 minutes upon rewetting and a gradual increase during the subsequent 90 minutes. Total $CO_2$ respired from soil exposed to two weeks of drought was 75.6% higher than soil exposed to one week of drought, consistent with the Birch effect (Birch, 1958).
Metabolomic analysis was used to characterize the microbiome response to the dehydration and rehydration processes. GC-MS based metabolomics analysis identified 336 metabolites over the course of the experiment. While most metabolites did not change significantly throughout the incubation, 20 metabolites decreased, and 19 metabolites increased during dehydration. Conversely, 23 metabolites decreased and10 increased during rehydration. These results illustrate changes that occurred in the soil metabolic profile in response to changing soil moisture levels. Specific metabolites changed during both phases of the experiment. For example, lyose, xylitol and other sugars accumulated during dehydration and decreased after rehydration, although mannitol increased throughout the experiment. These observations support the hypothesis that soil microbes accumulate sugars and other compounds during dehydration to protect against osmotic stress. The metabolites of interest are currently being mapped to metabolic pathways to identify potential metabolic strategies that are active during the dehydration and rehydration response.

These results indicate that soil microbial communities react rapidly to changing moisture conditions. Understanding the mechanisms underlying these changes will have important implications for carbon allocation within and between organisms in the soil environment.

**References/Publications**


1. Weitz K.K., Real-Time Mass Spectrometry Measurements of Respiration Rates in Biological Systems. *J. Am. Soc. Mass Spectrom.* **2021**, 32, 3, 648–652
2. Birch, H. F., Pattern of Humus Decomposition in East African Soils. *Nature* **1958,** *181* (4611), 788-788.



**Funding Statement:** PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 70880. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL.

**Coordination of species roles during chitin decomposition in a model soil microbial consortium**

Ryan McClure[1] (Ryan.mcclure@pnnl.gov), Yuliya Farris[1], Robert Danczak[1], William Nelson[1], Hyun-Seob Song[3,4], Aimee Kessell[3], Joon-Yong Lee[1], Sneha Couvillion[1], Chris Henry[5], Janet K. Jansson[1] and **Kirsten S. Hofmockel**[1*]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA, 99352
[3] Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE 68583
[4]Department of Food Science and Technology, Nebraska Food for Health Center, University of Nebraska-Lincoln, Lincoln, NE 68588
[5]Data Science and Learning Division, Argonne National Laboratory, Lemont IL

**Website URL:** https://www.pnnl.gov/projects/soil-microbiome/research

**Project Goals:** PNNL's Phenotypic Response of Soil Microbiomes SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions throughout the soil profile. Data are used to parametrize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in lab and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

**Abstract**: Soil microorganisms are the drivers underpinning key ecological functions, including plant growth promotion and nutrient cycling. These important processes rely on metabolic interactions between individual populations of the soil microbiome. Despite the importance of microbial interaction networks to the emergent properties of decomposition and nutrient cycling, the complexity of soil makes analysis and data interpretation difficult. Here, we focused on a simplified, defined, and representative community of eight bacterial species, MSC-2, as they interact to metabolize chitin, an abundant carbon source in soil.

Using a multi-omics approach we show that both species and community level processes during chitin decomposition were distinct when comparing monocultures of individual members to co-culture growth of the complete MSC-2 community. In addition, emergent properties of both specific species and the community were found. While certain members of MSC-2 showed poor growth on chitin in monoculture our metabolomic and metatranscriptomic analysis suggests that

these same species, when cultured within the context of the complete MSC-2 community, contribute to chitin metabolism. The dominant metabolically active members within MSC-2 were further evaluated to determine their specific roles during chitin breakdown. Intriguingly, the most highly abundant members of MSC-2 were not those that could metabolize chitin, but rather those that could grow best on the resulting breakdown products. The use of multi-omics also allowed us to map the path of chitin through this community, assigning certain species to certain steps of chitin breakdown. This mapping revealed which taxa are critical to different aspects of chitin breakdown and revealed how, in environments where chitin is a major nutrient source, carbon cycling could be disrupted through the loss of certain taxa.

This study not only greatly increased knowledge of how chitin is metabolized by soil microbial communities, but also provided new details underpinning interspecies interactions that are key to global carbon and nitrogen cycling in soil. These conclusions will be critical to our understanding of how native soil microbiomes process C sources, especially those such as chitin that drive interactions and metabolite sharing, and how these processes may shift as community membership changes as a function of both biotic and abiotic pressures. Application of these conclusions to the native soil microbiome will greatly expand our ability to identify what the keystone species are, who may have the greatest advantage for growth and how these communities are organized to promote C cycling in natural settings.

**Funding Statement:** PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 70880.A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL.

**Title:** Three Feet Deep: Abiotic and Biotic Drivers of Organic and Mineral Soil Carbon Cycling

**Authors:** Katherine I. Naasko (katherine.naasko@pnnl.gov),[1,2*] Dan Naylor, [1] Emily Graham, [1] Sneha Couvillion, [1] Joon Y. Lee, [1] Robert Danczak, [1] Nikola Tolic, [1] Carrie Nicora, [1] Steven Fransen, [3] Haiying Tao, [2] Janet K. Jansson,[1] and **Kirsten S. Hofmockel[1]**

**Institutions:** [1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA; [2]Department of Crop and Soil Sciences, Washington State University, Pullman, WA; [3]Department of Crop and Soil Sciences, Washington State University, Prosser, WA

**Website URL:** https://www.pnnl.gov/projects/soil-microbiome/research

**Project Goals:** PNNL's Phenotypic Response of Soil Microbiomes SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions throughout the soil profile. Data are used to parametrize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in lab and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

**Abstract Text:** This study focuses on the influence of soil moisture and plant cover on multi-omic, microbial, and chemical indicators of carbon (C) cycling at different depths through the soil profile. Soil C is known to be metabolized through the soil-plant-microbe continuum in interrelated processes that determine the balance of C sequestered in soil or released back into the atmosphere. Long-term storage of soil C is thus dependent upon resource use efficiency by soil microorganisms and the complex interplay with the physiochemical soil matrix of organic and mineral forms of C. In surface soils, soil organic C (SOC) cycling is regulated by soil pH, moisture, and plant cover. Deep soil profiles are characterized by resource gradients of soil pH, moisture, nutrient availability, and mobile plant and microbial residues. A unique opportunity to understand mineral and organic carbon interactions through the soil profile is presented in the case of calcareous soils – characterized by high soil pH and calcium throughout the soil profile with carbonate-containing horizons a few feet deep. Calcium carbonate-containing marginal soils hold the potential for long-term storage of select forms of C as presented in our study.

We hypothesized that soil biochemical diversity would increase with plant cover and decrease with depth into the soil profile. Furthermore, we hypothesized that differences in soil moisture would result in differences in soil C chemistry with depth. We tested our hypotheses using a multi-omics approach to determine the composition and diversity of the soil microbiome, and the metabolome, lipidome, and proteome from 0 to 1 m depth into the soil profile, and how these relate to soil

moisture and C chemistry. We analyzed soils from our tall wheatgrass (*Thinopyrum ponticum*) irrigation experiment located in the Columbia Basin of Washington state. Samples were collected from the drip irrigated plots and adjacent non-irrigated bare soil. Irrigated plots received surface drip irrigation weekly based on crop evapotranspiration, monitored by WSU, at a rate of 18.75% of the recommended application to impose drought stress.

In total, 336 polar metabolite features, 649 lipids, and 19,165 proteins were detected in 48 soil samples. Greater depths were characterized by higher soil pH and calcium, lower SOC and SOM, and higher abundances of select lipids and organic acids. Of the organic acids which responded to depth, calcium was most strongly positively correlated to benzoic acid and most strongly negatively correlated to linoleic acid. SOM was also strongly predicted by linoleic acid and metal micronutrients which suggests that linoleic acid interacts in complexes between metals, cations, and SOM.

Depth's effects on soil chemical and metabolome composition were variable by field treatment. Nonirrigated bare soils contained significantly greater TOC, TN, and trehalose from 0-5 cm and lower carbonate as a polar metabolite, calcium, and magnesium compared to the 48-100 cm horizon. By contrast, in irrigated bare soils, the lower 48-100 cm horizon contained greater soil moisture and benzoic acid, and lower potassium and trehalose compared to the 0-5 cm horizon.

Soil proteome structure was best explained by soil pH compared to all other soil chemical properties measured. Soil proteome composition differed by sampling depth: the top 0-5 and 5-15 cm were characterized by significantly greater di-haem oxidoreductase activity, bacterioferritin activity, bacterial lysR proteins, outer membrane receptor proteins involved in mostly Fe transport, and gram-negative porin protein abundance compared to the lower 48-100 cm horizon. Di-haem oxidoreductase activity was detected in 40 samples. Acetoacetate decarboxylase activity was detected in 39 samples and ranked as one of the most commonly abundant protein signatures in the 15-48 cm and 48-100 cm horizons.

Our results present a novel soil dataset revealing chemimicrobial interactions with SOM and minerals in calcareous, marginal soil systems. This rich dataset and multi-omics perspective highlights the contrasting importance of plant-microbe interactions in the formation of SOM in surface soils and salt-metal-mineral interactions in stabilizing deep soil C in arid marginal soils.

**Funding Statement:** PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U.S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 70880. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL.

**Title:** Visually Mapping Phenotypes and Community Interactions at the Microbial Scale

**Authors:** Natalie C. Sadler,[1*] (natalie.sadler@pnnl.gov), Hyun-Seob Song,[2,3] Joon-Yong Lee,[1] Arunima Bhattacharjee[1], Song Feng[1], Ryan Mcclure,[1] and **Kirsten S. Hofmockel[1]**

**Institutions:** [1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA; [2]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE; [3] Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, NE;

**Website URL:** https://www.pnnl.gov/projects/soil-microbiome/research

**Project Goals:** PNNL's Phenotypic Response of Soil Microbiomes SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions throughout the soil profile. Data are used to parameterize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in lab and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

**Abstract Text:** The physical and biochemical interactions between microbial community members drive the rate and extent of soil organic matter (SOM) decomposition. However, soil is a heterogeneous and porous environment in which microbial connectivity to other cells and nutrients is largely dependent on moisture levels. An important but challenging scientific objective is gaining a predictive understanding of how the dynamic interplay between key microbial features (e.g., motility and enzyme/metabolite production strategies) and soil aqueous phase variability affect SOM decomposition rates. The challenges associated with this objective are extensive due to the inherent chemical, physical, and biological complexity of soil. To address these challenges, we have developed a suite of tools that can generate spatially explicit cellular and molecular level insights into the microbial phenotypes and interactions involved in SOM degradation in a soil-like environment. The newly developed tools include 1) fluorescent protein (FP) expressing bacterial strains isolated from a naturally evolved soil microbial community (McClure et al. 2020) allowing for tracking of species abundance and co-aggregation in real time, 2) a porous and transparent artificial soil habitat, or Soil Chip, that is well suited for fluorescence based and MS imaging.

Using these tools, we generate empirical data that can be used to test and parameterize microbial explicit models (agent-based and reaction-diffusion models) and to test hypotheses. In our Soil Chips we hypothesize that both microbial functionality (motility and metabolic capacity) and C substrate complexity and solubility dictate strain growth profiles and assembly over time.

Specifically, motile strains will have greater biomass than non-motile strains when a C substrate is insoluble, and producer strains that synthesize C substrate degrading enzymes will have lower biomass over time than cheater strains. Additionally, competing strains will assemble to form larger aggregate patches in an attempt to privatize resources. To test these hypotheses, we conducted experiments designed to map strain level assembly patterns, growth and motility of a microbial consortium treated with insoluble chitin or freely diffuse chitin products. To do so, phylogenetically diverse FP tagged bacteria including *Rhodococcus sp. S2-17::mTagBFP2*, *Variovorax paradoxus::mScarlet-I*, and *Sphingopyxis fribergensis::mClover3,* were incubated for 4 days in saturated Soil Chips treated with either chitin, chitopentaose (chitin oligomer), or n-acetyl glucosamine (NAG; chitin monomer). Unlike NAG, chitin and chitopentaose require extracellular enzymes to degrade the substrates into products that can be taken up by cells and utilized for growth. Furthermore, chitin was localized to an area ~700 μm away from cell inoculum port, whereas NAG and chitopentaose diffused freely throughout the Soil Chips. Soil Chips were imaged regularly by confocal laser scanning microscopy.

Our results indicate that when treated with NAG, the consortium generated greater unmixed strain aggregate patches than when treated with chitopentaose and chitin. This suggests competitive interactions as all strains can assimilate NAG. When treated with chitopentaose, the chitinolytic strains, non-motile *Rhodococcus* and motile *V. paradoxus,* started to grow by day 1, whereas the non-chitinolytic and motile *S. fribergensis* grew to a high density and likely outcompeted *V. paradoxus* growth by day 4. With chitin, we see the most even strain distribution and the smallest strain patch sizes by day 4, evidence of more cooperative than competitive strain interactions. Our initial hypothesis is supported by these preliminary results and will be more rigorously tested through the comparison with simulation results from predictive spatiotemporal interaction models that are being developed. This research will serve as our benchmark for future studies designed to test moisture effects on microbial phenotypes and chitin decomposition rates.

**References**

1. McClure R, Naylor D, Farris Y, Davison M, Fansler SJ, Hofmockel KS, Jansson JK. Development and Analysis of a Stable, Reduced Complexity Model Soil Microbiome. Front Microbiol. 2020 Aug 26;11:1987. doi: 10.3389/fmicb.2020.01987. PMID: 32983014;

**Funding Statement:** PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 70880. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL.

**Title:** Omics-enabled global gapfilling (OMEGGA) for phenotype-consistent metabolic network reconstruction of microorganisms and communities

**Authors:** Hyun-Seob Song,[1,2*] (hsong5@unl.edu), Joon-Yong Lee,[3] Firnaaz Ahamed,[1] Aimee K. Kessell,[1] William C. Nelson,[3] David M. L. Brown Jr.,[4] Christopher S. Henry,[5] Janaka N. Edirisinghe,[5] and **Kirsten S. Hofmockel[3]**

**Institutions:** [1]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE; [2]Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, NE; [3]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA; [4]IT/Research Computing Directorate, Pacific Northwest National Laboratory, Richland, WA; [5]Division of Mathematics and Computer Science, Argonne National Laboratory

**Website URL:** https://www.pnnl.gov/projects/soil-microbiome/research

**Project Goals:** PNNL's Phenotypic Response of Soil Microbiomes SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions throughout the soil profile. Data are used to parametrize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in lab and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

**Abstract Text:** Metabolic network models of microorganisms help us to understand cellular metabolic capabilities, evolution, and ecological principles, as well as aid in the biotechnological design and management of microbial strains and consortia with desired functions. Because of the fundamental importance of metabolic network models in such a wide range of applications, the DOE Systems Biology Knowledgebase (KBase, http://kbase.us) provides a suite of apps and modules supporting the reconstruction, prediction, and design of metabolic models for microorganisms.

Construction of metabolic network models is facilitated by iterative implementation of three key steps: draft model building, gapfilling, and manual curation. Draft metabolic network models (i.e., initial models constructed from genomic or metagenomic data) typically contain incomplete biochemical pathways (i.e., have gaps) due to underlying knowledge gaps in gene function. Gapfilling – adding reactions to a metabolic model to reconcile with phenotypic data – is an essential step in model building because it augments the completeness and functionality of metabolic networks. Typical gapfilling algorithms (including the process currently implemented in KBase) correct one erroneous prediction at a time by iteratively adding new reactions to the

network. This approach, however, often leads to 'false positives' for other growth conditions (i.e., the model predicts growth, but experimental data show non-growth). Occurrence of false positive predictions is a greater problem in modeling communities, compared to isolates, due to the substantially larger pool of reactions available as options for gapfilling.

Based on the hypothesis that false positives are caused by identifying a minimum number of reactions to add to the network (parsimony) without accounting for their broader biological relevance, we propose a new advanced optimization algorithm (termed OMics-Enabled Global GApfilling or OMEGGA) that uses multi-omics data profiles, including amplicon, transcriptomic, proteomic, and intracellular metabolomic data, to simultaneously fit a draft model to all available phenotype data. This novel integration of amplicon, transcript, protein, and metabolite data into model refinement will yield more precise and predictive models and increase the accuracy of identification of active reactions. We will demonstrate the effectiveness of OMEGGA using condition-specific multi-omics and phenotype data from the Model Soil Consortia-2 (MSC-2) and associated isolated organisms developed through PNNL's Soil Microbiome SFA. These organisms were isolated from chitin enrichment cultures of a native soil microbiome. Data generated from various combinations of MSC-2 isolates is extremely valuable in testing the proposed algorithm under diverse contexts. We will accordingly develop generalized KBase apps using the KBase Software Development Kit (SDK) for implementation of all required gapfilling processes in OMEGGA.

Our new optimization algorithms will enable constructing high-quality metabolic networks that best match both molecular and phenotypic observations by avoiding time-consuming manual troubleshooting, which generally does not guarantee a successful outcome. Integration of omics data for gapfree model construction through simultaneous fit to multiple phenotype datasets is a novel idea that will fundamentally change the way we annotate genomes and build metabolic networks by allowing us to consider a more conservative threshold in predicting gene functions towards higher accuracy but lower coverage. This is because those gaps resulting from the conservative threshold can be filled in through the integration of confident experimental evidence (i.e., omics data) as proposed in this work. Therefore, our computational tools and KBase Apps will significantly improve the accuracy in metabolic network models of all complex biological systems including microorganisms, microbial communities, plants, and fungi.

**Funding Statement:** PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 70880 and FWP 78749.

**Title:** Functional and structural characterization of soil viral auxiliary metabolic genes

**Authors:** Ruonan Wu[1] (ruonan.wu@pnnl.gov), John R. Cort[1], Garry W. Buchko[1], Clyde A. Smith[2], Ian K. Blaby[3], Nikos C. Kyrpides[3], Yasuo Yoshikuni[3], Janet K. Jansson[1] and **Kirsten S. Hofmockel[1]**

**Institutions:** [1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99354, USA; [2]Stanford Synchrotron Radiation Lightsource, Stanford University, Menlo Park, CA 94025, USA; [3]Department of Energy, Joint Genome Institute, Berkeley, CA 94720, USA.

**Website URL:** https://www.pnnl.gov/projects/soil-microbiome/research

**Project Goals:** PNNL's Phenotypic Response of Soil Microbiomes SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions throughout the soil profile. Data are used to parameterize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in lab and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

**Abstract:** Recent research has revealed that viruses carry auxiliary metabolic genes (AMGs) that potentially contribute to soil metabolic processes while tuning the host machinery towards their own replication [1-3]. The majority of these AMGs are poorly characterized and only one, mannan endo-1,4-β-mannosidase, has been expressed and confirmed to be active in cleaving β-1,4-linked mannose, a plant-derived polysaccharide [1]. AMGs are commonly assigned potential functions based on their sequence similarity to annotated microbial genomic databases. These approaches overlook the critical motifs or catalytic sites that are key to determining if a protein is functional. Therefore, direct characterizations of AMG enzymatic functions and protein structures are of utmost importance for a mechanistic understanding of the ecological roles of soil viruses.

Here, our team focused on AMGs encoding chitosanases involved in the decomposition of chitin, the second most abundant structural polysaccharide after cellulose. We used a bioinformatics workflow that was previously developed [4] to identify a group of AMGs (glycosyl hydrolase family 75; GH75) from the largest global viral database (IMG/VR v3.0). The identified GH75 chitosanase-like AMGs were more prevalent in terrestrial environments, especially forest soils. Several chitosanase AMGs were synthesized and cloned at the Department of Energy (DOE)-Joint Genome Institute (JGI) into expression vectors. The cloned chitosanase AMGs were expressed and functionally characterized at the DOE-Environmental Molecular Sciences Laboratory

(EMSL). We were able to crystallize one of these proteins, verify its chitosanase activity and determine the first structure of any member of the GH75 chitosanase family using diffraction data collected at the Stanford Synchrotron Radiation Lightsource (SSRL). The resulting protein structure was determined at ultra-high resolution from crystals diffracting to better than 0.9 Å. The structure contained two domains: one domain containing a fold motif similar to other carbohydrate-hydrolyzing enzymes and the adjacent domain comprising a unique, previously uncharacterized fold. The active site, in a cleft between the two domains, was validated by the abolition of activity using single-site mutants of residues postulated to participate in the reaction mechanism. Further, a co-crystal structure of one of the site-directed mutants (E157Q) with chitohexaose revealed the substrate bound in the active site cleft. Because this viral chitosanase was found in forest soil and observed to be more active at an acidic pH, we speculate that viral AMG proteins may be under selection pressure to develop beneficial changes for adapting to the environment. AlphaFold2 modeling of the same GH75 chitosanase-like AMG generated a structure close to the experimental crystal structure (the root-mean-square deviation in Cα positions is 0.6 Å), demonstrating the potential of high-throughput predictions for selecting other chitosanase AMGs for functional validation. This study is the first thorough description of a soil viral AMG with detailed functional and structural characterizations. This new information will potentially enable us to better understand the metabolic contributions of soil viruses and their impact on microbiome.

**References/Publications**


1. Emerson, J.B., et al., *Host-linked soil viral ecology along a permafrost thaw gradient.* Nature microbiology, 2018. **3**(8): p. 870-880.
2. Trubl, G., et al., *Soil viruses are underexplored players in ecosystem carbon processing.* MSystems, 2018. **3**(5): p. e00076-18.
3. Wu, R., et al., *DNA Viral Diversity, Abundance, and Functional Potential Vary across Grassland Soils with a Range of Historical Moisture Regimes.* Mbio, 2021. **12**(6): p. e02595-21.
4. Wu, R., et al., *Moisture modulates soil reservoirs of active DNA and RNA viruses.* Communications biology, 2021. **4**(1): p. 1-11.



**Funding Statement:** PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 70880. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL.

**Synthetic Biology Driven Approach to Repurpose Polyamides**

Kate H. Kucharzyk[1*], Emma Beasley[1], Jacob Lilly[1], Edward Trigg[1], Marat Valiev[2], Margret [2], Jaydeep Bardhan[2], Yasuo Yoshikuni[3]

[1]Battelle Memorial Institute, Columbus OH
[2]William R. Wiley Environmental Molecular Sciences Laboratory, PNNL, Richland, WA
[3]DNA Synthesis Science Program Head, U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkley, CA

**Project Goals**: STORM aims to discover and design novel methods and enzymatic pathways for bioconversion and reuse of synthetic polymers. Nylon-6 (PA6) is a synthetic polymer highly recalcitrant to degradation, whose environmental and economic footprint could be reduced through polymer degradation and upcycling technologies. Currently, little understanding on PA6 biological degradation exists and no direct enzymatic degradation pathway has been identified. With an increase advancement in systems biology approaches, computational modeling, sequence-based homologue searching, and molecular docking will allow us to elucidate PA6 biological basis of degradation using several target enzymes classes. Here, we investigate a direct relationship between the PA6 material morphology and topology and efficiency of enzymatic degradation using several modeling approaches as well as experimental testing. The intercept between polymer morphology and enzyme kinetics will allow us to propose model pathway for degradation of PA6 previously unattainable by existing biological systems.

**Abstract:** STORM contains both polymer degradation and polymer morphology modeling approaches. For degradation, the research concept involves integration of computational modeling, elucidating of the biological catalyst for degradation via homologue searches and evolutionary analyses, experimental laboratory testing of enzyme efficacy, and finally, optimization of enzyme expression in model organisms. Enzymes of interest are fed into enzyme-substrate models to down-select potentially effective enzymes. Homology searches and alignments are then be used to identify target enzymes in currently available organisms. Laboratory testing are performed on the down-selected enzymes to determine degradation efficacy and parameters. These enzymes will then be engineered into a model organism for overexpression capable of degrading PA6 at higher rates. This work provides the first known technology for degradation and upcycling of recalcitrant PA6 polymers. Key risks include low yields of enzymes, cost of genetic engineering of organisms, and effective connection of the degradation portion of work. These identified risks are mitigated through proper enzyme down-selection prior to testing, experimental testing prior to overexpression, and the use of teaming partners at Debut Biotechnology and the U.S. Department of Energy's (DOE's) User Facilities, who provide technical expertise in these areas.

**Funding Statement**: Funding to this project is provided through the DOE Office of Science Biological and Environmental Research (BER), Award Number: DE-SC0022026.

**Title: Temperature sensitivity of soil bacterial networks from the Arctic to the Tropics**

**Authors:**
Bram W. G. Stone[1]* (bram.stone@pnnl.gov), Sheryl L. Bell[1], **Steven J. Blazewicz**[2], **Paul Dijkstra**[3,4], Michaela Hayer[3], Ben J. Koch[3], **Michelle C. Mack**[3,4], Jane C. Marks[3,4], Samantha N. Miller[3], **Ember M. Morrissey**[5], Juan Pineiro[5], Jeffrey S. Propster[3,4], Alicia M. Purcell[3], **Egbert Schwartz**[3,4], Chao Wang[5], **Kirsten S. Hofmockel**[1], **Jennifer Pett-Ridge**[2], **Bruce A. Hungate**.[3,4]

**Institutions:**
1. Earth and Biological Sciences Directorate, Pacific Northwest National Lab, Richland WA
2. Physical and Life Sciences Directorate, Lawrence Livermore National Lab, Livermore, CA
3. Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ
4. Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ
5. Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV

**Project Goals:** The work proposed here will integrate genomics- and isotope-enabled measurements of **G**rowth **R**ate, growth **E**fficiency, and the stoichiometry of **E**ssential **N**utrients during growth, an integration we call GREEN 'omics. Our **overarching objective** is to develop and apply 'omics approaches to investigate microbial community processes involved in nutrient cycling. The specific objectives of our proposed work are 1) to evaluate the microbial ecology of nutrient uptake, testing hypotheses about nutrient assimilation in response to temperature variation; 2) to evaluate the ecology of nutrient-use efficiency for soil microorganisms within a framework of ecological theory, and 3) to develop new isotope-enabled genomics and transcriptomics techniques that probe the microbial ecology of nutrient dissimilation. This work will push the frontier of isotope-enabled genomics by connecting quantitative stable-isotope probing to ecological theory about nutrient assimilation, nutrient-use efficiency, metabolic efficiency, and by applying these tools to understand the basic biology and ecology of soil microorganisms and how they transform nutrients in the environment.

**Abstract text:**
Temperature changes can have dramatic effects on the structure and function of soil microbial communities[1–4], though uncertainty exists as to the magnitude and predictability of carbon loss across different spatial and temporal extents[2,5]. Complex multi-species interactions which underpin the ecological functioning of a community can be summarized through network analysis[6,7]. We examined the soil bacteria from wide diversity of biomes representing Arctic, boreal, temperate, and tropical ecosystems in response to 5, 15, 25, and 35 °C temperatures. Network analyses typically infer interactions among taxa by detecting simultaneous changes in their relative abundances[8]. In ecological theory, interactions among taxa also alter rates of growth. Here, we constructed interaction networks using enrichment of $^{18}O$ in bacterial DNA – added via isotope-heavy water—as a measure of growth[9,10]. Previous analysis of these data indicated that the growth and respiration (i.e., the ecological functioning) of bacteria in these soils was most sensitive to changes from 5 to 15 °C[11]. By contrast, network size and composition was most severely altered at 35 °C in Arctic and boreal soils. At 35 °C, Arctic and boreal networks lost nearly all diversity except those bacteria that were present and active across all temperatures, forming a consistent, temperature-invariant core microbiome, and aligning with

patterns of lower sensitivity of growth and respiration at the community level at higher temperature[11]. Networks from temperate and tropical soils were composed of fewer temperature-sensitive bacteria. The temperature-invariant taxa remaining in Arctic and boreal networks at higher temperatures maintained less interactions per taxon leading to lower network density. With regard to the type of interactions, we saw an increase in the proportion of positive interactions from 5 to 15 °C, an indicator of stress and network instability via increased interdependency between bacterial taxa[12]. Community cohesion, a measure of total ecological connectivity[13], was highest at intermediate temperatures in all ecosystems and lowest at 5 and 15 °C. Taken together, these results help explain the sensitivity of different soils to warming conditions and demonstrate the strong links between ecological dynamics and soil carbon cycling.

**References/Publications:**


1. Allison, S. D., Wallenstein, M. D. & Bradford, M. A. Soil-carbon response to warming dependent on microbial physiology. *Nat. Geosci.* **3**, 336–340 (2010).
2. Crowther, T. W. *et al.* Quantifying global soil carbon losses in response to warming. *Nature* **540**, 104–108 (2016).
3. Ofiti, N. O. E. *et al.* Warming and elevated CO 2 promote rapid incorporation and degradation of plant-derived organic matter in an ombrotrophic peatland. *Glob. Chang. Biol.* 1–16 (2021) doi:10.1111/gcb.15955.
4. Walker, T. W. N. *et al.* Microbial temperature sensitivity and biomass change explain soil carbon loss with warming. *Nat. Clim. Chang.* **8**, 885–889 (2018).
5. Sulman, B. N. *et al.* Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry* **141**, 109–123 (2018).
6. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
7. Creamer, R. E. *et al.* Ecological network analysis reveals the inter-connection between soil biodiversity and ecosystem function as affected by land use across Europe. *Appl. Soil Ecol.* **97**, 112–124 (2016).
8. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).
9. Hungate, B. A. *et al.* Quantitative microbial ecology through stable isotope probing. *Appl. Environ. Microbiol.* **81**, 7570–7581 (2015).
10. Schwartz, E. Characterization of growing microorganisms in soil by stable isotope probing with H218O. *Appl. Environ. Microbiol.* **73**, 2541–2546 (2007).
11. Wang, C. *et al.* The temperature sensitivity of soil: microbial biodiversity, growth, and carbon mineralization. *ISME J.* (2021) doi:10.1038/s41396-021-00959-1.
12. Hernandez, D. J., David, A. S., Menges, E. S., Searcy, C. A. & Afkhami, M. E. Environmental stress destabilizes microbial networks. *ISME J.* **15**, 1722–1734 (2021).
13. Herren, C. M. & McMahon, K. D. Cohesion: A method for quantifying the connectivity of microbial communities. *ISME J.* **11**, 2426–2438 (2017).



**Funding statement:** *This research was supported by a grant from the Department of Energy's Biological Systems Science Division Program in Genomic Science (No. DE-SC0020172). Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.*

**Novel quantitative trait loci for leaf blight resistance in sorghum aids in understanding of *E. turcicum* pathosystem**

Sarah Lipps[1*] (slipps@illinois.edu), William Rooney[2], Santiago Mideros[1], **Tiffany Jamann[1]**

[1]University of Illinois, Urbana; [2]Texas A&M University, College Station

**Project Goals: Short statement of goals. (Limit to 1000 characters)**

Decreasing yield losses due to pathogens is crucial for any crop production. Sorghum leaf blight (SLB) and northern corn leaf, both caused by *Exserohilum turcicum*, are major diseases of sorghum and maize, respectively. In sorghum, yield losses as high as 50% have been reported. In contrast to maize, genetic architecture of resistance against *E. turcicum* in sorghum is poorly understood. By dissecting the genetic architecture of resistance to leaf blight in sorghum we can better understand resistance in both maize and sorghum. We had two overall objectives: i) exploit the genetic variation in two recombinant inbred line populations to identify quantitative trait loci associated with resistance and ii) identify shared resistance genes between sorghum and maize. We identified six loci, three in each population, significantly associated with resistance to SLB. The sorghum homologs of eight previously known maize resistance genes colocalized with the sorghum QTL. By understanding host resistance in the two species, we can optimize management strategies to create more durable resistance and protect bioenergy feedstock yields.

*This work is funded by DOE award number DE-SC0019189 (Plant Feedstocks Genomics).*

**Title:** Novel Systems Approach for Rational Engineering of Robust Microbial Metabolic Pathways

**Authors: Laura Jarboe**[1,*] (ljarboe@iastate.edu), Robert L. Jernigan[1], Peter C. St. John[2], Kejue Jia[1], Pranav M. Khade[1], Ambuj Kumar[1], Onyeka Onyenemezu[1], Jetendra K. Roy[1], Chao Wu[2]

**Institutions:** [1]Iowa State University, Ames, Iowa; [2]National Renewable Energy Laboratory, Golden, Colorado

**Project Goals:**

The goal of this project is to develop and implement a process for improving bioproduction under conditions that are appealing for industrial processes, such as high temperature and low pH. Our approach addresses the failure of metabolic reactions due to inhibition, denaturation, mis-folding or disorder of individual enzymes. We will develop and implement a framework that identifies these enzymes and then identifies their robust replacement enzymes from sets of extremophiles. The engineering strategy of replacing enzymes to improve bio-production is well-established, but rarely applied to system-wide stressors. We apply a systems genomics approach to improve bio-production, with *E. coli* as the model organism. Butanol production at high temperature and succinate production at low pH are the model systems. This approach is complementary to improvement of microbial robustness by engineering the cell membrane and has advantages relative to evolutionary-based organism improvement by prioritizing bio-production rather than growth.

**Abstract Text:**

**Aim 1: Metabolic Systems Modeling**. Flux balance analysis methods will be used to model changes in system temperature and intracellular pH through enzyme inhibition and changes to reaction thermodynamics. Models will be tuned with experimental data from Aim 3 and validated against existing literature data. An iterative process using these models, existing data and work from Aim 2 will be used to identify rate-limiting enzymes and evaluate the effects of their replacements.

**Aim 2: Computational Enzyme Assessment**. Enzymes from thermophilic organisms known to be robust under high temperature and low pH conditions will be queried as replacements for the rate-limiting enzymes identified in Aim 1. This approach utilizes the huge and rapidly growing body of knowledge regarding protein sequence, structure and evolution. Sequence matching will be used to assess protein vulnerability and to identify candidate replacement enzymes. Structures of sequence variants will also incorporate entropy evaluations to assess protein stabilities.

**Aim 3: Organism Characterization and Engineering**. Building on previously developed engineering strategies, the enzymes and replacements identified above will be recursively assessed for their impact on organism performance. Experimental data from Aim 3 will feed back to Aims

1 and 2. The final temperature-tolerant butanol producer and acid-tolerant succinate producer will both be characterized at the 0.5L-scale. Success in organism engineering will be judged in terms of the sensitivity of product titer, rate and yield to temperature or acidification.

The tasks span three length scales: enzyme sequence (Aim 2), the performance of metabolic pathways as enzyme networks (Aim 1), and the functional output of metabolic pathways in the form of organism activity (Aim 3). Our approach strikes out on a new path to build on existing knowledge, by complementing existing organism with enzyme replacement strategies, and at the same time is generic in terms of production organism and product identity. The ability to adjust metabolic models for changes in temperature and intracellular pH is also relevant to systems analysis of food spoilage organisms and pathogenesis. This framework could also be applied to other stressors that impact enzyme activity, such as salinity and concentration of alcohols or solvents. The proposed approaches for enzyme robustness and stability could likewise be applied for the selection of enzymes for pan-organism *in vitro* systems, such as in the production of bioprivileged molecules that can serve as precursors for drop-in petroleum replacement and novel molecules.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022090.
This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research under FWP ERW3526. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.

**Title:** The Scent of Senescence: Cell Wall Ester Modifications and Volatile Emission Signatures of Plant Responses to Abiotic Stress

**Authors:** *Rebecca A. Dewhirst[1], Joseph Lei[1], Cristina Castanha[1], Eliana Tucker[1], Yu Gao[1], Robert P Young[2], Miguel Portillo-Estrada[3], Luping Su[4], Silvano Fares[5], Jenny C. Mortimer[1,6], and* ***Kolby J. Jardine***[1*] *(kjjardine@lbl.gov)*

**Institutions:** *[1]Lawrence Berkeley National Lab, Berkeley, CA; [2]Pacific Northwest National Lab, Richland, WA*; *[3]University of Antwerp, Wilrijk, Belgium, [4]Tofwerk USA, Boulder, CO, [5]Institute of BioEconomy, National Research Council, Rome, Italy, and [6]University of Adelaide, Glen Osmond, SA, Australia*

**Website URL:** http://cellwallesters.pbworks.com/w/page/127623629/FrontPage

**Project Goals:** The Poplar Esterified Cell Wall Transformations and metabolic INtegration (PECTIN) project aims to study the metabolism of cell wall ester modifications and volatile intermediates, and their role in central physiological processes in the emerging biofuel species California poplar (*Populus trichocarpa*). A key goal of this research is to evaluate abiotic stress responses in plants with modified expression patterns of key genes involved in cell wall metabolism with altered amounts of methyl and acetyl groups present on cell walls. These genetic modifications will be evaluated for potential impacts on plant hydraulics, physiology, and stress responses. Understanding and manipulating the metabolism of cell wall modifications will not only provide important knowledge on the physiology and ecology of plants but will also allow the generation of engineered bioenergy crops such as poplar for sustainable production of biofuels and bioproducts, addressing BER's goal of developing renewable bioenergy resources.

**Abstract Text:** Growth suppression and defense signaling are simultaneous strategies that plants invoke to respond to abiotic stress. Here, we show that the drought stress response of poplar trees (*Populus trichocarpa*) is initiated by a suppression in cell wall derived methanol (meOH) emissions and activation of acetic acid (AA) fermentation defenses. Temperature sensitive emissions dominated by meOH (AA/meOH < 30%) were observed from physiologically active branches, detached stems, leaf cell wall isolations, and whole ecosystems. In contrast, drought treatment resulted in a suppression of meOH emissions and strong enhancement in AA emissions together with fermentation volatiles acetaldehyde, ethanol, and acetone. These drought-induced changes coincided with a reduction in stomatal conductance, photosynthesis, transpiration, and leaf water potential. The strong enhancement in AA/meOH emission ratios during drought (400-3,500%) was associated with an increase in acetate content of whole leaf cell walls, which became significantly $^{13}C_{1,2}$-labeled following the delivery of $^{13}C_{1,2}$-acetate via the transpiration stream. The results are consistent with central roles of acetate fermentation in regulating plant defense and metabolic responses to drought, and suggest that cell wall *O*-acetylation may be reversible

allowing plants to rapidly respond to drought stresses by down regulating methyl ester hydrolysis and growth processes while enhancing *O*-acetylation. We suggest that AA/meOH emission ratios could be used as a new highly sensitive non-destructive sensor to discriminate between thresholds of rapid plant growth and biomass accumulation and negative drought stress impacts on carbon metabolism, water use, and net primary productivity.

**References/Publications**


- Dewhirst R, Handakumbura P, Clendinen C, Arm E, Wang W, Washton N, Young R, Mortimer J, McDowell N, Jardine K (2021) Temperature acclimation of leaf gas exchange and photochemistry in *Populus trichocarpa*, ***ACS Earth and Space Chemistry*****,** 5, 8, 1813–1828.
- Dewhirst R, Lei J, Afseth C, Castanha C, Wistrom C, Mortimer J, Jardine K (2021) Are Methanol-Derived Foliar Methyl Acetate Emissions a Tracer of Acetate-Mediated Drought Survival in Plants? ***Plants***, 10, 411.
- Dewhirst R, Mortimer J, Jardine K (2020) Do cell wall esters facilitate forest response to climate? **Trends in Plant Science**. https:/doi.org/10.1016/j.tplants.2020.05.011
- Dewhirst R, Afseth C, Castanha C, Mortimer J, and Jardine K (2020) Cell wall *O*-acetyl and methyl esterification patterns of leaves reflected in atmospheric emission signatures of acetic acid and methanol, ***PLoS ONE*** 15(5), e0227591.



**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), Early Career Research Project (ECRP) grant no. FP00007421, the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by contract DE-AC02-05CH11231, and the Next-Generation Ecosystem Experiments–Tropics Project (NGEE-Tropics) under contract No. DE-AC02-05CH11231.

**Title: Mutations for Improved Enzyme Functionality at High Temperatures and Low pH**

**Authors:** Ambuj Kumar[1*] (ambuj@iastate.edu), Kejue Jia,[1] Pranav M. Khade,[1] Onyeka Onyenemezu,[1] Jetendra K. Roy,[1] Robert L. Jernigan,[1] **Laura Jarboe**[1]

**Institutions:** [1]Iowa State University

**Project Goals:** The initial goal is to computationally screen protein mutants to identify ones that will be efficient at high temperatures and low pH for greater fermentation efficiency. This will be followed by experimental selection for the best of these.

**Abstract Text:** Amino acid mutations can introduce functional changes ranging from loss of function to adaptation to extreme environmental conditions. Mutations have frequently been used to design thermostable proteins to retain function at high temperatures. Furthermore, predictions of damaging protein mutations have been widely applied in cancer research and drug discovery. Here we are designing a new machine learning classifier with the xgboost algorithm [1] to estimate mutation-induced changes in protein stability (both stabilizing and destabilizing) for a wide range of temperatures and pHs. A total of 4302 mutation induced protein stability changes were collected from Chen et al, 20019 [2], which was split into 80 percent for training and 20 percent for testing. Our results yield 91.1% accuracy and a 0.69 Mathews correlation coefficient, which is relatively higher than other sequence-based methods such as I-Mutant 2.0 [3], Mupro [4], iPTREE-STAB [5], and iStable 2.0 [2]. Since our model is based on information extracted from sequences alone, it can even be used on proteins where structure information is not available. We will continue developing this approach to include additional protein information. This initial approach informs us about how to take the next steps on our project "Novel Systems Approach for Rational Engineering of Robust Microbial Metabolic Pathways"

## References


1. T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 13-17-August-2016 (2016) 785–794. https://doi.org/10.1145/2939672.2939785.

2. C.W. Chen, M.H. Lin, C.C. Liao, H.P. Chang, Y.W. Chu, iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules, Comput. Struct. Biotechnol. J. 18 (2020) 622–630. https://doi.org/10.1016/j.csbj.2020.02.021.

3. E. Capriotti, P. Fariselli, R. Casadio, I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure, Nucleic Acids Res. 33 (2005). https://doi.org/10.1093/nar/gki375.

4. J. Cheng, A. Randall, P. Baldi, Prediction of protein stability changes for single-site mutations using support vector machines, Proteins Struct. Funct. Genet. 62 (2006) 1125–1132. https://doi.org/10.1002/prot.20810.

5. L.T. Huang, M.M. Gromiha, S.Y. Ho, iPTREE-STAB: Interpretable decision tree based method for predicting protein stability changes upon mutations, Bioinformatics. 23 (2007) 1292–1293. https://doi.org/10.1093/bioinformatics/btm100.



**Funding Statement:** This project is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), DE-SC0022090

**Phylogenomics of Solvent-Producing *Clostridium* Species To Enable Carbon-Negative Production of Acetone and Isopropanol**

Steven D. Brown,[1*] (steve.brown@lanzatech.com), Rasmus O. Jensen[1], Frederik Schulz[2], Fungmin (Eric) Liew,[1] Robert Nogle,[1] Tanus Abdalla,[1] Blake J. Rasor,[4] James Winkler[1], Shilpa Nagaraju,[1] Vinicio Reynoso,[1] Dawn M. Klingeman,[3] Zamin K. Yang,[3] Christina Canter,[1] Lan Wang,[1] Jonathan Strutz,[1] Payal Chirania,[3] Sashini De Tissera[1], Alexander P. Mueller[1], Zhenhua Ruan,[1] Allan Gao,[1] Loan Tran,[1] Nancy L. Engle,[3] Nicole Shapiro[2], Natalia Ivanova[2], T. B. K. Reddy[2], Daniel Udwary,[2] Sarah Moraïs,[5,6] Itzhak Mizrahia,[5,6] Sagar Utturkar,[7] Jason C. Bromley,[1] James Daniell,[1] Robert Conrado,[1] Timothy J. Tschaplinski,[3] Richard J. Giannone,[3] Robert L. Hettich,[3] Ashty S. Karim,[4] Edward A. Bayer,[5,6] David Paez-Espino,[2] Wayne P. Mitchell,[1] Tanja Woyke,[2] Nikos C. Kyrpides,[2] Nigel Mouncey,[2] Séan D. Simpson,[1] Ching Leang,[1] Michael Köpke[1] and **Michael C. Jewett[4]**

[1]LanzaTech, Skokie, IL; [2]DOE Joint Genome Institute, Berkeley, CA; [3]Oak Ridge National Laboratory (ORNL), Oak Ridge, TN; [4]Northwestern University, Evanston, IL; [5]Ben-Gurion University of the Negev, Beer-Sheva, Israel; [6]The Weizmann Institute of Science, Rehovot, Israel; and [7]Purdue University Center for Cancer Research, Purdue University, West Lafayette, IN.

https://www.energy.gov/science/ber/articles/cell-free-technology-accelerates-industrial-biotechnology

**Project Goals:** Non-model organisms have unique traits and offer significant advantages and benefits for biomanufacturing. One example is gas fermenting acetogens capable of converting low-cost waste feedstocks to fuels and chemicals, deployed today at commercial scale for conversion of steel mill emissions to ethanol. Yet, engineering these non-model organisms is challenging due to lower transformation and recombination efficiencies, longer cycle times and a more limited set of genetic tools compared to model organisms *E. coli* or yeast. Cell-free systems can guide and accelerate non-model organism strain development. We are establishing a new interdisciplinary venture, the clostridia Foundry for Biosystems Design (cBioFAB) that combines advancements in cell-free and *Clostridium* engineering metabolic engineering to develop industrial-robust production strains for conversion of lignocellulosic biomass to next-generation biofuels and bioproducts.

**Abstract:** We recently described optimized gas-fermenting *Clostridium autoethanogenum* strains for continuous at scale production of acetone or isopropanol from syngas at rates of up to ~3 g/L/h with ~90% selectivity. Life cycle analysis (LCA) confirmed the process is carbon negative and offers >160% greenhouse gas (GHG) savings over current production routes[1]. The multi-disciplinary approach involved screening a historical industrial strain collection for superior enzymes, omics analyses, kinetic modelling, cell-free prototyping to optimize flux, and fermentation scale-up to an industrial pilot. Gas fermentation using carbon-fixing microorganisms offers an economically viable and scalable solution with unique feedstock and product flexibility that has been commercialized recently[2].

The mixed acetone-butanol-ethanol (ABE) fermentation was one of the first industrial fermentation processes for chemicals production and its peak global production reached >500 metric tons per annum. The largest and most comprehensive strain collection of solvent-producing clostridia was assembled by Prof. David Jones (University of Otago, New Zealand), which form the basis of this genome resource project. Genome sequences for this collection were generated

and determined to facilitate biosystems design for biofuel and chemical production and provide insights into bacteriophage infection and adaptive immunity via CRISPR-Cas and other systems.

We describe the genomes for seven *C. acetobutylicum*, 194 *C. beijerenckii*, five *C. butyricum*, 57 *C. saccharobutylicum*, four *C. saccharoperbutylacetonicum* three *C. tetanomorphum* and a new candidate species, which were classified using genome-wide average nucleotide identities comparisons and phylogenetic tree analysis. For each species the number of new genomes has been increased significantly and this study increases number of core and accessory protein families, with ~19% more protein families overall and a concomitant increase in phylogenetic diversity across the genus. We describe 221 amino acid sequences for core acetone-butanol-ethanol (ABE) genes that were unique within the collection and new sequences not previously in public databases. We generated a combinatorial library of the acetone-forming genes thiolase, acetoacetate CoA-transferase and acetoacetate decarboxylase that we randomly screened for increased production in gas-fermenting *C. autoethanogenum* and were able to find a set that led to a 22-fold improved performance over designs with the reference genes. Along with acetone, we observed production of unwanted side products such as 3-hydroxybutyrate due to interaction of the heterologous acetone pathway and native *C. autoethanogenum* metabolism. We retrofitted a cell-free prototyping system we had previously developed (iPROBE)[3] to identify respective candidate genes which were subsequently deleted.

To facilitate future genome editing, we further mined the collection for new CRISPR systems. CRISPR arrays were unevenly distributed and were broadly classified into 2 types: a complete Type I-B CRISPR-Cas system, and a partial Type I-B CRISPR-Cas system, which lacks genes associated with CRISPR spacer integration. An analysis of CRISPR array sequences supports phages infecting industrial clostridia described here are mostly distinct from previously isolated *Clostridium* phages. Spacer hits indicate that CRISPR-based phage defense was mostly used against phages infecting other members of the same species, and rarely against phages infecting other related species of clostridia. Cellulosomal elements were retrieved only from *C.acetobutylicum* and *C. saccharoperbutylacetonicum*, as expected. The 271 high-quality genome sequences for solvent producing clostridia are a resource[4] that will enable synthetic biology and strain development for a range of production systems.

**References:**


1. Liew, F. *et al.* Carbon-negative, scaled-up production of acetone and isopropanol by gas fermentation. *Nat. Biotechnol.* **in press**, (2022).
2. Fackler, N. *et al.* Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. *Ann. Rev. Chem. Biomol. Eng.* **12**, 439–470 (2021).
3. Karim, A. S. et al. *In vitro* prototyping and rapid optimization of biosynthetic enzymes for cellular design. *Nat. Chem. Biol.* **16**, 912-919 (2020).
4, Brown, S. D., et al. Available via JGI DOI: 10.46936/10.25585/60000855 and at NCBI.



**Funding Statement:** We acknowledge the Biosystems Design program of the U.S. Department of Energy (DOE), Office of Science (SC), Office of Biological and Environmental Research (BER) under Award Number DE-SC0018249 and the BER BioEnergy Science Center (BESC). The work conducted by the DOE Joint Genome Institute is supported by SC under contract no. DE-AC02-05CH11231. The DOE Office of Energy Efficiency and Renewable Energy (EERE) supported Illumina data generation under Award Number DE-EE0007566. ORNL is managed by UT-Battelle, LLC, for DOE under contract DE-AC05- 00OR22725.

**Temperature drives substantial metabolic changes in the gas fermenting *Clostridium autoethanogenum* as revealed by multi-omics characterization**

Payal Chirania[1,2*] (pchirani@vols.utk.edu), Richard J. Giannone,[1] Nancy L. Engle,[1] David T. Reeves,[1,5] Hunter Zeleznik,[3] Vicki Liu,[3] Loan Tran,[3] Steven D. Brown,[3] Michael Köpke,[3] Timothy J. Tschaplinski,[1] Robert L. Hettich,[1] and **Michael C. Jewett**[4]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]UT-ORNL Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN; [3]LanzaTech Inc, Skokie, IL; [4]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL; [5]The Bredesen Center, University of Tennessee, Knoxville, TN.

https://www.energy.gov/science/ber/articles/cell-free-technology-accelerates-industrial-biotechnology

**Project Goals:** The interdisciplinary clostridia Foundry for Biosystems Design (cBioFAB) project addresses the complex challenge of designing, building, and optimizing biosynthetic pathways in biological systems by combining efforts from university, government, and industry partners. The goal of the project is to accelerate engineering efforts in non-model organisms through in vitro and in vivo metabolic pathway prototyping, computational modeling, and integrated omics analysis. Through these diverse approaches, the project seeks to provide the tools to enable high-level synthesis of next-generation biofuels and bioproducts from lignocellulosic biomass and expand the breadth of platform organisms that meet DOE bioenergy goals.
Non-model organisms have unique traits and offer significant advantages and benefits for biomanufacturing. One example is gas fermenting acetogens capable of converting low-cost waste feedstocks to fuels and chemicals, deployed today at commercial scale for conversion of steel mill emissions to ethanol. Yet, engineering these non-model organisms is challenging due to lower transformation and recombination efficiencies, longer cycle times and a more limited set of genetic tools compared to model organisms *E. coli* or yeast. Cell-free systems can guide and accelerate non-model organism strain development. This interdisciplinary venture, cBioFAB, combines advancements in cell-free and *Clostridium* engineering metabolic engineering to develop industrial-robust production strains for conversion of lignocellulosic biomass to next-generation biofuels and bioproducts.

**Abstract Text:** The fermentation of waste gases or syngas (a mixture of $H_2$/CO/$CO_2$) is a sustainable alternative for producing commodity chemicals and biofuels[1]. A small subset of acetogens, such as *Clostridium autoethanogenum*, can directly convert syngas into non-petroleum-based fuels, including ethanol, butanol, and other industrially-relevant chemicals. However, multiple factors such as pH, temperature, media composition, etc., affect end-product titers and process stability, and thus must be understood to better control product output. Studies on other acetogens have reported increases in ethanol production at temperatures lower than the organism's optimum[2]. However, few studies have examined the response of *C. autoethanogenum* outside its optimal growth temperature. Here, *C. autoethanogenum* cultures were grown at either 30°C or 40°C and characterized at a molecular level using a multi-omics approach (metabolomics, proteomics, and lipidomics). Notably, the product profiles varied between the two temperatures; at 40°C, the proportion of ethanol in liquid product was reduced to ~56% compared to ~73% at

30°C, with more carbon diverted towards the production of acetate (~27% compared to ~9% at 30°C) at the higher temperature. Initial omics analyses revealed that enzymes from both ethanol producing pathways (i.e., from acetate or acetyl-CoA) exhibited significantly reduced abundances at 40°C. In contrast, enzyme abundances for other products (acetate, lactate, and 2,3-butanediol) increased at the higher temperature. High temperature caused a dramatic metabolic shift in the microbe, whereby ~63% of the quantified proteins showed a significant difference in abundance. Among these, proteins from predicted bacterial microcompartment (BMC) gene clusters exhibited the highest fold changes with temperature. At 40°C, multiple proteins from the relatively smaller BMC cluster CAETHG_3273-3290 (18 genes) decreased, whereas those from the larger BMC cluster CAETHG_1810-1841 (32 genes) increased in abundance. Whereas both correspond to glycyl radical enzyme microcompartments (GRMs), limited information is available about the role of these temperature responsive BMCs. Interestingly, the larger cluster is a GRM1-type BMC which is predicted to metabolize choline into acetyl-phosphate or ethanol, which could further explain the increased acetate levels observed at 40°C. Although not grown on choline, the upregulation of these proteins could affect lipid metabolism and may explain the altered carbon partitioning leading to the large and diverse accumulation of alkyl-glycerol conjugates, including 1-myristoyl-glycerol, 1-palmitoleoyl-glycerol, and 1-myristoleoyl-glycerol, at 40°C as revealed by metabolomics. Overall, the acetogen undergoes significant metabolic changes to adapt to the temperature change, all of which provide key insights into its metabolism. Since this chassis organism can thrive at a range of temperatures (20°C - 44°C), the adjustment of fermentation operating parameters, combined with omics-guided metabolic engineering efforts, offer opportunity to direct the generation of a range of products.

**References**

1. Fackler, N.; Heijstra, B. D.; Rasor, B. J.; Brown, H.; Martin, J.; Ni, Z.; Shebek, K. M.; Rosin, R. R.; Simpson, S. D.; Tyo, K. E.; Giannone, R. J.; Hettich, R. L.; Tschaplinski, T. J.; Leang, C.; Brown, S. D.; Jewett, M. C.; Kopke, M., Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. *Annu Rev Chem Biomol Eng* **2021,** *12*, 439-470.
2. Kundiyana, D. K.; Wilkins, M. R.; Maddipati, P.; Huhnke, R. L., Effect of temperature, pH and buffer presence on ethanol production from synthesis gas by "*Clostridium ragsdalei*". *Bioresource Technol* **2011,** *102* (10), 5794-5799.

*This work is sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. Award number DE-SC0018249.*

**Rapid Prototyping for Development of A Novel Gas-to-1,3-Butanediol Bioprocess**

Fungmin (Eric) Liew,[1*] (eric.liew@lanzatech.com), Joseph Ni,[2] Bastian Vögeli,[2] Joanna Cogan,[1] June Gramling,[1] Rasmus O. Jensen,[1] Rebecca Ostien,[1] Monica MacDonald,[1] James Daniell,[1] Séan D. Simpson,[1] Steve D. Brown,[1] Linda J. Broadbelt,[2] Keith E. Tyo,[2] Michael Köpke[1] and **Michael C. Jewett[2]**

[1]LanzaTech, Skokie, IL; [2]Northwestern University, Evanston, IL

https://www.energy.gov/science/ber/articles/cell-free-technology-accelerates-industrial-biotechnology

**Project Goals:** Non-model organisms have unique traits and offer significant advantages and benefits for biomanufacturing. One example is gas fermenting acetogens capable of converting low-cost waste feedstocks to fuels and chemicals, deployed today at commercial scale for conversion of steel mill emissions to ethanol. Yet, engineering these non-model organisms is challenging due to lower transformation and recombination efficiencies, longer cycle times and a more limited set of genetic tools compared to model organisms *E. coli* or yeast. Cell-free systems can guide and accelerate non-model organism strain development. We are establishing a new interdisciplinary venture, the clostridia Foundry for Biosystems Design (cBioFAB) that combines advancements in cell-free and *Clostridium* engineering metabolic engineering to develop industrial-robust production strains for conversion of lignocellulosic biomass to next-generation biofuels and bioproducts.

**Abstract:** We recently described optimization and scale-up of gas-fermenting *Clostridium autoethanogenum* strains for continuous at scale production of acetone or isopropanol from syngas at rates of up to ~3 g/L/h and ~90% selectivity and >160% greenhouse gas (GHG) savings over current production routes[1]. This was achieved through an interdisciplinary approach of combinatorial pathway analysis, cell-free prototyping, multi-omics analysis, genome-scale kinetic modeling and process development as a blueprint for accelerated development of new bioproducts. Here we validate and refine this blueprint for production of 1,3-butanediol (1,3-BDO), a precursor for butadiene used in nylon and rubber production with a $20 billion USD/yr market.

While optimizing the acetone and isopropanol pathway, we also observed production of 3-hydroxybutyrate (3-HB) as by-product through native reductase and thioesterase activity interacting with the heterologous pathway [1]. The genome of *C. autoethanogenum* further encodes two aldehyde::ferredoxin oxidoreductase (AOR) enzymes[2] which could further reduce 3-HB to 3-hydroxybutanal, which subsequently could be reduced to 1,3-butanediol (1,3-BDO) via native alcohol dehydrogenase activity.

In a first step to build a route to 1,3-BDO, we aimed at enhancing 3-HB production by prototyping heterologous thiolase (ThlA) and acetoacetyl-CoA reductase enzymes (Hbd/PhaB) using a cell-free system (iPROBE). An optimal set of ThlA-Hbd were able to improve 3-HB *in vivo* production by 8-fold to 13 g/L from syngas and also allowed us to detect low levels of 1,3-

BDO of up to 0.5 g/L[3]. To further optimize 3-HB formation, we next aimed at replacing the native thioesterase reaction with promiscuous phosphate butyryltransferase (Ptb) and butyrate kinase (Buk) enzymes that would allow for ATP generation via substrate level phosphorylation. Genome scale modelling confirmed improved growth coupling and yields. We performed cluster analysis to identify potential Ptb and Buk variants from UniProt database. In total, 30 Ptb and 10 Buk variants were selected and synthesized into a modular cell-free to cell vector system we recently developed, which allows for cell-free prototyping and also serve as donor vectors for Golden Gate combinatorial assembly for *in vivo* workflow in *C. autoethanogenum*[4]. Cell-free prototyping in *Escherichia coli* revealed that Buk enzyme is not the bottleneck so we focused our effort on screening Ptb variants. Four Ptb variants were found to improve *in vitro* 3-HB biosynthesis by up to 6-fold. Based on this result, we performed combinatorial analysis using down-selected Ptb, Buk, ThlA and Hbd/PhaB in *C. autoethanogenum* to identify strains with optimal flux towards 3-HB and 1,3-BDO while growing on gas. Fermentation optimization in continuously stirred tank reactor was performed, demonstrating increased 1,3-BDO production titers using gas as feedstock.

**References**


1. Liew, F. *et al.* Carbon-negative, scaled-up production of acetone and isopropanol by gas fermentation. *Nat. Biotechnol.* **in press**, (2022).
2. Liew, F. *et al.* Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Met. Eng.* **40**, 104-114 (2017)
3. Karim, A. S. *et al. In vitro* prototyping and rapid optimization of biosynthetic enzymes for cellular design. *Nat. Chem. Biol.* **16**, 912-919 (2020).
4. Karim, A. S. *et al.* Modular cell-free expression plasmids to accelerate biological design in cells. *Synthetic Bio*. Ysaa019 (2020)



**Funding Statement:** We acknowledge the Biosystems Design program of the U.S. Department of Energy (DOE), Office of Science (SC), Office of Biological and Environmental Research (BER) under Award Number DE-SC0018249. DNA synthesis for the gene libraries was supported by the Joint Genome Institute (JGI) Community Science Program (CSP) under Award No. CSP-503280.

**Title:** Dynamic Kinetic Models Capture Cell-Free Metabolism for Improved Metabolic Engineering

**Authors:** Jacob Martin* (jacobmartin2022@u.northwestern.edu), Blake Rasor, Ashty Karim, Linda Broadbelt, **Keith Tyo**

**Institutions:** Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL

**Website URL:** N/A

**Project Goals:** We aim to develop a predictive model of metabolism in bacterial cell-free systems for the purpose of rapidly prototyping heterologous metabolic pathways. This model may be used to both optimize the production of metabolites in cell-free systems, as well as to understand how results in these systems should inform design in related cell-free metabolic pathways or even metabolism in living organisms.

**Abstract:** The optimization of biosynthetic production remains a challenge in metabolic engineering. This is particularly true for products made via longer heterologous pathways, which may require manual tuning of all component reactions. While cell-free systems rapidly increase the experimental throughput of testing pathway combinations, they remain complex systems and produce large amounts of difficult-to-interpret data. Toward this goal, we are developing a dynamic kinetic model to better understand this complex system and enable rapid data analysis for pathway optimization.

We are currently using this model to study butanol production via acetyl-CoA in *E. coli* cell-free systems. Because these experiments have exhibited complex dynamics, wherein the transient behavior of the heterologous butanol pathway interacts with core metabolism and vice versa, our model is both mechanistic and dynamic to robustly capture these experimental phenomena and predict optimal engineering solutions. However, compared to steady-state models, dynamic models have additional degrees of freedom that demand non-stationary flux measurements and different parameterization methods than those typically used for models of living cellular steady-state metabolism. To this end, we have developed a dynamic modeling framework which utilizes a variety of literature kinetic values, thermodynamic calculations, and Monte Carlo methods for parameter sampling. An ensemble of models is first pruned according to their fit to timecourse metabolomics, and the resulting top-performing models are fine-tuned by local parameter optimization algorithms. Each model also simulates several phenomena unique to cell-free systems, including gas-liquid equilibrium and transient pH measurements. By using this framework, we have successfully captured complex dynamic behavior, such as shifts in core metabolism that were experimentally observed when butanol production varied. Because these models are mechanistic in nature, detailed analysis was able to be performed to understand the metabolic causes of many of these dynamic behaviors. Lastly, the final ensemble of models was

used to provide experimental recommendations for metabolite and enzyme level changes to improve butanol production and has additionally identified a key bottleneck in the butanol pathway.

In future work, we plan to refine this model by retraining on data in which butanol pathway enzymes were individually adjusted, instead of simply knocked out. By preserving the parameters associated with core metabolism learned in the previous work and adjusting only the parameters of enzymes within the butanol pathway, we aim to learn detailed kinetics of the butanol pathway while minimizing model refitting or experimental measurements. Ultimately, we aim to translate these trained mechanistic parameters into models of *in vivo* production strains, which will accelerate model-building and product optimization workflows.

**Funding Statement:** This poster is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018249.

**Unearthing Enzyme Promiscuity with Cheminformatics to Design Biosynthetic Pathways Towards Novel Biomolecules**

Zhuofu Ni* (joseph.ni@u.northwestern.edu), Jonathan Strutz, Kevin Shebek, Linda Broadbelt, **Keith Tyo**

Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL

**Project Goals: Enzyme promiscuity, where an enzyme may catalyze a range of side reactions in addition to its main reaction, is a widely recognized yet largely unexplored phenomenon in biological systems. Cheminformatics workflows that can learn from native enzymatic reactions, enumerate novel reactions based on relevant enzymatic transformations, assess predicted biosynthetic pathways, as well as visualize predicted pathways in a user-friendly way, is crucial for pushing the boundaries of biomanufacturing. This could open up vast possibilities for bioproduction of valuable chemicals not natively produced by biology, where biosynthetic pathways can be constructed based on enzymes with desired promiscuous reactions. On the other hand, this also enables us to understand the implications of introducing heterologous enzymes on host organisms.**

We have developed the enzymatic reaction ruleset "JN1224min", an open-source retrobiosynthesis package Pickaxe v2.0, and the novel metabolite explorer Metabolic In-silico Network Expansions (MINEs) database v2.0. These are tools that incorporate enzyme promiscuity into the design of novel biosynthetic pathways. "JN1224min" is a minimal yet comprehensive set of 1224 enzymatic reaction rules that describes all common enzymatic transformations. These reaction rules, which specify reaction-center transformations of enzymatic reactions, can enumerate the largest possible number of reactions using the least number of rules. Pickaxe v2.0 utilizes these reaction rules to generate novel enzymatic reaction networks iteratively over many generations, in order to find promising pathways that lead from feedstocks to valuable products. We have updated Pickaxe to allow for on-the-fly filtering of enumerated reactions, based on criteria including chemical similarity, molecular weight, thermodynamics, as well as any custom filters. These implementations allow us to explore the largest possible enzymatic reaction space with improved efficiencies. MINE v2.0 is an update of the original MINE v1.0, a resource that

allows users to query novel metabolites that could be products of a promiscuous reactions from KEGG, E.coli, and yeast metabolites. MINEs are deployed as a graphical user interface, which allows users to easily propose candidate structures from untargeted metabolomics. Using our updated ruleset, MINE v2.0 features more than 15 times the number of potential metabolite structures compared to v1.0. Therefore, we have provided powerful tools that enable the metabolic engineering community to utilize enzyme promiscuity and improve pathway design towards biomanufacturing of a wider array of novel biomolecules.

*This poster is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018249.*

**Mechanistic Insights into Cell-free Gene Expression from an Integrated -Omics Analysis of Extract Preparation Methods**

Blake J. Rasor[1,2,3*] (b.rasor@u.northwestern.edu), Payal Chirania[5,7], Grant A. Rybnicky[2,3,4], Richard J. Giannone[5], Nancy L. Engle[6], Ashty S. Karim[1,2,3], Timothy J. Tschaplinski[6], Robert L. Hettich[5], and **Michael C. Jewett**[1,2,3,8,9]

[1]Department of Chemical and Biological Engineering, [2]Chemistry of Life Processes Institute, [3]Center for Synthetic Biology, and [4]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL; [5]Chemical Sciences Division and [6]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [7]Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN; [8]Robert H. Lurie Comprehensive Cancer Center and [9]Simpson Querrey Institute, Northwestern University, Chicago, IL

**Project Goals: We employ *E. coli* cell-free systems for enzyme synthesis and metabolic pathway prototyping to inform strain engineering in non-model organisms. Although cell-free gene expression has become a more common technique, relatively little is known about the shifts in metabolism and composition that occur when cells are lysed and processed to generate clarified cell extracts, which have different capacities for expression based on the specific steps employed during extract preparation. We sought to better characterize the behavior and composition of differentially processed extracts at the molecular level by performing kinetic analysis of transcription and translation in parallel with proteomic and metabolomic analysis.**

Cell-free systems originated as simplified platforms for studying biological processes, and they have developed into widespread biotechnology tools for gene expression and high-throughput prototyping. Recent efforts have optimized cell-free systems from numerous organisms for applications such as biosensing, protein production, and metabolite synthesis by refining cell lysate preparation methods and cell-free reaction composition. However, the applied nature of this optimization toward improved rates or titers often limits investigation into the physiological mechanisms behind changes in process and chemical composition in cell-free synthetic biology. In this work, we assessed changes in transcription and translation activity for *E. coli* cell extracts prepared with acetate or glutamate buffer and the common post-lysis processing steps of a runoff incubation and dialysis. We applied proteomic and metabolomic analysis to uncover potential mechanisms behind these changes in gene expression with the processing steps separately and in tandem, highlighting the impact of runoff incubation on the proteome and the role of buffer composition on central metabolism. Better understanding the shifts in activity and composition of cell extracts will inform future cell-free biology efforts with significant implications for gene expression and biochemical conversion in pathway prototyping and biomanufacturing.

*We acknowledge the Department of Energy grant DE-SC0018249 for funding of this project.*

**Title**: Sequence Entanglement with Post-Entanglement Modifications Enhances Functionality and Biosecurity of Entanglement Pairs

**Authors**: Jennifer L. Chlebek[1*] (chlebek1@llnl.gov), Dan Park,[1] and **Yongqin Jiao[1]**

**Institutions**: [1]Lawrence Livermore National Laboratory

https://sc-programs.llnl.gov/biological-and-environmental-research-at-llnl/secure-biosystems-design

**Project goals:** The overall goal of this project is to establish genetic sequence entanglement as a generalizable biocontainment strategy to improve genetic stability and to prevent horizontal gene transfer of genetically modified organisms. Here, we specifically sought to improve the function and robustness of an entanglement pair through post-entanglement modifications in an environmentally relevant organism.

**Abstract**:

For the past two decades, synthetic biologists have sought to genetically engineer microorganisms (GEMs) for a wide range of applications including therapeutic treatment and delivery, drug manufacturing, biofuel production, mineral extraction and waste degradation. Additionally, the microbiota that colonizes the rhizosphere of plant roots has been genetically engineered to enhance nutrient acquisition and drought resistance of agriculturally important crops. In order to ensure environmentally deployed GEMs do not proliferate uncontrollably, transfer their genetically modified genes horizontally to neighboring bacteria or cause unforeseen ecological consequences, biocontainment strategies must be implemented within these GEMs. For example, kill-switches are a common biocontainment strategy which control cell proliferation by using genetically engineered sense-and-respond modules to control the expression of a lethal actuator, such as a toxin. However, kill-switches and many other current biocontainment strategies (e.g., auxotrophies, codon recoding, etc.) are vulnerable to genetic mutations and horizontal gene transfer (HGT). We sought to use synthetic gene entanglement—a technique in which two genes are encoded within the same DNA sequence but translated from different reading frames—to increase the mutational robustness and prevent HGT of kill-switch systems in environmentally relevant plant symbionts, such as *Pseudomonas protegens*.

As an initial approach, we started with a previously developed [1], but poorly functional, entanglement pair comprised of a toxin (*relE*) embedded within a conditionally essential gene (*ilvA*). The gene *relE* is a small ~300bp sequence encoding a mRNA-degrading toxin and the gene *ilvA* is a larger ~1,000 bp sequence encoding the enzyme threonine deaminase which is required for isoleucine biosynthesis. When testing the function of this entangled pair in *P. protegens*, we found that *ilvA*/*relE* partially rescued the growth of an Δ*ilvA* mutant in minimal media but was not lethal to cells, which suggests that the threonine deaminase enzyme produced by the gene *ilvA* in the entanglement was partially functional, but the RelE toxin was not. In designing this entanglement pair, significant missense mutations were forced within the *ilvA* sequence in order to accommodate a WT amino acid sequence for RelE. Because of this, we hypothesized that the *relE* gene entangled within *ilvA* was poorly expressed due to lack of an

apparent upstream RBS within *ilvA*. To increase the expression of *relE*, we altered the ribosomal binding site (RBS) upstream of *relE* within the *ilvA*/*relE* entanglement. These enhanced *relE* RBSs improved the toxicity of entangled *relE* and surprisingly, did not reduce entangled *ilvA* function.

Utilizing the same *ilvA*/*relE* entanglement with the improved RBS, we sought to test whether the entangled *relE* toxin could prevent the transfer of the plasmid hosting the entanglement to other common soil Pseudomonads. Indeed, we found that the plasmid harboring *ilvA*/*relE* with an enhanced *relE* RBS dramatically reduced transformation efficiency among a variety of bacterial species compared to a plasmid hosting *ilvA/relE* lacking an improved RBS. These data suggest that synthetically entangling a toxin with a gene of interest can mitigate HGT and that genetic sequence entanglement can be used as a biocontainment strategy in an environmentally relevant microbe. Work is ongoing to determine if *ilvA* can provide long-term sequence fidelity to the embedded toxin, *relE*.

**References**

1. Blazejewski, T., Ho, H. I., & Wang, H. H. (2019). Synthetic sequence entanglement augments stability and containment of genetic information in cells. *Science*, *365*(6453), 595-598.

**Funding Statement:** This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Lawrence Livermore National Laboratory Secure Biosystems Design SFA “From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms”. Work at LLNL is performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

**Characterization of Toxins and Their Cognate Inactivators as Kill Switch Actuators in Plant-Beneficial *Pseudomonas fluorescens***

Halvorsen T.,[1]* (halvorsen1@llnl.gov), Chlebek J.,[1] Park D.,[1] Jiao Y.,[1] and **Yung M.C.[1]**

[1]Lawrence Livermore National Laboratory, Livermore, CA

*https://sc-programs.llnl.gov/biological-and-environmental-research-at-llnl/secure-biosystems-design*

**Project Goals: The LLNL Secure Biosystems Design Scientific Focus Area (SFA) aims to develop robust biosecurity tools at the sequence, cellular, and population levels to safeguard the deployment of genetically engineered, plant growth-promoting soil bacteria for environmental applications. In this portion of the project, we are focused specifically on the development of stable, effective kill switch circuits in plant-associated Pseudomonads.**

**Abstract:** Genetically engineered microorganisms (GEMs) are microbes programmed to treat disease, degrade pollutants, sustain agriculture, and produce a wide range of commodity chemicals. Often these engineered functions require release into the environment, where uninhibited GEM proliferation or horizontal transfer of recombinant DNAs to native organisms could result in unintended negative consequences. To prevent this, biocontainment strategies are needed that restrict GEM growth to defined conditions. One containment method uses protein toxins to kill the GEM via cytosolic expression from an engineered genetic circuit (i.e., a kill switch). While a vast number of new bactericidal protein toxins have been characterized in *E. coli*, with several being used to design kill switches, few toxins overall have been tested in industrially relevant microbes.

In this study, we investigate the functionality and robustness of a variety of kill switch toxins in the agriculturally relevant *Pseudomonas fluorescens* SBW25*,* a plant growth-promoting soil bacterium. Using a dual regulatory system to independently control the expression of each toxin and its cognate inactivator in SBW25, we characterized the cell-killing efficacy of 1) toxin-antitoxin modules (*ccdB-ccdA, parE-parD, relE-relB, hicA-hicB, ralR-ralA*), 2) a restriction endonuclease-methylase system (*ecoRIR-ecoRIM*), and 3) bactericidal toxin-immunity pairs normally deployed for direct killing of neighboring cells (*colicin E1-immunity E1, colicin E2-immunity E2, tse2-tsi2*).

While the majority of toxins we tested inhibit growth of SBW25, two are not effective in the dual circuit (RalR*,* Colicin E1). Interestingly, all functional toxins exert different levels of cellular burden when expressed under the same promoter; ParE and RelE exhibit the most burden and require expression of their cognate antitoxins to achieve wild type growth rates, while EcoRI, Colicin E2, and Tse2 exhibit the least burden. When we tested each kill switch under conditions that mimic the variable inducer levels a deployed GEM might face in the environment, we find that cell-killing by toxin systems with lower burden is less effective overall, which we show is

related to the relative stability and activity of the inactivator protein. In support of this, we find that although EcoRI is a toxic effector in *P. fluorescens,* its toxicity is entirely prevented in a dual circuit due to complete protection by basal expression of the cognate methylase. Collectively, these results demonstrate that inactivator proteins are powerful drivers of kill switch tolerance, a characteristic that can either be optimized for ON-OFF circuit switching behavior or exploited for biocontainment applications requiring irreversible protection. These results will help to guide the design of kill switches with robust cell-killing activity in uncontrolled environments, such as the plant rhizosphere.

**Funding:** Work at LLNL is performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-ABS-XXXXX). This program is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Lawrence Livermore National Laboratory SFA "From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms"

**Title:** Learning Protein Fitness Models from Evolutionary and Experimental Data

**Authors:** Chloe Hsu (chloehsu@berkeley.edu),[1]*, Hunter Nisonoff,[1] Clara Fannjiang,[1] and **Jennifer Listgarten**[1]

**Institutions:** [1]University of California, Berkeley

**Website URL:** https://sc-programs.llnl.gov/biological-and-environmental-research-at-llnl/secure-biosystems-design

**Project Goals:** Many protein design applications, including synthetic gene entanglement designs, rely on protein fitness models that predict protein function based on their amino acid sequence. This project aims to reduce the amount of data that the model requires to make reliable functional predictions for a protein by including sequences of evolutionarily related proteins as additional input.

**Abstract Text:**

There are several approaches to predict functional properties of a given protein from the protein's amino acid sequence. Existing machine learning-based models of protein fitness typically learn from either unlabeled, evolutionarily related sequences or variant sequences with experimentally measured labels. To reduce the amount of data that the model requires to make reliable functional predictions for a protein, recent work has suggested methods for combining both sources of information including evolutionary and experimental data.

Toward that goal, we propose a simple combination approach that is competitive with, and on average outperforms more sophisticated methods. Our approach uses ridge regression on site-specific amino acid features combined with a probability density feature from modeling the evolutionary data. Within this approach, we find that a variational autoencoder-based probability density model showed the best overall performance regardless which evolutionary density model was used. Moreover, our analysis highlights the importance of systematic evaluation and sufficient baseline. In addition to evolutionary and assay-labeled data, we also demonstrate that our combination approach can be extended to include protein structure information to further improve fitness prediction.

**References/Publications**

1. Hsu, C., Nisonoff, H., Fannjiang, C. and Listgarten, J.. " Learning protein fitness models from evolutionary and assay-labeled data." in press, Nature Biotechnology

**Funding Statement:**
This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Lawrence Livermore National Laboratory Secure Biosystems Design SFA "From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms". Support was also provided by the Chan Zuckerberg

Investigator program, C3.ai, the National Library of Medicine of the National Institutes of Health, and the National Science Foundation Graduate Research Fellowship Program.

**Title:** CAMEOX: Enhanced Computational Design of Overlapping Genes to Prolong Synthetic Device Function and Limit Horizontal Gene Transfer

**Authors:** Jose Manuel Martí[1]* (martimartine1@llnl.gov), Sean P. Leonard,[1] Christina Kang-Yun,[1] Jennifer L. Chlebek,[1] Dante P. Ricci,[1] Dan Park,[1] Charlotte Rochereau,[2] Tomasz Blazejewski,[2] Harris Wang,[2] Chloe Hsu,[3] Jennifer Listgarten,[3] **Yongqin Jiao[1]** and **Jonathan E. Allen[1]**

**Institutions:** [1]Lawrence Livermore National Laboratory; [2]Columbia University; [3]University of California, Berkeley

**Website URL:** https://sc-programs.llnl.gov/biological-and-environmental-research-at-llnl/secure-biosystems-design

**Project Goals:** The LLNL Secure Biosystems Design Scientific Focus Area (SFA) aims to develop robust biosecurity mechanisms at the sequence, cellular, and population levels to safeguard the deployment of genetically engineered, soil bacteria for environmental applications. In this portion of the project, we are focused on augmenting the algorithm for sequence entanglement design with an ML-driven design loop such as active learning feedback.

**Abstract Text:** When engineered bacteria with synthetic biology designs are used in natural environments, their engineered functions can be lost rapidly due to genetic instability or can be horizontally acquired by native bacteria. One way to prolong a desired function and to limit horizontal gene transfer (HGT) is to design genes with overlapped reading frames (i.e., "sequence entanglement"). A recently published software, CAMEOS (Constraining Adaptive Mutations using Engineered Overlapping Sequences) [1], can design and optimize overlapped genes. CAMEOS is a computationally intensive sequential code requiring several precomputed inputs. To automate the entire process and enhance the computational scalability, we developed CAMEOX (CAMEOs eXtended), a computational biology pipeline based on CAMEOS. CAMEOX improves CAMEOS by implementing multi-thread parallelism, broadening its application to any target host, and enabling multiple coordinated runs. In addition, the entire sequence entanglement design process has been automated, starting from the target gene sequence alignment used to estimate models of protein fitness along with visualization tools for sequence selection, all integrated in a HPC (High Performance Computing) pipeline. Since a critical step in protein fitness estimates requires generating a high-quality Multiple Sequence Alignment (MSA), depending on the number of homology sequences available for a specific target protein, the CAMEOX pipeline offers two different methods for building MSA. One method is based on UniProt's UniRef100 to allow for a larger set of input sequences, and the other based on OrthoDB for starting with a curated set of gene orthologs, which are presumably functionally closer to the target gene [2].

With the enhanced performance of CAMEOX, we performed a computational screen of 4,418 entangled gene pairs based on all conditionally essential genes present in *Escherichia coli* to assess the entanglement feasibilities. We selected the best performing gene pairs and then generated 20,000 putative entanglement constructs for each pair. Furthermore, CAMEOX was used to assess the entanglement feasibilities of other targets —specifically membrane proteins and toxins. Membrane proteins provide flexible regions that are amenable to gene entanglement,

including the transmembrane regions and exposed trans loops, where a gene of interest (GOI) and toxin can be entangled to promote desired function in our system while effectively preventing HGT. Moreover, some membrane proteins are essential for growth, and thus can be maintained in the engineered organisms over generations. Seven genes encoding for membrane proteins from our model plant-commensal bacterium, *Pseudomonas protegens*, and four type II toxins were selected to test their propensity for entanglement and potentially be developed into a platform to entangle various GOIs.

So far, we have focused our experimental efforts on a single overlapping pair containing an essential and conditionally essential gene: *infA* and *aroB*. The essential gene *infA* is a small ~200 bp sequence which encodes for the translation initiation factor 1 (IF-1) and the conditionally essential gene *aroB* is a larger ~1,000 bp sequence encoding for 3-dehydroquinate synthase which is involved in aromatic amino acid biosynthesis. Using CAMEOX, we have generated thousands of *infA*/*aroB* sequence entanglements solutions, and our near-term goal is to perform a high-throughput genetic screen to select for functional variants. The first step in the CAMEOX process involves generating an MSA based on *aroB* and *infA* orthologs pulled from public protein databases, OrthoDB in this case. When generating this MSA, however, we found that most of the orthologs had <60% sequence identity to the WT sequence. We hypothesized that these lower sequence identity orthologs would likely be non-functional and may not provide meaningful data to the MSA and hence to CAMEOX algorithm. To test the utility of the orthologs in the MSA with low sequence identity, we randomly sampled ~20 *infA* and *aroB* orthologs and tested their function *in vivo*. Our results showed that the majority of both *infA* and *aroB* orthologs tested were fully or partially functional in *P. protegens* and therefore that relative sequence identity does not correlate with ortholog functionality *in vivo*. These data suggest using lower sequence identity (<60%) *infA* and *aroB* orthologs in the MSA may provide useful information for generating functional entanglements. Ongoing high-throughput screen will determine if pseudolikelihood scores and other selection criteria provided by a trained MRF (Markov Random Field) fitness model, are better indicators of variant functionality.

**References**

1. Blazejewski, T., Ho, H. I., & Wang, H. H. (2019). Synthetic sequence entanglement augments stability and containment of genetic information in cells. *Science*, *365*(6453), 595-598.

2. Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids research, 47(D1), D807-D811.

**Funding Statement:** This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Lawrence Livermore National Laboratory Secure Biosystems Design SFA "From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms". Work at LLNL is performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-ABS-XXXXX).

**Title:** Redesigning the *Escherichia coli* genome with a 19-Amino Acid Alphabet

**Authors:** Guillaume Urtecho[1*] (gu2144@cumc.columbia.edu), Liyuan Liu[1], Yiming Huang[1], Charlotte Rochereau[1], Avi Swartz[2], Tomasz Blazejewski[1], Sergey Ovchinnikov[2], Calin Plesa[3], and **Harris Wang**[1]

**Institutions:** [1]Department of Systems Biology, Columbia University, New York, NY; [2]Knight Campus, University of Oregon, Eugene, Oregon; [3]John Harvard Distinguished Science Fellowship Program, Harvard University, Cambridge, MA

**Website URL:** https://sc-programs.llnl.gov/biological-and-environmental-research-at-llnl/secure-biosystems-design

**Project Goals:**

**AIM 1. Computational redesign of all *E. coli* proteins with a 19 amino acid alphabet (Ec19).** The overarching goal of this aim is to develop the necessary computational models and designs leading to the generation of functional proteins utilizing a reduced amino acid alphabet (Ec19).

**AIM 2. Systematic high-throughput (HT) experimental testing of 19-AA gene designs for all essential and highly expressed *E. coli* genes.** The goal of this aim is to develop and implement a HT platform to experimentally test the function of individual gene designs (from Aim 1) directly in cells.

**Abstract Text:**

The amino acid alphabet of life is universally conserved from bacteria to eukaryotes. As such, all living organisms on Earth require at least 20 amino acids (AAs) to grow and reproduce. This project seeks to answer the following central question: Are all 20 canonical amino acids (AAs) essential for life, or can life be built with fewer than 20 amino acids? Computational analysis of all *E. coli* genes and their orthologs shows that Ile (I), and Val (V) have the lowest frequencies of strong evolutionary conservation. Furthermore, these amino acids exhibit highly similar biochemical properties and are rarely found in active sites of enzymes. Altogether, these observations suggest that Ile or Val may be suitable residues for global substitution. Here, we explore the hypothesis that Ile or Val is dispensable to biological life. We have developed five protein design strategies, ranging in degree of sophistication, to redesign 396 essential or highly-expressed *E. coli* genes where one or both of these amino acids have been globally replaced. These approaches yield 1606 redesigned variants, which we have synthesized using DropSynth and uniquely barcoded for downstream characterization in pooled formats. Lastly, we perform the direct replacement of native *E. coli* genes with their cognate variants using a lambda recombinase-based approach and characterize the viability of redesigned genes in multiplexed growth assays. We anticipate these experiments will generate a rich set of data that can be used to improve and refine protein models developed in Aim 1 and to provide a platform for iterative design improvements for all genes in *E. coli*.

**Funding Statement:**
This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Lawrence Livermore National Laboratory Secure Biosystems Design SFA “From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms”, grant no. 20-1777; the National Science Foundation, Grant no. MCB-2032259; and the Howard Hughes Medical Institute Hanna H. Gray Fellows Program, grant no. GT15182.

**Title:** Programmed Lysis of Cells in Response to Electrogenetic Inputs

**Authors:** Eric VanArsdale[1,2,3*] (esv5009@terpmail.umd.edu), Monica Chu[1,2,3], Tiffany Halvorsen[4], Ali Navid[4], Gregory F. Payne[2,3], Yongqin Jiao[4], Mimi Cho Yung[4], **William E. Bentley[1,2,3]**

**Institutions:** [1]Fischell Department of Bioengineering, University of Maryland, College Park, MD 20742; [2]Institute of Bioscience and Biotechnology Research, University of Maryland, College Park, MD 20742; [3]Fischell Institute of Biomedical Devices, University of Maryland, College Park, MD 20742; [4]Systems and Synthetic Biology Group, Lawrence Livermore National Laboratories, Livermore, CA 94550

**Website URL:** https://genomicscience.energy.gov/research/sfas/llnlseqcellpop.shtml

**Project Goals:** The LLNL Secure Biosystems Design Scientific Focus Area (SFA) aims to develop robust biocontainment mechanisms at the sequence, cellular, and population levels to safeguard the deployment of genetically engineered, soil bacteria. In this portion of the project, we are focused on developing robust and generalizable cellular and population level containment mechanisms for improving the safer use of plant benefiting microbes in the rhizosphere.

**Abstract Text:** Cell death and lysis play an important role in a wide array of biological niches. By genetically "programing" microbial cell lysis, we might provide a new methodology for ensuring biocontainment of cells within natural ecosystems as well as for engineered microbial consortia. Understanding how lysis and its control give rise to observed population dynamics is challenging as few methods exist that enable precise measurement and control. To this end, we have developed tools that link an externally applied electronic potential to initiate, measure, and even control genetic circuits that biologically mediate cell lysis. We first explore direct electronic activation of cell lysis within a clonal population of cells through OxyR-mediated expression of lysis protein E (here abbreviated LysisE), a lytic protein from φX174 phage that disrupts peptidoglycan synthesis. This genetic circuit can plug directly into electrogenetic control through detection of hydrogen peroxide produced at a gold electrode surface. Next, we show that by transforming electronically produced hydrogen peroxide into native biological signals that induce quorum sensing, and, in turn, LysisE expression in a separate population, we can dramatically enhance cell lysis overall, preventing significant regrowth or remodeling of the overall population for at least 6 hours. Lastly, we show that by combining the hydrogen peroxide induced synthesis of quorum sensing signals and the quorum sensing induced expression of LysisE in the same population, we can create a self-lysing transmitter of lytic signals that can simultaneously lyse neighboring bacteria. Each of these methods should enable deployment in agricultural (i.e., rhizosphere) or industrial process settings.

**Funding Statement:** This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Lawrence Livermore National Laboratory Secure Biosystems Design SFA "From Sequence to Cell to Population: Secure and

Robust Biosystems Design for Environmental Microorganisms", as well as the NSF (ECCS 1807604, CBET 1805274), DTRA (HDTRA1-19-0021) and the Semiconductor Research Corporation (SRC 2483.001) through the University of Maryland.

# Transforming our understanding of chloroplast-associated genes through comprehensive characterization of protein localizations and protein-protein interactions

Lianyong Wang*[1](lianyong@princeton.edu), Weronika Patena[1], Kelly A. Van Baalen[1], Yihua Xie[1], Sophia Gavrilenko[1], Michelle Warren-Williams[1], Henry Harrigan[1], Emily R. Singer[1], Vinh Ton[1], Alexandra Wilson[1], Linqu Han[2], Bibin Paulose[2], Christoph Benning[2], **Danny J. Schnell[2]**, and **Martin C. Jonikas[1]**

[1]Princeton University, Princeton, NJ; [2]Michigan State University, East Lansing, MI

**Project Goals: Our project aims to generate a map of protein localizations and protein-protein interactions for 5,906 genes associated with the chloroplast. We use two synergistic organisms, the unicellular model green alga *Chlamydomonas reinhardtii* and the dedicated biofuels oilseed crop *Camelina sativa*. Objectives 1 and 2 are to generate a searchable online resource of protein localizations and protein- protein interactions for nearly all chloroplast-associated proteins. We seek to achieve these objectives by leveraging high-throughput protein tagging, microscopy and affinity purification-mass spectrometry in *Chlamydomonas*. Objective 3 is to illustrate the value of this resource to biofuel crops by validating high-priority localizations and protein-protein interactions in *Camelina* and by building on the newly generated knowledge to advance our understanding of protein interaction networks that impact yield and stress resistance.**

Our efforts are focused around the chloroplast because of the organelle's central roles in photosynthesis, metabolism and intracellular signaling, all of which are targets of ongoing biofuels crop engineering efforts. Furthermore, chloroplast-associated genes are particularly underrepresented in existing systems-level datasets because most high-throughput studies to date were performed in model systems that lack chloroplasts. As demonstrated in yeast, protein localization and protein-protein interaction data transform our understanding of the genes under study by immediately generating specific hypotheses about the mechanism of action of their protein products.

So far, we have determined the localization of 1,048 proteins. Of these proteins, 585 localized to the chloroplast. Intriguingly, 305 of the proteins that localized to the chloroplast were also observed in other subcellular compartments, suggesting proteins with possible signaling roles, dual functions, and possible alternative targeting routes. We have collected protein-protein interactions of 960 tagged proteins and we are currently identifying protein-protein interactions of the rest and the second biological replicate by affinity purification-mass spectrometry pipeline. We anticipate that the localization and protein-protein interaction data will provide key information on the functions of thousands of uncharacterized proteins, many of which have no

recognizable protein motifs. The project will also have a long-term impact as the scientific community utilizes the resource of strains, constructs, and data.

*This research is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, grant no. DE-FOA-0002060.*

**Title:** Modeling Environmental Influences on Biomass Composition of Diverse Switchgrass Genotypes

**Authors:** Demeke Mewa Bayable,[1]* (demeke.bayable@wsu.edu), Mahbobeh Lesani,[2] Jason Bonnette,[3] Edward Wolfrum,[4] Robert W. Heckman,[3] Li Zhang,[3] John Lovell,[5] **Thomas Juenger**,[3] Laura E. Bartley[1]* (laura.bartley@wsu.edu)

*presenting authors

**Institutions:** [1]Institute of Biological Chemistry, Washington State University, Pullman, WA; [2] Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK; [3] Department of Integrative Biology, University of Texas, Austin, TX; National Renewable Energy Lab, Golden, CO; [5]Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Project Goals:** This project aims to understand the environmental and genetic influences on switchgrass composition towards increasing sustainability of switchgrass production for biorefining by developing generalist and specialist plant ideotypes that maximize biomass yield and composition, stress tolerance, and carbon sequestration capacity.

**Abstract Text:** Biomass composition influences biorefining yield and conversion efficiency. Towards optimizing composition and related physiological traits of switchgrass, this project seeks to understand the genetic and environmental determinants of switchgrass biomass composition. Switchgrass consists of multiple ecoptypes, akin to races, that possess distinctive phenotypes. Compared to the northern upland ecotype, the southern lowland ecotype is characterized by greater biomass production but reduced abiotic stress tolerance, and in some studies, compositional differences, including lower lignin, ash, and nitrogen content. Here we report results of ten near-infrared reflectance spectroscopy-predicted cell wall composition traits, including in vitro dry matter digestibility (IVDMD), lignin, glucan and xylan, for $F_2$ plants of a cross of lowland and upland genotypes grown in common gardens at 10 sites across central North America. For biomass harvested at the end of the season, all trait means varied significantly among genotypes and locations, though heritable variation was only 20-40% on average. Still, we found a total of 68 quantitative trait loci (QTL) across all traits, of which 47 QTL (70%) showed gene-by-environment (GxE) interactions, meaning that the magnitude of the QTL effects varied with location. Several QTL were pleiotropic. Consistent with single loci controlling multiple traits, we found many strong negative and positive genetic correlations. In contrast to the phenotypic correlations, which except for those with xylan were fairly strong and consistent across sites, genetic correlations often varied, even showing different signs at different sites. For example, glucan was positively genetically correlated with IVDMD at most sites, but negatively correlated at two southern sites. This is consistent with the strong environmental influence on composition. To better understand the components of the environment that influence composition, we are

conducting a meta-regression analysis of weather on trait variation across sites using principal components (PCs) of seven weather statistics (e.g., coldest, warmest, driest, and wettest 14-day spans) in each of three intervals of the growing season. Across all modeled genotype groups, variation in xylan was highly influenced by weather PC1; but for most traits, models of the influence of weather on composition greatly varied with genotype. In one $F_0$ lowland genotype, WBC, 48% of the total weather variation (PC1 and PC4) explained lignin, which was mutually exclusive with the 25% of the variation (PC2 and PC6) that explained glucan. On the other hand, for the $F_0$ upland genotype VS16, lignin was not significantly influenced by PC1 (38% weather variation). WBC lignin was highly positively related to the mean daily minimum temperature and negatively related to the driest span; whereas, glucan was highly positively related to precipitation in the wettest spans. Understanding the interaction of climatic drivers with genetic covariance and QTL that are antagonistic across sites may help to minimize composition tradeoffs. Thus, this study provides insights and expands the foundation for optimal manipulation of cell wall chemistry to improve biorefining.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021126. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract No DE-AC02-05CH1123

**Title:** Exploring Switchgrass Genetic Diversity with Multiple Reference Genomes

**Authors: Thomas E. Juenger**,[1] John T. Lovell[2,3]* (jlovell@hudsonalpha.org), Jane Grimwood,[3] and Jeremy Schmutz[2,3]

**Institutions:** [1]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA; [2]Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; [3]Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Website URL:** https://sites.cns.utexas.edu/juenger_lab/switchgrass

**Project Goals:** Overall, we are striving to improve bioenergy feedstock production by understanding the genetic basis of plant-environment interactions. This goal includes testing for climate adaptation, modeling beneficial and stressful biotic interactions, and exploring the mechanisms of abiotic stress responses. During our earlier work (e.g. [1-2]), we discovered a massive amount of physiological and molecular variation in switchgrass. While this diversity is the raw material that allows breeders to improve feedstock production, making use of this variation is very challenging — the immense DNA differences between some switchgrass genotypes means that traditional methods to explore genetic diversity simply do not work. Under the work presented here, we developed and employed multiple genome resources that span this diversity to provide the foundation for molecular characterization of switchgrass biomass production, stress responses and biotic interactions.

**Abstract Text:** A single haploid reference genome gives breeders the resources to connect alleles to traits; a significant step towards accelerating crop improvement. However, breeding programs often leverage highly diverged germplasm, which contain large-scale variants that are not readily identified by a single reference genome. For example, in switchgrass, the fast-growing southern lowland AP13 genotype (which serves as the reference genome [1]) is ~1 million years diverged from the cold-tolerant northern upland gene pool. To assist breeding and gene discovery efforts, we have developed four total reference genomes, two northern uplands and two southern lowlands. Each chromosome-scale genome is fully outbred representing both meiotically homologous chromosomes. Here, we present these genomes and a detailed comparison of gene presence absence, structural and copy number variation. These variants can serve as *a priori* targets for ongoing molecular breeding efforts to make switchgrass a more economically and ecologically sustainable biofuel feedstock.

**References/Publications**

1. Lovell, John T., Alice H. MacQueen, Sujan Mamidi, Jason Bonnette, Jerry Jenkins, Joseph D. Napier, Avinash Sreedasyam, et al. 2021. “Genomic Mechanisms of Climate Adaptation in Polyploid Bioenergy Switchgrass.” *Nature* 590 (7846): 438–44.
2. Lowry, David B., John T. Lovell, Li Zhang, Jason Bonnette, Philip A. Fay, Robert B. Mitchell, John Lloyd-Reilley, et al. 2019. “QTL × Environment Interactions Underlie

Adaptive Divergence in Switchgrass across a Large Latitudinal Gradient." *Proceedings of the National Academy of Sciences of the United States of America* 116 (26): 12933–41.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021126. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract No DE-AC02-05CH1123

**Title:** A Generalist-Specialist Tradeoff between Switchgrass Cytotypes Impacts Climate Adaptation and Geographic Range

**Authors:** Joseph D. Napier[1*+](Joseph.Napier@austin.utexas.edu)**,** Paul P. Grabowski[2+], John T. Lovell[2], Jason Bonnette[1], Sujan Mamidi[2], Maria Jose Gomez-Hughes[1], Acer VanWallendael[3], Xiaoyu Weng[1], Lori H. Handley[2], Min K. Kim[2], Arvid R. Boe[4], Philip A. Fay[5], Felix B. Fritschi[6], Julie D. Jastrow[7], John Lloyd-Reilley[8], David B. Lowry[3], Roser Matamala[7], Robert B. Mitchell[9], Francis. M Rouquette[10], Yanqi Wu[11], Jenell Webber[2], Teresa Jones[2], Kerrie Barry[12], Jane Grimwood[2], Jeremy Schmutz[2], & **Thomas E. Juenger**[1]

+Co-first authors

**Institutions:** [1]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA; [2]Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; [3]Department of Plant Biology, Michigan State University, East Lansing, MI, USA; [4]Department of Agronomy, Horticulture and Plant Science, South Dakota State University, Brookings, SD, USA; [5]Grassland, Soil and Water Research Laboratory, USDA-ARS, Temple, TX, USA; [6]Division of Plant Sciences, University of Missouri, Columbia, MO, USA; [7]Environmental Science Division, Argonne National Laboratory, Lemont, IL, USA; [8]Kika de la Garza Plant Materials Center, USDA-NRCS, Kingsville, TX, USA; [9]Wheat, Sorghum, and Forage Research Unit, USDA-ARS, Lincoln, NE, USA; [10]Texas A&M AgriLife Research and Extension Center, Texas A&M University, Overton, TX, USA; [11]Department of Plant and Soil Sciences, Oklahoma State University, Stillwater, OK, USA; and [12]Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

**Project Goals:**

- **Explore the natural genetic and phenotypic variation found in octoploids sampled from across the geographic range of switchgrass**
- **Quantify differences in genetic diversity, habitat preference, adaptability, and fitness responses across climatic gradients between tetraploid and octoploid switchgrass**
- **Identify potentially novel combinations of genetic diversity present in octoploids that might be linked to climate adaptation and range expansion, which would subsequently represent a valuable breeding resource for enhancing the resilience and sustainability of switchgrass feedstock production**

**Abstract Text:**

Polyploidy is the result of whole-genome duplication (WGD) and can have pronounced ecological and evolutionary implications. Approximately half of all angiosperms are polyploids, and polyploidy is thought to be a major component of adaptation and speciation within this group. Quantifying how transitions to higher ploidy can generate shifts in fitness and adaptability

is difficult when comparing across taxa because the role of ploidy is hard to disentangle from the roles of other genetic changes that accumulated after taxa diverged. However, different ploidy levels, or cytotypes, can exist within a single species, which provides a natural experiment to test how ploidy variation alters genomic diversity, fitness, and adaptability.

Switchgrass, *Panicum virgatum*, is a widespread, perennial C4 grass in North America with multiple naturally occurring cytotypes, primarily tetraploids (4X) and octoploids (8X). The genetic and morphological variation present in 4X switchgrass has been closely studied (i.e. Lovell et al. 2021), but there has been considerably less focus on the 8X and little is known about how this shift to higher order ploidy has altered fitness, genetic composition, or niche breadth. Here we contrast the molecular and quantitative genetic diversity of 4X and 8X switchgrass across naturally-occurring genotypes and 10 common gardens to evaluate the basis of ploidy-associated shifts in admixture, adaptive potential, ecological niche, and fitness. Specifically, we discovered (1) recurrent and evolutionary distinct genesis of 8X populations containing novel combinations of genetic diversity, (2) similar morphological/ecotypic divisions within 4X and 8X cytotypes but divergent cytotype fitness clines indicating a generalist-specialist tradeoff, and (3) niche evolution between 4X and 8X linked to climate adaptation. Combined, our results indicate that mixed-ploidy systems can be used as valuable tools to bolster the resilience of natural and agronomic systems by providing insight on how ploidy variation enables niche divergence, fitness tradeoffs, and range expansion.

**References/Publications**

1. Lovell, John T., et al. "Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass." *Nature* 590.7846 (2021): 438-444.

**Funding Statement:** *This research was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0014156 and DE-SC0021126 to T.E.J.*

**Title:** Enrichment of Soil Bacterial Taxa in Switchgrass (*Panicum virgatum*) Cropping Systems Across a Latitudinal Gradient

**Authors:** Michael P. Ricketts [1]* (mricketts@anl.gov), Joseph A. Edwards,[2] Julie D. Jastrow,[1] Roser Matamala,[1] Felix B. Fritschi,[3] Philip Fay,[4] and **Thomas E. Juenger [2]**

**Institutions:** [1]Environmental Science Division, Argonne National Laboratory, Lemont, IL; [2]University of Texas at Austin, Austin, TX; [3]Division of Plant Sciences, University of Missouri-Columbia, Columbia, MO; [4]United States Department of Agriculture – Agricultural Research Service, Temple, TX

**Project Goals:** The primary objective of this project is to gain a better understanding of the genetic drivers and mechanisms contributing to variation in switchgrass (*Panicum virgatum*) production and fitness across a broad climatic gradient. We have established experiments at multiple sites across the United States to evaluate switchgrass as a potential source of sustainable bioenergy feedstock and study its effects on ecosystem processes. Specifically, we aim to identify beneficial traits, genes, and microbial associates that may provide locally adaptive benefits to switchgrass and improve establishment, growth, and resiliency. Here we focus on soil bacterial community dynamics in switchgrass cultivars during a 3-year period following initial plant establishment.

**Abstract text:** Plants have evolved alongside soil microorganisms over hundreds of millions of years, resulting in the establishment of many mutualistic relationships. Soil microbial associates can provide protection from pathogenic microbes and regulate nutrient accessibility through organic matter decomposition and nutrient transformation processes. Additionally, variations in root morphology and/or root metabolite exudation may facilitate the active recruitment and enrichment of beneficial microbes, leading to the development of unique consortia of microbial associates, which may vary both between- and within-species depending on the local soil environment. Understanding the relationship between switchgrass and soil microbes, and how it influences the surrounding soil environment, may contribute to establishing switchgrass as an economically viable and sustainable bioenergy crop.

To study these relationships, we collected samples from a large, multi-institutional collaborative project where thirty plots of monoculture switchgrass stands were established at each of three sites across a latitudinal gradient (Fermilab, IL; Columbia, MO; and Temple, TX). Six cultivars were selected to represent a broad range of genetic diversity within switchgrass and five replicate 6m x 6m plots were arranged in a randomized complete block design. At each site, six soil cores were collected using a stratified random design in spring 2016 to characterize the soil environment prior to switchgrass planting. In 2018 and 2019, soil microbial communities in the root zone were targeted by collecting soil cores adjacent to switchgrass crowns at multiple time points throughout the growing season. Samples were homogenized and roots removed prior to subsampling for DNA extraction. Amplicon sequencing of the 16S rRNA gene was performed at DOE's Joint Genome Institute (JGI) to determine the bacterial community structure within each sample.

Principal coordinates analyses suggest the primary drivers differentiating soil bacterial community structure in the root zone were site and collection date (differentiated by year in MO and TX, and by early *vs*. late season in IL). Although soil bacterial community structure in the root zone did not vary distinctly among switchgrass cultivars, we did find evidence that suggests switchgrass may recruit and enrich certain bacterial taxa. Using differential analysis (DESeq2), we identified 2,638 bacterial amplicon sequence variants (ASV's) that were significantly ($p$<0.001) more abundant in samples from 2019 (3 years after switchgrass planting) relative to samples from 2016 (before switchgrass planting). Of those, we found 96 ASV's with log2 fold changes greater than 20 (or > 1,048,576-fold change). Some taxa were enriched at multiple sites, indicating that regardless of the soil environment, these microbes share a tight association with switchgrass. For instance, we identified 3 ASV's belonging to the genus *Sphingomonas* that were highly enriched (>20 log2 fold change) at all three sites. *Sphingomonas* likely play a role in plant pathogen resistance and may be a valuable member of the switchgrass microbial consortia by aiding in disease suppression.[1][2] Altogether, these results reveal specific soil bacteria that were affected by the presence of switchgrass and provide a first step in understanding the relationships between soil microorganisms and switchgrass. However, many questions remain.

Future efforts will explore whether switchgrass cultivars are locally adapted to recruit soil microbes specific to their site of origin, and determine what, if any, functional benefits these microbes may provide. Using shotgun metagenomics within the context of a nitrogen fertilization experiment, we will focus on microbially mediated belowground processes, especially pathways that regulate plant available nutrient supply, and examine the potential for microbially-mediated pathogen resistance. We are also using high-throughput bacterial isolation to establish a catalog of diverse switchgrass-associated microbes. To date, we have isolated over 1,200 unique bacterial strains from three switchgrass planting sites, with whole genome sequences being assembled for each isolate. The collection is diverse, encompassing over 68 bacterial genera from 5 different phyla. We are compiling synthetic communities to test the impact on plant nutrition, root architecture, and disease resistance. This collection will be an important shared resource for the switchgrass microbiome research community and could help to improve efforts towards sustainable switchgrass production.

**References/Publications**


1. Innerebner, G., Knief, C., Vorholt, J.A., 2011. Protection of Arabidopsis thaliana against leaf-pathogenic *Pseudomonas syringae* by Sphingomonas strains in a controlled model system. Applied and Environmental Microbiology 77, 3202–3210. doi:10.1128/AEM.00133-11
2. Matsumoto, H., Fan, X., Wang, Y. *et al.* 2021. Bacterial seed endophyte shapes disease resistance in rice. Nat. Plants 7**,** 60–72. doi:10.1038/s41477-020-00826-5



**Funding Statement:** This research was supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers DE-SC-0014156 and DE-FOA-0002214 to T.E.J. The work conducted by Argonne National Laboratory was supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

# *Agrobacterium*-mediated Transformation of *P. hallii*

Anthony Trieu[1] (ttrieu@hudsonalpha.org), Xiaoyu Weng[2], Rebekah Wood[1], **Thomas Juenger[2]**, Kankshita Swaminathan[1]

[1]HudsonAlpha Institute for Biotechnology, Huntsville, AL; [2]The University of Texas at Austin, Austin, TX

https://sites.cns.utexas.edu/juenger_lab/switchgrass

## Project Goals

*Panicum hallii* Vasey (Hall's panicgrass) is a diploid, self-compatible, compact, perennial $C_4$ grass in the family Poaceae, which has the potential to serve as a model for perennial bioenergy grasses like switchgrass (*Panicum virgatum* L.). Our goal is to develop tools and resources that will enable us to use *P. hallii* to understand the role and function of genes of interest in switchgrass.

## Abstract

Switchgrass (*Panicum virgatum L.*) is native to the tallgrass prairie of North America. The high yielding potential and the ability to grow well in marginal lands make switchgrass an ideal species as a dedicated bioenergy feedstock. However, its large genomes with varied ploidy levels from diploid (2x) to dodecaploid (12x) with tetraploid and octoploid being the most common, large stature, long-life cycle, and self-incompatiblity, makes genetic and gene function analysis difficult. *Panicum hallii* is closely related to switchgrass, has a compact diploid genome, a shorter generation time, is perennial, and self-compatible. These features permit replicated experimentation in a laboratory setting. *P. hallii* also shows massive genetic variation in many important traits related to biomass production, including timing of flowering, growth rate, disease susceptibility, and drought tolerance. This combination of genomic and physiological attributes makes *P. hallii* an ideal genetic model for switchgrass and other C4 perennial grasses. Here we describe the development of genetic transformation for *P. hallii* using the FIL2 genotype (*P. hallii* var. filipes), a representative of the lowland ecotype of *P. hallii*.

## Funding

This research was supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers DE-SC0021126.

**Carbon Conserving Redox Balanced Co-Utilization of Aromatics and Sugar by Engineered *Pseudomonas putida***

Deepika Awasthi*[1] (dawasthi@lbl.gov), Blake A. Simmons[1] and Steven W. Singer[1] and **Jay Keasling[1,2]**

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA
[2] University of California, Berkeley, CA

www.jbei.org

**Project goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.**

Lignocellulosic biomass is presently an underutilized resource due to lack of a microbial host that can consume all the simple and complex carbohydrates present in the biopolymer. Here we present a rewired central carbon metabolism approach that integrates phosphoketolase bypass in *P. putida* strain KT2440 mutants for ED (Entner-Doudoroff) pathway. We see its impact on reviving growth on glucose minimal medium and co-utilization of aromatics and glucose, to support synthesis of a key cellular intermediate (acetyl-CoA) for biofuel/chemical production.

Funding: Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

# Title: Liquid chromatography-mass spectrometry analysis of lignin depolymerized products from plant hydrolysates

Ramu Kakumanu,[1] Edward E. K. Baidoo[1]* (eebaidoo@lbl.gov), Hemant Choudhary,[2] John Gladden,[2] Blake Simmons,[1] Christopher J. Petzold,[1] Paul D. Adams,[1] and **Jay D. Keasling**[1,3]

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]Sandia National Laboratory, Livermore, California; [3]University of California, Berkeley, California

https://www.jbei.org/

**Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.**

There is a growing demand to produce renewable and sustainable energy from plant-based biomass. However, the structural support tissue of vascular plants, i.e. lignin, is recalcitrant under ambient conditions, which poses a considerable technical challenge to the conversion of lignin to fuels and value-added coproducts. Consequently, lignin requires chemical (e.g., base catalysis and ionic liquid, IL) and/or enzymatic pretreatment for effective depolymerization. In this regard, the physicochemical properties of ILs enable the efficient deconstruction and fractionation of biomass needed for conversion of lignocellulosic feedstocks to fuels and chemicals and offers the potential for lignin valorization towards enhanced economic viability of lignocellulosic biorefineries. As a result, it is imperative that the resulting hydrolysates from such refineries be characterized. To this end, we have developed a high-throughput liquid chromatography-mass spectrometry (LC-MS) method to quantify products of lignin depolymerization and/or residual phenolic monomers from hydrolysates produced via cholinium lysinate ([Ch][Lys]) IL pretreatment and enzymatic saccharification. The method showed good linearity ($R^2 = \geq 0.99$) and retention time repeatability (% RSD = < 0.5) for the analytes tested. The method revealed different phenolic profiles for poplar and sorghum hydrolysates, with 4-hydroxybenzoic acid being prominent in both. This suggests that the method can be used to determine the phenolic contents of hydrolysates, assess phenolic compound viability (i.e., extent of degradation) and uptake by microbes in the biorefinery, and potentially quantify the amounts of value-added coproducts produced.

Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

# Multiproduct Cellulosic Biorefinery Enables Market-Competitive Gasoline and Jet Fuel Blendstocks

Nawa Raj Baral,[1*](nrbaral@lbl.gov) Minliang Yang,[1] Deepanwita Banerjee,[1] Thomas Eng,[1] Aindrila Mukhopadhyay,[1] John Gladden[2], Blake Simmons[1], **Corinne D. Scown**[1]

[1]Lawrence Berkeley National Laboratory; [2]Sandia National Laboratories

**Project Goals:** Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

## Abstract

Establishing sustainable cellulosic biorefineries requires the production of renewable biofuels and bioproducts at market-competitive prices while reducing their carbon footprints. Achieving these targets is challenging with a single-fuel biorefinery warranting a cellulosic biorefinery with multiple fuels and products equivalent to a petroleum refinery. Here, we modeled a multiproduct biorefinery combining high-energy-density renewable gasoline and jet fuel blendstocks as well as nonhazardous biomass-derived indigoidine (blue dye). The model biorefinery utilizes ensiled biomass sorghum feedstock. Results show market-competitive selling prices of fuels of $0.6/L-petroleum-equivalent and indigoidine of $5.7/kg at the current state-of-the-technology. At present, the multiproduct biorefinery generates annual revenue of $434±10/bone-dry-metric ton (bdt) of biomass—141 to 156% more than single-fuel biorefinery—while meeting the carbon footprint reduction mandate of cellulosic biofuels. With an improved biomass conversion—achieving targeted sugar and lignin utilizations including 95% glucose, 85% xylose, and 95% lignin monomers—the annual revenue and the total carbon footprint reduction, respectively, could increase to $697±18/bdt and 94±3%, but envisioned when the biobased dye fully substitutes the synthetic dye.

**One-pot deconstruction and conversion of dry and ensiled sorghum**

Julius Choi[1,2]* (juliuschoi@lbl.gov), Seema Singh[1,2], Blake Simmons[1,3], and John Gladden[1,2]

[1]Joint Bioenergy Institute, Lawrence Berkeley National Laboratory, 5885 Hollis St, Emeryville, CA 94608, USA; [2]Sandia National Laboratories, 7011 East Ave, Livermore, CA 94551, USA; [3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

URL: www.jbei.org

**Project Goal: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts**

The development of low cost and high efficiency biomass deconstruction methods is a crucial step in the commercialization of biorefinery. Ensiling is an anaerobic biomass storage method producing some acids like lactic acids and acetic acids beneficial for biomass conservation. Furthermore, these acids generated during ensiling storage could serve as an in-situ mildly acidic pretreatment leading to improved deconstruction efficiency[1,2]. [Ch][Lys] based-one pot process has been demonstrated as an efficient method to deconstruct biomass[3]. The integration of the two processes is expected to improve the efficiency of the biomass pretreatment step. This approach has been investigated with different sorghum hybrids

**References**

1) Essien et al. (2018), Ensiled wet storage accelerates pretreatment for bioconversion of corn Stover, *Front. Bioeng. Biotechnol*, 6, 195
2) Thomsen et al. (2016) Combination of ensiling and fungal delignification as effective wheat straw pretreatment, *Biotechnol Biofuels*, 9, 16
3) Sun et al.(2017), One-pot integrated biofuel production using low-cost biocompatible protic ionic liquids*, Green Chemistry*, 19, 3152-3163

Funding Statement

Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

# Field testing of transgenic Sorghum variants, overexpressing 4 types of aromatic compounds

Jutta Dalton (jdalton@lbl.gov)[1]*, Yang Tian[1], Chris DeBen[#], Albert Kausch[3], Chien-Yuan Lin[1], Aymerick Eudes[1], Daniel Putnam[1a,2],Corinne Scown[1a], Henrik Scheller[1] **, Jay Keasling[1]**

[1] Feedstocks Division, [1a]LEAD-Lifecycle Economy and Agronomy Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, [2] University of California Davis, Davis, [3] University of Rhode Island, West Kingston

www.jbei.org

Project goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Sorghum is a unique crop for Bioenergy production, as all its parts, the grain and the stover can be used as sustainable feedstocks through various production routes. It also has low input requirements, making it an ideal candidate to improve on those natural qualities to increase feedstock production. Aromatic compounds are of particular interest for biofuel production in plants since they increase Octane ratings and the catalysis in plants provides an economic way to produce such desirable co-products. In this study, we compare the performance of 4 transgenic variants, expressing different aromatic compounds for biomass yield under non-controlled conditions: coumaric acid, Hydrobenzoic acid, Protocatechuic acid and aromatic metabolites from the shikimate pathway. After one growing season in the field, biomass yield of both grain and stover was analyzed by weight. A significant increase in total yield was only determined for lines expressing protocatechuic acid, with the majority of that increase coming from grain. These lines are therefore excellent candidates for large scale biofuel production and analysis of these lines is ongoing.

Funding statement: Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**Understand the Nanoarchitecture of Native and Engineered Plant Cell Wall via Multi-dimensional Solid-state NMR**

Yu Gao[1,2]* (yugao@lbl.gov), Andrew Lipton[3], Coyla Munson[4], Yuuki Wittmer[4], Dylan Murray[4], Mark Bowden[3], Henrik Scheller[1,2,5], Jenny Mortimer[1,2,6], and **Jay D. Keasling**[1,2,5]

[1]Joint BioEnergy Institute, Emeryville, CA, USA; [2]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [3]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA; [4]Department of Chemistry, University of California, Davis, CA, USA; [5]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA; [6]School of Agriculture, Food and Wine, University of Adelaide, Glen Osmond, SA, Australia

**Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.**

**URL:** www.jbei.org

**Abstract:** Lignocellulosic biomass will be a major sustainable feedstock for the burgeoning bioeconomy. Understanding the biosynthesis and nanoarchitecture of the plant cell wall provides indispensable insights to enable predictive cell wall engineering for bioenergy crops. Multi-dimensional solid-state NMR spectroscopy allows detailed investigation of the composition and organization of biopolymers in the intact wild-type and engineered plant cell walls, which allows us to better understand the impact of genetical modification on the native plant cell wall architecture and refine our genetical models of the plant cell wall. Here, by successfully generating mature plants, including Arabidopsis, sorghum, switchgrass and poplar, with high $^{13}C$ incorporation (>90 %) in a customized growth chamber, we employed a series of multi-dimensional ssNMR experiments (e.g., refocused INADEQUATE, PDSD, T1 relaxation measurements) to understand the arrangement of biopolymers in the native secondary cell wall of these plants and understand the architecture of their intact cell walls. Then, using the plants engineered to have altered biomass, we tested the effects of these modifications, which will provide insights for future plant engineering strategies.

**References/Publications**

1. Gao, Y., Lipton, A.S., Wittmer, Y. *et al.* A grass-specific cellulose–xylan interaction dominates in sorghum secondary cell walls. *Nat Commun* **11,** 6081 (2020). https://doi.org/10.1038/s41467-020-19837-z
2. Gao, Y., Lipton, A.S., Munson, C. *et al*. Elongated galactan side-chains mediate cellulose-pectin interactions in engineered Arabidopsis secondary cell walls. *The Plant J.*, in revision

*This work was conducted as part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. Part of this work was conducted by the Environmental Molecular Sciences Laboratory (grid.436923.9), a DOE Office of Science scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at PNNL under contract DE-AC05-76RL01830.*

**Engineering and optimization of lignin catabolic pathways in *Rhodosporidium toruloides***

Valentina E. Garcia,[1,2*] Peter B. Otoupal,[1,2,3*] (PeterOtoupal@lbl.gov) Gina M. Geiselman,[1,2,3*] Joonhoon Kim,[2,3,4] Hyun Gyu Lim,[1,5] Ramu Kakumanu,[1,6] Jon K. Magnuson,[2,3,4] Edward E.K. Baidoo,[1,6] Adam M. Feist,[1,5] Jay D. Keasling,[1,6] Blake A. Simmons,[1,6] & **John M. Gladden**[1,2,3] (jmgladd@sandia.gov)

[1]Biomanufacturing and Biomaterials Department, Sandia National Laboratories, Livermore, CA; [2]Deconstruction Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA; [3]Agile BioFoundry, Department of Energy, Emeryville, CA; [3]Chemical and Biological Processing Group, Pacific Northwest National Laboratory, Richland, WA; [5]Department of Bioengineering, University of California, San Diego, CA; [6]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA

https://www.jbei.org/research/deconstruction/fungal-biotechnology/

**Project Goals:**

The vision of JBEI is that bioenergy crops can be converted into economically-viable, carbon-neutral, biofuels and renewable chemicals currently derived from petroleum, and many other bioproducts that cannot be efficiently produced from petroleum. JBEI's mission is to establish the scientific knowledge and new technologies in feedstock development, deconstruction and separation, and conversion needed to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts. When fully scaled, JBEI's technologies will enable the production of replacements for petroleum derived gasoline, diesel, jet fuel, and bioproducts. In doing so, JBEI will reduce the nation's dependence on fossil fuels, significantly reduce the amount of carbon added to the atmosphere, reduce contamination of the environment, and provide the scientific tools and knowledge required to transform the bioenergy marketplace.

**Abstract:**

The oleaginous basidiomycete, *Rhodosporidium toruloides* (also known as *Rhodotorula toruloides*), is rapidly gaining traction as an industrial host for converting lignocellulosic biomass into value-added products. The yeast is naturally capable of co-utilizing multiple carbon sources such as pentoses, hexoses, and lignin-derived aromatics. It also naturally accumulates substantial amounts of acetyl-CoA and malonyl-CoA, the building blocks for many desirable biochemicals, within its cytosol. Many labs within the Department of Energy and academia have leveraged these properties to use *R. toruloides* to valorize biomass into compounds such as bisabolene.

However, these efforts have largely focused on utilization of pentose and hexose feedstocks, while neglecting the carbon in biomass that is contained within aromatic compounds. To fully utilize the carbon available in lignocellulosic feedstocks, efforts must be directed to valorize these aromatics. Here, we describe our efforts with the Joint BioEnergy Institute to do so. We focus primarily on the aromatic compound *p*-coumarate, as the metabolic pathway to this compound in *R. torulolides* is well-established in through the enzymes phenylalanine ammonia lyase and 4-coumarate-CoA ligase.

We first build upon our metabolic model of *p*-coumarate degradation in *R. toruloides* by employing CRISPR editing to delete key steps within the pathway. We show that this leads to substantial accumulation of key beachheads, protocatechuate (PCA, 6.7 ± 2.2 g/L) and 4-

hydroxybenzoate (4HBA, 4.6 ± 0.1 g/L) in minimal media. We also show considerable buildup of 4HBA (1.3 ± 0.1 g/L) during fermentation of one of these strains in hydrolysates derived from lignocellulosic biomass. We next describe our attempts to integrate four heterologous pathways for production of curcuminoids, naringenin, resveratrol, and 2-pyrone-4,6-dicarboxylic acid (PDC). We obtain the most promising results with the later compound, observing signs of its precursor compound 4-carboxy-2-hydroxymuconate-6-semialdehyde (CHMS) as well as 0.4 g/L of PDC itself. Finally, we describe our efforts to adapt *R. toruloides* to better utilize *p*-coumarate by employing Tolerance Adaptive Laboratory Evolution (TALE), generating an evolved strain that robustly grows in media with 20 g/L *p*-coumarate as the sole carbon source. Taken together, these results describe how the Joint BioEnergy Institute has sought to maximize the productive potential of *R. toruloides* as an industrial chassis for fully valorizing the carbon in lignocellulosic biomass.

*This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science. Work conducted at the DOE Joint BioEnergy Institute was supported by the U.S. DOE, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02- 05CH11231*

**A media optimization pipeline for improving TRY with machine learning**

Matthew Incha[1,2,3*](incha@lbl.gov), Tijana Radivojević[2,3], Vincent Blay Roger[2,3], Matthias Schmidt[2,3,4], Hector Garcia Martin[2,3], & **Jay D. Keasling[1,2,3,5]**

[1]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, 14 USA; [2]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA 94608, USA; [3]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, 10 CA 94720, USA; [4]Institute of Applied Microbiology (iAMB), Aachen Biology and Biotechnology (ABBt), RWTH 12 Aachen University, Aachen, Germany; [5]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 20 94720, USA

**Project Goal:**
Improve the biosynthetic performance of an engineered microorganism via machine-learning guided optimization of media components. Using a robotics platform, many unique media formulations can be tested. Titers achieved in each formulation will inform a Bayesian model over 5-10 iterations which will then design new improved media formulations.

**Abstract:**
Bioprocess optimization is often nonintuitive and requires many iterations of the canonical design, build, test, learn (DBTL) cycle. One method for improving biosynthetic performance is by altering the media conditions with design of experiment (DoE) tools. Using robotics and colorimetric screening, these processes can be conducted quickly and effectively. To this end, we chose the red compound, flaviolin, as our first product for this pipeline. A genome integrated production strain with a constitutive promoter driving flaviolin biosynthesis (*rppA*) was developed. This strain demonstrated highly reproducible titers in minimal media and was used as the reporter for media performance. A robotics platform was further developed to rapidly iterate over diverse media formulations and assay production via simple spectrophotometry. Following platform development 22 media formulations based on MOPS minimal medium were designed using Latin Hypercube Sampling (LHS)[1]. Using the LHS results as training data, a machine learning model developed in JBEI, the Automated Recommendation Tool (ART)[2], designed new formulations. Following 5 DBTL cycles, ART consistently designed media with 2x titer improvement over MOPS. Following strain construction, these improved media formulations were developed in less than 2 months, illustrating the cost & time effectiveness of this machine learning media optimization pipeline.

**References**
[1]McKay, M. D., R. J. Beckman, and W. J. Conover. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code." *Technometrics* 21, no. 2 (1979): 239–45. https://doi.org/10.2307/1268522.
[2]Radivojević, T., Costello, Z., Workman, K. et al. A machine learning Automated Recommendation Tool for synthetic biology. Nat Commun 11, 4879 (2020). https://doi.org/10.1038/s41467-020-18008-4

**Funding Statement**
Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**Title:** Multifaceted Adaptive Laboratory Evolution Approaches for Improving Host Phenotypes

**Authors:** Hyun Gyu Lim (hyl003@ucsd.edu),[1,2*] Thomas Eng,[2,3] Xi Wang,[2,3] Myung Hyun Noh,[1] Deepanwita Banerjee,[2,3] Russel Menchavez,[2,3] Blake A. Simmons,[2,3] Bernhard O. Palsson,[1,2] Take Soon Lee,[2,3] Aindrila Mukhopadhyay,[2,3] Adam M. Feist,[1,2] and **Jay Keasling**[2,3]

**Institutions:** [1]University of California San Diego; [2]Joint BioEnergy Institute, Berkeley, CA; [3]Lawrence Berkeley National Laboratory, Berkeley, CA

**Website URL:** http://www.jbei.org

**Project Goals:** Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

**Abstract Text:**

Natural microorganisms are not optimized for industry-relevant conditions and therefore often show unsatisfactory performance. In this study, multifaceted Adaptive Laboratory Evolution (ALE) approaches were applied for *Pseudomonas putida* KT2440 to improve its phenotypes to achieve efficient biochemical production. Firstly, we constructed efficient xylose and galactose catabolism by continuously evolving engineered KT2440 stains in a constant or weaning-off condition. Secondly, we significantly improve tolerance against isoprenol by growing cells with gradually increased isoprenol concentrations. With these multiple ALE strategies successfully generated strains with improved phenotypes. Whole-genome sequencing, transcriptome sequencing, and reverse engineering unveiled mutational mechanisms, enabling the improved phenotypes. Collectively, ALE proved to be an effective method for strain optimization and knowledge generation which can be utilized to develop efficient microbial bioprocesses.

**References/Publications**

1. Lim et al. (2021) Generation of *Pseudomonas putida* KT2440 Strains with Efficient Utilization of Xylose and Galactose via Adaptive Laboratory Evolution, *ACS Sustainable Chem. Eng*. 9(34), 11512–11523
2. Lim et al. (2020) Generation of Ionic Liquid Tolerant *Pseudomonas putida* KT2440 Strains via Adaptive Laboratory Evolution. *Green Chem*. 22(17), 5677–5690.
3. LaCroix et al. (2017) A Model for Designing Adaptive Laboratory Evolution Experiments. *Appl. Environ. Microbiol*., 83(8):e03115-16.

**Funding Statement:** This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**Title:** LigCHIP Technology for Bond-Specific Analysis of Lignocellulose Deconstructing Enzymes

**Authors:** Jenny R. Onley[1,2]* (jronley@lbl.gov), Kai Deng[1,2], Noel S. Ha[1], Markus de Raad[1], Le Thanh Mai Pham[1,2], Kenneth L. Sale[1,2], Paul D. Adams[1], Trent R. Northen[1], **Jay D. Keasling**[1]

**Institutions:** [1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA, USA; [2]Sandia National Laboratories, Livermore, California, USA

**Website URL:** https://www.jbei.org/

**Project Goals:** Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

**Abstract Text:** Lignocellulosic biomass has great potential to provide a sustainable source of fuel, and enzymatic deconstruction of lignocellulose is a key component of biofuel production. Lignocellulose is composed of cellulose, hemicellulose, and lignin. While deconstruction of cellulose and hemicellulose has improved greatly in recent years, further optimization is required to make the process efficient enough for biofuels to compete in the energy market. Lignin is composed of a variety of subunits linked together with a variety of bond types. We have a limited understanding of the enzymatic mechanisms for the breakage of different lignin bonds. Lignin is highly recalcitrant, and breakdown produces a complex, heterogenous mixture of products. Further understanding of the activities of enzymes that degrade cellulose, hemicellulose, and especially lignin is required to lower biofuel production costs and increase efficiency. We utilized the unique mass spectrometry assay (nanostructure-initiated mass spectrometry or "NIMS")[1]. Briefly this platform uses lignin model compounds containing target lignin bonds enabling bond-specific kinetic analysis of enzyme activities. In addition to the lignin linkages, these substrates have perfluorinated and cationic moieties enabling rapid separation from complex mixtures and efficient ionization from small samples. Each substrate has a unique mass enabling them to be analyzed from mixtures. Quantification is achieved using $^{13}C$ labeled internal standards. NIMS can be scaled up when it is combined with mass spectrometry imaging (MSI) and open-source MSI software[2], providing rapid measurements of multiple samples. Additionally, coupling mass spectrometry methods with microfluidics further increases the high throughput power of these technologies.[3] We constructed a microfluidics device for combinatorial screening of enzyme activities though droplet microfluidics.[4] A commercial droplet generator is used to prepare cocktails of droplets containing different enzymes and substrates. Our system spontaneously merges random pairs of droplets in parallel, replacing deterministic liquid handling operations and providing the potential for testing many more combinations of enzymes and substrates. The microfluidics device was coupled with NIMS, depositing the reaction products onto the NIMS chip for MSI analysis. We demonstrated the use of these technologies with model lignocellulose compounds that represent common linkages in lignin (β-O-4', β-β', and 5-5'), cellulose, and hemicellulose. In the past we observed

cleavage of the β-O-4' bond, testing lignolytic enzymes alone and in combination with glycoside hydrolases. Through collaboration with the Deconstruction Division at JBEI, we presented detailed studies for the degradation of a phenolic β-O-4 dimeric model compound by lignin peroxidase isozyme H8 as a function of pH.[5] Here we used our microfluidics device to gather performance data for an array of lignocellulolytic enzymes. Interestingly we demonstrated the oxidative cleavage of β-β'-linked lignin by fungal laccases, a phenomenon that is poorly characterized in the literature. Further work with the LigCHIP (our unique technological combination of NIMS, MSI, droplet microfluidics, and lignin substrates) will include scaling up and expanding these technologies to a greater number of enzymes and substrates.

**References/Publications**


1. Deng, K. *et al.* Rapid characterization of the activities of lignin-modifying enzymes based on nanostructure-initiator mass spectrometry (NIMS). *Biotechnol. Biofuels* **11**, 1–9 (2018).

2. De Raad, M. *et al.* OpenMSI Arrayed Analysis Toolkit: Analyzing Spatially Defined Samples Using Mass Spectrometry Imaging. *Anal. Chem.* **89**, 5818–5823 (2017).

3. Ha, N. S. *et al.* Faster, better, and cheaper: harnessing microfluidics and mass spectrometry for biotechnology. *RSC Chem. Biol.* **2**, 1331–1351 (2021).

4. Ha, N. S., de Raad, M. & Northen, T. R. Picoliter droplet sample processing and deposition for mass spectrometry. 1–21 (2021). Patent 17/062,609.

5. Pham, L. T. M. *et al.* Experimental and theoretical insights into the effects of pH on catalysis of bond-cleavage by the lignin peroxidase isozyme H8 from *Phanerochaete chrysosporium*. *Biotechnol. Biofuels* **14**, 1–11 (2021).



**Funding Statement:** The Joint BioEnergy Institute is managed by Lawrence Berkeley National Laboratory for the U.S. Department of Energy under contract no. DE-AC02-05CH11231. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research.

**Microbiomes that metabolize lignin fragments obtained by chemical pretreatment of biomass**

Mee-Rye Park[1,2] (mrpark@lbl.gov), Jipeng Yan[2,3], Jeffrey G. Pelton[4], Ning Sun[2,3], Blake A. Simmons[1,2], Steven W. Singer[1,2], **Jay Keasling**[1,2,5]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA [2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Advanced Biofuels Process Demonstration Unit, Emeryville, CA 94608; [4]QB3 Institute, University of California-Berkeley, Berkeley, CA 94720; [5]Department of Chemical and Biomolecular Engineering, Department of Bioengineering, University of California, Berkeley, CA, USA

http://jbei.org

**Project Goals**: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

**Abstract**
Overcoming the recalcitrance of lignin and developing conversion strategies for aromatics are key goals to maximize conversion of carbon in plant biomass. While monoaromatics are metabolized by a variety of bacteria, the metabolism of aromatic oligomers remains largely undescribed. Several Sphingomonad are known to depolymerize dimeric lignin model compounds. Beyond the dimeric model compounds, using HMW lignin directly extracted from lignocellulosic biomass would be beneficial to explore diverse lignin-degrading microbes and their enzymes. Here we generated aromatic mixtures of monomers and oligomers from IL pretreatment (referred to BCD liquor) then enriched microbiomes originated from diverse soils on the BCD liquor. Chemical analysis (GPC and 2D HSQC NMR) demonstrated lignin depolymerization by cleavage of $\beta$-O-4 linkage during 2-week incubation. Time series metatranscriptomics revealed that predominant microbes produce enzymes involved in lignin depolymerization to aromatic catabolism and ring cleavage.

**Reference**
Park, Mee-Rye, et al. "Response of *Pseudomonas putida* to Complex, Aromatic-Rich Fractions from Biomass." ChemSusChem 13.17 (2020): 4455-4467.

**Funding statement**: Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**Integrated chemical and biological catalysis for the valorization of ionic liquid-based biorefinery lignin**

Le Thanh Mai Pham[1,2]* (thanhmaipl@lbl.gov), Hemant Choudhary[1,2], Steven W. Singer[1,4], Blake A. Simmons[1,4], John M. Gladden[1,2], Kenneth L. Sale[1,3], Seema Singh[1,2], **Jay Keasling**[1]

[1]Deconstruction Division, Joint BioEnergy Institute, Emeryville, CA; [2]Department of Biomaterials and Biomanufacturing, Sandia National Laboratories, Livermore, CA; [3]Department of Computation Biology and Biophysics, Sandia National Laboratories, Livermore, CA; [4]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA

**Website URL:** www.jbei.org

**Project Goals:** Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Lignin, a complex and heterogenous polymer composed of phenylpropane moieties, is the most abundant bio-renewable source of aromatics. Due to its heterogeneity and recalcitrance to chemical or enzymatic depolymerization, lignin is currently under-utilized and routinely combusted to generate process heat in the paper and pulp industry. Chemical routes to lignin depolymerization and subsequent valorization are efficient, but can generate toxic chemicals, making the process unsustainable. Enzymatic depolymerization is selective but inefficient at degrading lignin in solid state. We have developed an integrated Chem-Bio route that combines the higher efficiency of chemical lignin depolymerization with the higher selectivity of enzymes to produce targeted value-added chemicals and compounds amenable to biological upgrading. We first studied depolymerization of Pd/ZrP catalyzed hydrodeoxygenation (HDO) of ionic liquid (IL) pretreated lignin from poplar. GC-MS analysis and molecular weight distribution profiling showed the process efficiently cleaved C-O bonds as a function of temperature and time. Based on analysis of lignin content, the process produced high yields (>29.3%) of a lignin oil with minimum char formation (<15%). In general, products obtained after HDO of lignin such as phenols and guaiacols are not biocompatible, so we investigated the potential of highly active and thermostable laccases from Cerrena unicolor to upgrade the HDO products to biocompatible compounds. We observed that laccases could detoxify up to 89.0% non-biocompatible HDO products to such as syringic acid and vanillin with a selectivity of 48.2% and 40.0%, respectively. Thus, our integrated chemical and biological approach enables the conversion of lignin oil into platform molecules.

**Funding Statement:** This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

# High-Throughput Screening of Wild Type and Engineered Plants using the JBEI Feedstocks-To-Fuels (F2F) Pipeline

Venkataramana R. Pidatala[1*](ramana@lbl.gov), Stephen Tan[1], Chris Petzold[1], Hector Garcia Martin[1], Corinne Scown[1], Seema Singh[2], John Gladden[2], Blake Simmons[1], **Jay Keasling**[1]

[1]BioSciences, Lawrence Berkeley National lab, Berkeley, California
[2]Sandia National laboratories, Livermore, California

www.jbei.org

**Project goals:** Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

**Abstract:**
Efficient conversion of lignocellulosic biomass into biofuels and bioproducts requires bioengineering of plants, microbes. Biomass is deconstructed using ionic liquids and enzymes, resulting sugars are converted by microbes into desired bioproducts. Testing the saccharification and conversion efficacy of hundreds of transgenic plants, various strains of microbes and Ionic liquids with different possible combinations is a daunting task. Evaluation and identification of best candidate plants, microbes at bench scale are cumbersome and time taking. A high-throughput screening platform (F2F: Feedstocks-To-Fuels) is developed to address these challenges. With end-to-end automated processes combined with machine learning tools the pipeline enables the flexibility to adjust multiple experimental parameters and predictive deconstruction of various biomass samples. Methods and components of the Pipeline and saccharification and fermentation results from screening of wild type and engineered sorghum plants are presented.

**Funding statement:**
Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**Structural characterization of lignin degrading enzyme PmdC, involved in the synthesis of polymer precursor 2-pyrone 4,6 dicarboxylic acid (PDC)**

Andria Rodrigues[1,2*](avrodrigues@lbl.gov), Jose Henrique Pereira[1,2], Andy DeGiovani[1,2], Ramu Kakumanu[1,3], Chien-Yuan Lin[1,4], Edward Baidoo[1,3], Aymerick Eudes[1,4], Paul Adams[1,2,5] and **Jay Keasling** [1,3,5,6]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley - CA, [2]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, [3]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, [4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, [5]Department of Bioengineering, University of California, Berkeley, [6]Department of Chemical and Biomolecular Engineering, University of California, Berkeley.

www.jbei.org

**Project Goals:** Establish scientific knowledge & new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Lignin is one of the most abundant materials available on earth, with 300 billion tons available globally and 20 million tons added annually providing an extensive carbon source for sustainable manufacturing of products. However, lignin remains largely under-utilized due to its tough, woody nature and inherent recalcitrance. Efficient utilization of plant biomass especially lignin, is crucial for producing biofuels to scale. Enzymatic degradation of lignin is currently an area of intense research focus as bacteria and fungi employ a variety of ligninolytic enzyme systems to breakdown lignin. The aromatics derived from lignin breakdown can serve as precursors for synthetic pathways for the production of biofuels and high value compounds such as food additives or polymers. Here, we report the protein structure of PmdC *(Comamonas)* a homolog of LigC *(Sphingomonas)*. This protein catalyzes the conversion of lignin breakdown intermediate 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase (CHMS) from the protocatechuate cleavage pathway to yield 2-pyrone-4,6-dicarboxylic acid (PDC), a precursor for the synthesis of useful polymers such as polyamides, polyesters and polyurethane. Metabolic engineering efforts are already underway for creating biosynthetic routes for the synthesis of PDC in appropriate host organisms. The structure of PmdC will yield valuable insight into the mechanism of catalysis of the enzyme and will support protein engineering efforts to boost PDC production.

**References:**

1) Efficient production of 2-pyrone 4,6-dicarboxylic acid as a novel polymer-based material from protocatechuate by microbial function. Otsuka et al., Biotech. Products & Process Engineering 2005
2) In-planta production of the biodegradable polyester precursor 2-pyrone-4,6-dicarboxylic acid (PDC): Stacking reduced biomass recalcitrance with value-added co-product. Lin et al., Metabolic Engineering 2021

*This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

# Comparison of Isoprenol Production from Sorghum Biomass Hydrolysates using Engineered Microbial Hosts

Aparajitha Srinivasan[1*] (asrinivasan@lbl.gov), Ian Yunus[1], Russel Menchavez[1], Deepanwita Banerjee[1], Asun Oka[1], Ning Sun[1], Blake Simmons[1], Taek Soon Lee[1], Thomas Eng[1], Aindrila Mukhopadhyay[1] and **Jay D. Keasling[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley, California

www.jbei.org

**Project Goals**: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

**Abstract:**
Biologically produced Isoprenol can be chemically converted to DMCO[1], an aviation fuel. While many microbial hosts have been engineered to produce isoprenol on defined growth media[2,3], it is essential to understand the host response to renewable carbon feedstocks. In this study we compare a range of engineered microbial hosts - *Pseudomonas putida*, *Corynebacterium glutamicum* and model microbes such as *Escherichia coli*, for production of isoprenol using ionic liquid-pretreated, dry and ensiled sorghum biomass hydrolysates supplemented growth media. Growth of all the three hosts was not significantly affected and the highest production of ~ 1 g/L isoprenol from ensiled biomass hydrolysate supplementation was observed. Residual sugar analysis also enabled understanding of host capability to utilize the sorghum biomass components for growth and production. The observed metabolic flux rerouting on hydrolysates needs further investigation using functional genomics. Improving fitness using adaptive laboratory evolution will also help to optimize this bioconversion process for sustainable production of biofuels from real world carbon streams.

**References**

1. Baral NR, Yang M, Harvey BG, et al. (2021). Production Cost and Carbon Footprint of Biomass-Derived Dimethylcyclooctane as a High-Performance Jet Fuel Blendstock. ACS Sustainable Chemistry & Engineering, 9 (35), 11872-11882.
2. Kang A, Mendez-Perez D, Goh EB, et al. (2019). Optimization of the IPP-bypass mevalonate pathway and fed-batch fermentation for the production of isoprenol in *Escherichia coli*. Metabolic engineering, 56, 85-96.
3. Sasaki Y, Eng T, Herbert RA, et al. (2019). Engineering *Corynebacterium glutamicum* to produce the biogasoline isopentenol from plant biomass hydrolysates. Biotechnology for biofuels, 12(1), 1-15.

**Funding statement**: This research was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**High Throughput Bioengineering Using a Microfluidic Platform**

Jess Sustarich[1,2]*(jsustarich@lbl.gov), Swarnagowri Vaidyanathan[1,2]*(swarna_vaidy@lbl.gov), Lauren Washburn[1,2], William R. Gaillard,[1,2] Kosuke Iwai,[1,2] Peter W. Kim,[1,2] Kai Deng,[1,2] Stephen Tan,[1,3] Trent R. Northen,[1,4,6] Nathan J. Hillson,[1,3,6] Hector Garcia Martin,[1,3,4] Paul D. Adams,[1,4,5,7] and **Anup K. Singh**[1,2]

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Biological and Material Science, Sandia National Laboratories, Livermore, CA; [3]Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [6]DOE Joint Genome Institute, Walnut Creek, CA; and [7]University of California, Berkeley, CA

https://www.jbei.org/research/divisions/technology/microfluidic-assays/

**Project Goals: The JBEI mission is to conduct basic and applied research to enable cost-effective conversion of lignocellulosic biomass into biofuels and bioproducts. The goal of this project, performed in the Microfluidic Assays group in the Technology Division at JBEI, is to develop a robust and easy-to-use droplet microfluidic platform to automate the steps involved in engineering of metabolic pathways to produce biofuel molecules.**

Synthetic biology offers a promising approach to produce biofuel and other chemicals. Optimization of metabolic pathways however, requires conducting a large number of experiments that are labor-intensive with repetitive pipetting and plating and require large amounts of expensive reagents. Robotic liquid handling stations represent a solution to automate genetic engineering processes. However, they still require a large volume of reagents and their high equipment and maintenance cost can be prohibitive to many users. Microfluidic platforms offer a promising alternative as they provide improvement over their macroscale counterparts in cost, amounts of reagents required, speed, and integration.

We are developing microfluidic devices for biofuel research applications including enzyme screening, enzyme evolution, and optimization of metabolic pathways. Our droplet-based microfluidic platforms use digital microfluidic (DMF) format where nanoliter aqueous droplets suspended in oil are manipulated on an electrode array using electrowetting on dielectric concept.[1-5] The systems can handle large numbers of droplets at once as well as actively manipulate droplets in a programmable manner, and are capable of multiple steps of droplet manipulation including formation of aqueous droplets and encapsulation of reagents and cells, electric-field driven merge and split of the droplets to add or remove liquid, on-chip electroporation, and incubation steps with localized temperature control. The device uses an array format with 100 elements, each containing sets of electrodes for two electric field actuated operations- electrowetting for merging droplets and electroporation for transformation. Reagents are introduced into the chip by dispensing droplets, are kept separate until ready to mix, mixed on-demand by merging droplets by electrowetting, and transformation of cells by on-chip electroporation.

A novel platform for high throughput electroporation is being developed in an automatable format. A 384 well microtiter plate having a novel electrode geometry is subjected to PEGylated Thiol-

Au chemistry to render the wells hydrophilic. They have individually addressable electrodes to improve the electric field transport for introducing the recombinant DNA into *E. coli* [6] as a part of the CRISPR-based MAGE as an example of how our microfluidic platform provides an alternative solution to the cumbersome traditional methods. Furthermore, with our platform, recovery, incubation and screening can be performed on the same chip. The configuration of the chip uses a 384-well template and is easily integrable with liquid handling robots. We validate our microfluidic chip by performing targeted genomic changes through CRISPR-based MAGE (CRMAGE) recombineering for the biosynthetic pathway producing the sustainable pigment indigoidine in *E. coli*.[7] The automated platform for multiplexed transformation holds the promise of accelerating the design-build-test-learn cycle.[8,9]

## References


1. P.C. Gach, K. Iwai, P.W. Kim, N.J. Hillson, and A.K. Singh, "Droplet Microfluidics for Synthetic Biology," *Lab Chip*, 2017, **17**, 3388-3400.
2. P.C. Gach, S.C.C. Shih, J. Sustarich, J.D. Keasling, N.J. Hillson, P.D. Adams and A.K. Singh, "A Droplet Microfluidic Platform for Automating Genetic Engineering," *ACS Synth. Biol.*, 2016, **5**, 426-433.
3. S.C.C. Sigh, G. Goyal, P.W. Kim, N. Koutsoubelis, J.D. Keasling, P.D. Adams, N.J. Hillson and A.K. Singh, "A Versatile Microfluidic Device for Automating Synthetic Biology," *ACS Synth. Biol.*, 2015, **4**, 1151-1164.
4. S.C.C. Shih, P.C. Gach, J. Sustarich, B.A. Simmons, P.D. Adams, S. Singh and A.K. Singh, "A Droplet-to-Digital (D2D) Microfluidic Device for Single Cell Assays," *Lab Chip*, 2015, **15**, 225-236.
5. P.C. Gach, S.C.C. Shih, J. Sustarich, J.D. Keasling, N.J. Hillson, P.D. Adams and A.K. Singh, "A Droplet Microfluidic Platform for Automating Genetic Engineering," *ACS Synth. Biol.*, 2016, **5**, 426-433.
6. C. Ronda, L. E. Pedersen, M. O. Sommer, and A. T. Nielsen, "CRMAGE: CRISPR Optimized MAGE Recombineering", *Sci. Rep.*, 2016, **6**, 19452.
7. M. Wehrs, J.M. Gladden, Y. Liu,ac Lukas Platz, J.P. Prahl, J. Moon, G. Papa, E. Sundstrom, G.M. Geiselman, D. Tanjore, J.D. Keasling, T.R. Pray, B.A. Simmons and A. Mukhopadhyay, "Sustainable bioproduction of the blue pigment indigoidine: Expanding the range of heterologous products in *R. toruloides* to include non-ribosomal peptides", *Green Chem.*, 2019, **21**, 3394-3406.
8. T. Radivojević, Z. Costello, K. Workman, and H.G. Martin, (2019) ART: A machine learning Automated Recommendation Tool for synthetic biology. *arXiv*.
9. R.D. King, K.E. Whelan, F.M. Jones, P.G.K. Reiser, C.H. Bryant, S.H. Muggleton, D.B. Kell, and S.G. Oliver, (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature 427*, 247–252.



*This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.*

## Comparing *In Planta* and Microbial Production of Bioproducts

Minliang Yang[1,2*] (minliangyang@lbl.gov), Di Liu[3,4], Nawa Raj Baral[1,2], Chien-Yuan Lin[1,2], Blake A. Simmons[1,2], John M. Gladden[1,2,4], Aymerick Eudes[1,2], Corinne D. Scown[1,2,5], and **Jay Keasling**[1,2,5]

1. Joint BioEnergy Institute, Emeryville, California 94608, USA
2. Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
3. Department of Energy, Agile BioFoundry, Emeryville, California 94608, USA
4. Sandia National Laboratories, Livermore, CA, 94551, USA
5. University of California, Berkeley, California 94720, USA

www.jbei.org

**Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.**

Plants and microbes share common metabolic pathways for producing a range of bioproducts that are potentially foundational to the future bioeconomy. However, *in planta* accumulation and microbial production of bioproducts have never been systematically compared on an economic basis to identify optimal routes of production. Detailed technoeconomic analysis using four biochemicals as exemplar compounds (4-hydroxybenzoic acid (4-HBA), catechol, muconic acid, and 2-pyrone-4,6-dicarboxylic acid (PDC)) is conducted with the highest reported yields and accumulation rates to identify economically-advantaged platforms and calculate break-even targets for plants and microbes. The results indicate that *in planta* mass accumulation ranging from 0.1 dry weight % (dwt%) to 0.3 dwt% achieve costs comparable to microbial routes with 40–55% of maximum theoretical yield. At small volumes and high market prices consistent with specialty chemicals ($20-50/kg), *in planta* accumulation rates in the 0.1 dwt%–0.3 dwt% are sufficient to be cost-competitive. At prices more consistent with commodity chemicals, an order of magnitude increases in accumulation rates for plants and and/or yields nearing theoretical maxima for microbial platforms are needed. This comparative analysis revealed that the accumulation rates of 4-HBA (3.2 dwt%) and PDC (3.0 dwt%) in engineered plants vastly outperform microbial routes, even if microbial platforms were to reach theoretical maximum yields. Their recovery and sale as part of a lignocellulosic biorefinery could enable biofuel prices to be competitive with petroleum. Muconic acid and catechol, in contrast, are currently more attractive when produced microbially. Ultimately both platforms can play an important role in replacing fossil-derived products.

*Funding statement*
*Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

**Isoprenol production in *P. putida* KT2440**

Ian S. Yunus[1,2,*] (ianyunus@lbl.gov), Xi Wang[1,2], Deepanwita Banerjee[1,2], Joonhoon Kim[1,2], Yan Chen[1,2], Jennifer Gin[1,2], Chris J. Petzold[1,2], Paul D. Adams[1,3,5], Aindrila Mukhopadyay[1,2], **Taek Soon Lee**[1,2], Jay D. Keasling[1,2,5,6] (Project PI)

[1] Joint BioEnergy Institute (JBEI), 5885 Hollis St., Emeryville, CA 94608, USA
[2] Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[3] Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[4] Department of Molecular & Cell Biology, University of California, Berkeley, CA 94720, USA
[5] Department of Bioengineering, University of California, Berkeley, CA 94720, USA.
[6] Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720, USA.

**www.jbei.org**

**Project Goals:** Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Isoprenol (3-methyl-3-buten-1-ol) is a candidate for a drop-in biofuel and a precursor for commodity chemicals (e.g., rubber), and more recently it was demonstrated as a precursor for a promising sustainable aviation fuel (SAF) compound DMCO (1,4-dimethylcyclooctane)[1,2]. Isoprenol has been produced in various engineered microorganisms, including *E. coli, C. glutamicum,* and *S. cerevisiae*. Recently, *Pseudomonas putida* KT2440 has gained interest as a promising host for bioproduction of drop-in biofuels as it can utilize carbon sources generated from inexpensive plant biomass. However, the potential of *P. putida* KT2440 for isoprenol production remained unexplored. In this study, we aim to engineer *P. putida* KT2440 for isoprenol production. First, we employed opt-approaches and constrained minimal cut sets (cMCS) to identify gene knockout targets to maximize isoprenol production. Secondly, we established an "IPP-bypass" isoprenol pathway by utilizing a promiscuous mevalonate diphosphate decarboxylase to alleviate the toxicity imposed by an intermediate metabolite (isopentenyl diphosphate, IPP). Next, we further improved the isoprenol titer by optimizing the pathway protein expression. Lastly, we employed targeted proteomics to identify the bottlenecks in the pathway. Altogether, the highest isoprenol production titer of 1.1 g/L was achieved from the engineered *P. putida* KT2440 strain.

**Reference**

[1]Kang, A. *et al. Metab. Eng.* 56, 85–96 (2019)
[2]Baral, N. R. *et al. ACS Sustain. Chem. Eng.* 9, 11872–11882 (2021)

**Funding Statement:** *Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

**Characterization and Stress Response of Monolignol *p*-Hydroxybenzoyltransferase in Poplar**

Yunjun Zhao[1,2]* (yunjunzhao@bnl.gov), Xiao-HongYu,[1,3] Pui-Ying Lam,[4] Kewei Zhang,[1,5] Yuki Tobimatsu,[4] Chang-Jun Liu,[1,2] and **Jay Keasling**[2]

[1]Biology Department, Brookhaven National Laboratory, Upton, NY; [2]Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Present address: Biochemistry and Cell Biology Department, Stony Brook University, Stony Brook, NY; [4]Research Institute for Sustainable Humanosphere, Kyoto University, Gokasho, Japan; [5]Present address: Institute of Plant Genetics and Developmental Biology, Zhejiang Provincial Key Laboratory of Biotechnology on Specialty Economic Plants, College of Chemistry and Life Sciences, Zhejiang Normal University, Jinhua, P. R. China.

www.jbei.org

**Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.**

Lignin in many species including *Populus* spp. is decorated with *p*-hydroxybenzoate. However, the molecular basis for such structural modification remains largely undetermined. Here, we report the identification and characterization of a *Populus* BAHD family acyltransferase that catalyzes monolignol *p*-hydroxybenzoylation, thus controlling the formation of *p*-hydroxybenzoylated lignin structures. With *in vitro* systematic screening and biochemic characterization of *Populus* BAHD family acyltransferases, we reveal that a *Populus* acyltransferase PHBMT1 kinetically preferentially uses *p*-hydroxybenzoyl-CoA to acylate syringyl lignin monomer sinapyl alcohol *in vitro*. Consistently, disrupting *PHBMT1* in *Populus* via CRISPR–Cas9 gene editing nearly completely depletes *p*-hydroxybenzoates of stem lignin; conversely, overexpression of *PHBMT1* enhances stem lignin *p*-hydroxybenzoylation, suggesting that PHBMT1 functions as a prime monolignol *p*-hydroxybenzoyltransferase *in planta*. Altering lignin *p*-hydroxybenzoate modification substantially changes the lignin solvent dissolution rate, indicative of its structural significance on lignin physicochemical properties. *PHBMT1* is highly induced with mechanical stress or gravistimulation along with the tension wood formation. Altering lignin *p*-hydroxybenzoylation changes plant's autotropism and gravitropism behaviors. Identification of monolignol *p*-hydroxybenzoyltransferase offers a valuable tool for tailoring lignin structure and physicochemical properties and for engineering the industrially important platform chemical in woody biomass.

**Publications**

1. Zhao, Y., *et al.* (2021) Monolignol acyltransferase for lignin *p*-hydroxybenzoylation in *Populus*. *Nature plants* 7, 1288-1300

2. Zhao, Y., *et al.* (2021) The Inducible Accumulation of Cell Wall-Bound *p*-Hydroxybenzoates Is Involved in the Regulation of Gravitropic Response of Poplar. *Front Plant Sci* 12, 755576

*Funding statement*: *This research was supported by Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

**Deep Neural Network-Guided Design of Orthogonal Trans-Elements for Plant Synthetic Biology**

Andy Zhou[1,2]* (andyzhou@lbl.gov), Niklas F. C. Hummel[1], Patrick M. Shih[1,2], **Jay Keasling[1,2]**

[1]The Joint Bioenergy Institute, Emeryville CA; [2]University of California, Berkeley CA

https://www.jbei.org

**Project goals:**
Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Synthetic biology tools for integrating complex gene circuits are a necessity for creating high-yielding, deconstructable bioenergy crops. Introducing complex gene circuits and functions into bioenergy crops will require the coordination of multiple cis and trans elements with minimal off-target effects on the plant. Transcription-activator-like effectors (TALEs) offer a useful scaffold for designing targetable synthetic trans-elements for bioenergy crop improvement due to the modular encoding of its DNA-binding site in the protein sequence. We assess the benefits of using different TALEs by targeting orthogonally designed DNA-sequence space and compare its impact on endogenous gene regulatory networks versus using literature-established *trans*-elements such as yeast Gal4. Using machine learning to assist in the identification and design of orthogonal DNA-binding sites, we are able to build synthetic trans-elements and characterize their ability to regulate transgene expression in a targeted manner in model and crop plant systems.

*Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

# Two Shifts in Evolutionary Lability Explain Many Independent Origins of Nitrogen-Fixing Nodulation Symbiosis in A Single Clade of The Plant Tree of Life

Heather R. Kates[1*] (hkates@ufl.edu), Ryan A. Folk[2], Pamela S. Soltis[1,3,4], Robert P. Guralnick[1,5], Brian O'Meara[6], Raphael LaFrance[1], **Matias Kirst**[7], Douglas E. Soltis[1,3,4,5]

[1]Florida Museum of Natural History, University of Florida, Gainesville, FL, United States; [2]Department of Biological Sciences, Mississippi State University, Mississippi State, MS, United States; [3]Department of Biology, University of Florida, Gainesville, FL, United States; [4]Biodiversity Institute, University of Florida, Gainesville, FL, United States; [5]Genetics Institute, University of Florida, Gainesville, FL, United States; [6]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1610, USA; [7]Forest Resources and Conservation, University of Florida, Gainesville, FL, United States

www.nitfix.org

**Project Goals: The collaborative project "Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar" is focused on identifying the genomic novelties that enable the symbiotic relationship between nodulating plants and N-fixing bacteria to support genetically engineering this capability into bioenergy crops. The first aim of this project is a comparative phylogenomic study of the nodulating clade to uncover the genomic novelties that were required for the evolution of these root nodule symbioses. A massively improved phylogenetic framework for the N-fixing clade that includes genetic data for nodulation genes for all species in the phylogeny will provide a robust, revised understanding of the exact ancestral origin of nodulation and the evolution of the predisposition to nodulate and inform a series of experiments comparing close relatives that nodulate or do not nodulate to identify the genes underlying nodulation. To achieve these goals we have completed a cutting edge phylogenetic study with an unprecedented sampling effort and used this to identify the origins of the symbiotic relationship between nodulating plants and N-fixing bacteria as illustrated by the project described below.**

Symbiotic nitrogen fixation via root nodulation (SNFN) is a complex trait that requires coordinated control of multiple plant host and bacterial symbiont genes. Pinpointing the evolutionary origins of SNFN is critical for understanding its genetic basis but is complicated by data limitations and intermittent presence in a single species-rich clade of flowering plants, the nitrogen-fixing clade (NFC). Here we use the largest purpose-built phylogeny for any lineage, along with an enhanced trait database, to reconstruct the evolution of SNFN. We show that shifts among heterogeneous evolutionary rates can explain how a trait can arise many times across a large phylogeny by identifying the evolutionary pathway to SNFN gain. Our analyses suggest a two-step process where an ancestral deep precursor state gives rise to a more labile state from which SNFN was easily gained at certain points in the NFC. Our reconstruction of ancestral

states illustrates how this two-step pathway could have led to 15 independent gains of SNFN, in contrast to single-gain and losses hypotheses. SNFN may be an example of multi-level convergent evolution, thus requiring scaling up of genome-phenome mapping to fully elucidate mechanisms enabling SNFN.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-1916632.

**Title:** Single-cell discovery of nodulation regulators in *Medicago* and evaluation of the functional role of homologs in *Populus*

**Authors:** Daniel Conde,[1] Sanhita Chakraborty,[2] Thomas Irving,[2] Sara Knaack,[3] Lucas Gontijo Silva Maia,[2] Kelly Balmant,[1] Matthew Crook,[4] Christopher Dervinis,[1] Heather Kates,[5] Wendell Pereira,[1] Ryan Folk,[6] Robert Guralnick,[5] Douglas Soltis,[5,7]Pamela Soltis,[5] Jean-Michel Ane,[2] Sushmita Roy,[3,8] **Matias Kirst**,[1]* (mkirst@ufl.edu)

**Institutions:** [1]School of Forest, Fisheries and Geomatic Sciences, University of Florida, Gainesville, FL; [2]Departments of Bacteriology and Agronomy, University of Wisconsin at Madison, Madison, WI; [3]Wisconsin Institute for Discovery, University of Wisconsin at Madison, Madison, WI; [4] Department of Microbiology, Weber State University, Ogden, UT; [5] Florida Museum of Natural History, University of Florida, Gainesville, FL; [6] Department of Biology, Mississippi State University, Mississippi State, MS; [7] Department of Biology, University of Florida, Gainesville, FL; and [8] Department of Biostatistics and Medical Informatics, University of Wisconsin at Madison, Madison, WI

**Website URL:** http://NitFix.org

**Project Goals:** The collaborative project "Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar" is focused on identifying the genomic novelties that enable the symbiotic relationship between nodulating plants and N-fixing bacteria to support genetically engineering this capability into bioenergy crops. As part of this effort, we are identifying new nodulation regulators using single-cell transcriptome analysis of *Medicago* roots treated with rhizobia. In parallel, we are characterizing the function of *Populus* homologs of known nodulation genes, such as the transcription factor Nodule INception (NIN).

Legumes can establish a symbiotic relationship with nitrogen-fixing rhizobia by developing nodules, a modified lateral root organ. Nodule development initiates with anticlinal divisions in a subset of pericycle cells after root exposure to lipochito-oligosaccharides secreted by the bacteria. To discover new regulators of the early differentiation of pericycle cells during nodule formation, we isolated and sequenced the transcriptome of *Medicago truncatula* single nuclei derived from roots, 24 hours after inoculation with rhizobia. Clusters of cells were generated and annotated based on expression enrichment for known marker genes (Fig. 1). Clusters representing all the cell types expected in roots were detected. Next, we re-grouped the pericycle cells at higher resolution and detected two distinct sub-types of cells. Further analysis indicated that one of these clusters involved cells responding to the rhizobia. By following the developmental trajectory of pericycle cells transitioning from a steady to a response-activated state, we identified a homolog of the Arabidopsis STY family proteins as highly differentially expressed. STY proteins are positive regulators of auxin biosynthesis, as regulators of the pericycle activation during nodule initiation. As expected, the RNAi down-regulation of STY resulted in a significant decrease in the number of nodules generated in transgenic *Medicago* plants. While the combined activation of auxin and cytokinin had been previously recognized, several of the specific genetic component activated early in the pericycle differentiation have remained largely unknown. These genetic triggers may be necessary to introduce nodule development in non-legume crops. They are currently being tested for their potential role in inducing the formation of nodule-like structure in *Populus*.

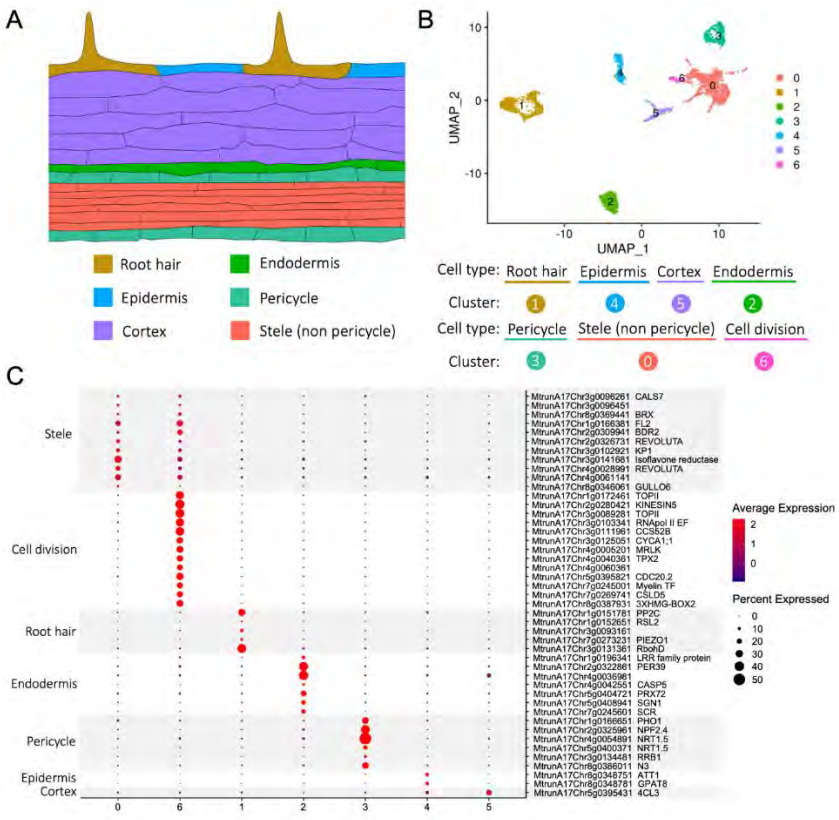



**Figure 1. Single-cell analysis of rhizobia treated *M. truncatula* roots.** (A) Depiction of a longitudinal section of the *M. truncatula* root portion and its cell types. (B) UMAP projection plot showing dimensional reduction of the distribution of 4,000 individual cell transcriptomes. Cell transcriptomes are clustered in seven groups. (C) Expression of known cell-type marker genes across cells reveals the identity of clusters. Dot diameter displays the proportion of cluster cells expressing a given gene; color shows mean expression in cells in that cluster.

In addition to discovering new nodulation regulators, we are exploring the role in *Populus* of genes previously shown to regulate this process in legumes. The transcription factor Nodule INception (NIN) is critical for establishing root nodule symbioses in legumes and other plant groups within the nitrogen-fixing clade (NFC). Plants within the NFC that have lost the ability to develop nodules strongly correlate with a loss of NIN. Interestingly, NIN homologs are present in plants outside the NFC, suggesting its role in other developmental processes. We discovered that *Populus* sp. contains eight copies of NIN falling into three distinct sub-families. Lipo-chitooligosaccharides (LCOs) produced by rhizobia and a wide range of fungi, including mycorrhizal ones, can trigger root nodule symbiosis as well as lateral root formation. We observed that LCOs induced all three members of the PtNIN2 sub-family, with PtNIN2b showing a rapid induction within 15 minutes of treatment. We further observed that PtNIN2b promotes lateral root formation. Constitutive expression of PtNIN2b overcame the inhibition of lateral root development by cytokinin under high nitrate. Finally, we found that the LCO-induced, NIN-dependent lateral root formation is conserved in legumes. Our study suggests that the primitive function of NIN was to modify root architecture in response to symbiotic signals, which later evolved into more specialized functions in root nodule symbiosis.

**Funding Statement:** This work is supported by the Biosystems Design Program from the Biological and Environmental Research (BER) Office of Science at the U.S. Department of Energy (grant #DE-SC0018247).

**Title:** Dynamic change in chromatin accessibility predicts regulators of nodulation in *Medicago truncatula*

**Authors:** Sara A. Knaack[1]†, Daniel Conde[2]†, Sanhita Chakraborty[3], Kelly M. Balmant[2], Thomas B. Irving[3], Lucas Gontijo Silva Maia[3], Paolo M. Triozzi[2], Christopher Dervinis[2], Wendell J. Pereira[2], Junko Maeda[3], Henry W. Schmidt[2], Jean-Michel Ané[3,4], **Matias Kirst**[2,5*], Sushmita Roy[1,6,7*]

**Institutions:** [1] Wisconsin Institute for Discovery, University of Wisconsin, Madison; [2] School of Forest Resources and Conservation, University of Florida; [3] Department of Bacteriology, University of Wisconsin; [4] Department of Agronomy, University of Wisconsin, Madison; [5] Genetics Institute, University of Florida, Gainesville; [6] Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison; [7] Department of Computer Sciences, University of Wisconsin, Madison; † These authors contributed equally to this work

**Website URL:** http://www.nitfix.org

**Project Goals:** Our research goal is to identify host genomic elements involved in the symbiotic relationship between legume roots and nitrogen ($N_2$)-fixing rhizobium bacteria. In order to identify such elements, we measured and analyzed novel RNA-seq and ATAC-seq time-course data obtained from *Medicago truncatula* roots subjected to treatment with the rhizobial symbiotic signals, lipo-chitooligosaccharides (LCO). The gene regulatory network involved in the response to LCOs was predicted by using a novel computational method that defines dynamically transitioning genes and predicts key regulators of these genes. Prioritized regulators and their target genes are now validated experimentally with RNAi experiments. The RNAi results reveal that *Ethylene Insensitive 3 (EIN3)* and *Ethylene Response Factor 1 (ERF1)* are essential for rhizobium-legume symbiosis.

**Abstract Text:** Rhizobia can establish symbiotic associations with legume roots resulting in the formation of root organs called nodules. This symbiosis triggers extensive genome and transcriptome remodeling in the host plant, yet the extent of chromatin changes and impact on gene expression is not well known. We profiled the chromatin accessibility (ATAC-seq) and transcriptome (RNA-seq) dynamics of *M. truncatula* roots treated with rhizobial LCOs over a 24 hr period (0 (control), 15 and 30 min, and 1, 2, 4, 8, and 24 hours).

Using a novel approach, Dynamic Regulatory Module Networks,[1] we predicted gene expression as a function of chromatin accessibility and accessible *cis*-regulatory elements within 10 kbp upstream to 1 kbp downstream of gene transcription start sites. This approach identified the *cis*-regulatory elements and associated transcription factors that most significantly contribute to transcriptomic changes triggered by LCOs. Regulators involved in auxin (IAA4-5, SHY2), ethylene (EIN3, ERF1*)* and abscisic acid (*ABI5*) hormone response, as well as histone and DNA methylation (IBM1), emerged among those most predictive of transcriptome dynamics.

We validated our prioritized set of regulators by knocking down regulators with RNAi and measuring the number of nodules in mutant and wild type plants. The knockdown of *EIN3* and *ERF1* was found to reduce the nodule number in *M. truncatula* roots relative to the empty vector control suggesting that these regulators are important for nodulation. These findings are corroborated by previous results in *L. japonicus*.[2] The molecular mechanisms by which *EIN3* and *ERF1* regulate rhizobium-legume symbiosis remain to be explored. Taken together, our dataset and results provide novel insights into the regulation of gene expression in LCO induced responses in legumes and should be useful to the broad community of researchers interested in mechanisms underlying symbiotic relationships between plants and microbes.

**References/Publications**

1. Alireza Fotuhi Siahpirani, Deborah Chasman, Morten Seirup, Sara Knaack, Rupa Sridharan, Ron Stewart, James Thomson, Sushmita Roy, Dynamic regulatory module networks for inference of cell type specific transcriptional networks. BioRxiv: https://doi.org/10.1101/2020.07.18.210328

2. E. Asamizu, Y. Shimoda, H. Kouchi, S. Tabata, S. Sato, A Positive Regulatory Role for *LjERF1* in the Nodulation Process Is Revealed by Systematic Analysis of Nodule-Associated Transcription Factors of *Lotus* japonicus. *Plant Physiol.* **147**, 2030–2040 (2008).

**Funding Statement:** This work was supported by the Department of Energy Office of Science Biological and Environmental Research (DE-SC0018247). We also acknowledge the support of the Center of High Throughput Computing at UW-Madison to enable computational experiments.

**Title:** Sustainable Production of Acrylic Acid via 3-Hydroxypropionic Acid from Lignocellulosic Biomass

**Authors:** Sarang S. Bhagwat[1,2], Yalin Li[1,3], Yoel R. Cortés-Peña[1,2], Emma C. Brace[1,3], Teresa A. Martin[1,4], Huimin Zhao[1,4], **Jeremy S. Guest**[1,2,3]

**Institutions:** [1]DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI), University of Illinois at Urbana-Champaign, 1206 W. Gregory Drive, Urbana, IL 61801, USA; [2]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 3221 Newmark Civil Engineering Laboratory, 205 N. Mathews Avenue, Urbana, IL 61801, USA; [3]Institute for Sustainability, Energy, and Environment (iSEE), University of Illinois at Urbana-Champaign, 1101 W. Peabody Drive, Urbana, IL 61801, USA; [4]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 215 Roger Adams Laboratory, 600 S. Mathews Avenue, Urbana, IL 61801, USA

**Website URL:** http://pubs.acs.org/doi/10.1021/acssuschemeng.1c05441

**Project Goals:** The goals of this project were (i) to characterize, under uncertainty, the financial viability (via techno-economic analysis, TEA) and environmental impacts (via life cycle assessment, LCA) of biorefineries producing acrylic acid via 3-hydroxypropionic acid (3-HP) from lignocellulosic feedstocks for the current state of technology as well as across potential technological improvements, (ii) to benchmark the performance of these biorefineries against conventional and alternative production processes, and (iii) to set specific technology performance targets to identify and prioritize research needs.

**Abstract Text:** Lignocellulosic biomass is a promising renewable feedstock for the sustainable manufacturing of biofuels and bioproducts. Among emerging bioproducts, 3-HP is of particular interest as a platform chemical to produce commercially significant chemicals such as acrylic acid. In this study, BioSTEAM[1]—an open-source platform—was leveraged to design, simulate, and evaluate (via TEA and LCA) biorefineries producing acrylic acid via fermentation of sugars (glucose and xylose) to 3-HP. At the current state of technology, the biorefinery could produce high-purity (>99.5 wt%) acrylic acid at a minimum product selling price (MPSP) of $\$1.83 \cdot kg^{-1}$ (baseline) with a range of $\$1.72–2.08 \cdot kg^{-1}$ [$5^{th}$–$95^{th}$ percentiles, hereafter shown in brackets], a 100-year global warming potential ($GWP_{100}$) of 3.90 [3.42−4.63] kg $CO_2$-eq$\cdot kg^{-1}$, and a fossil energy consumption (FEC) of 51.4 [43.1−62.1] $MJ \cdot kg^{-1}$. The environmental impacts ($GWP_{100}$ and FEC) for the current state of technology were shown to be significantly lower than those of conventional (fossil-derived) acrylic acid and similar to those of algal glycerol-derived acrylic acid. Advancements in key technological parameters (fermentation yield, titer, and saccharification solids loading) could greatly enhance the biorefinery's performance (MPSP of $\$1.29–1.52 \cdot kg^{-1}$ with ~88% probability of market-competitiveness, $GWP_{100}$ of 3.00 [2.53−3.38] kg $CO_2$-eq$\cdot kg^{-1}$, FEC of 39.9 [31.6−45.1] $MJ \cdot kg^{-1}$). Alternative fermentation regimes (neutral/low-pH fermentation across titer, yield, and productivity combinations) and alternative feedstocks

(first/second generation feedstocks across price and sugar/carbohydrate content) were evaluated to map the sustainability of the biorefinery across the selected design and technology space under uncertainty. Overall, this research highlights the ability of agile TEA-LCA to screen promising biorefinery designs, navigate sustainability tradeoffs, prioritize research needs, and establish a roadmap for the continued development of bioproducts and biofuels.

**References**

(1) Cortés-Peña, Y.; Kumar, D.; Singh, V.; Guest, J. S. BioSTEAM: A Fast and Flexible Platform for the Design, Simulation, and Techno-Economic Analysis of Biorefineries under Uncertainty. *ACS Sustainable Chem. Eng.* **2020**, *8* (8), 3302–3310. https://doi.org/10.1021/acssuschemeng.9b07040.

**Funding Statement:** This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

# Refining Metabolic Engineering Strategies for Hyperaccumulation of Triacylglycerol in Oilcane

Viet Cao[1,2]* (caodangviet@ufl.edu), Baskaran Kannan[1,2], Shelby Korynta[1,2], Sofia Cano[1,2], Hui Liu[1,3], Steve Long[1,4], John Shanklin[1,3], and **Fredy Altpeter[1,2]**

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation, Gainesville FL; [2]University of Florida, Gainesville, FL; [3]Brookhaven National Laboratory, Upton, NY; [4]University of Illinois, Urbana, IL.

## Project Goals:

In this project, we:

1) Explore the field performance of oilcane, a sugarcane that was engineered to produce an abundance of lipids in the form of triacylglycerol which can be converted into biodiesel, biojet fuel, and bioproducts.
2) Refine strategies for multi-gene engineering to elevate lipid yields per land area.

These project goals support CABBI's "plants-as-factories" paradigm, in which biofuels, bioproducts, high-value molecules, and foundation molecules for conversion are synthesized directly in plant stems.

## Abstract:

Sugarcane is an ideal target crop to fuel the emerging bioeconomy. It combines superior biomass production and photosynthetic efficiency with hyperaccumulation of sucrose in its stem, which offers great prospects for diversion to alternative products. We recently reported the generation of oilcane, a sugarcane which has been metabolically engineered for hyperaccumulation of triacylglycerol (TAG) in its vegetative biomass [1,2].

Using constitutive promoters for metabolic engineering approaches has the advantage that analysis can be performed early in plant development. This is beneficial when hyperaccumulation of target products like TAG depend on combined expression and catalytic performance of four or more lipogenic factors to (1) increase fatty acid synthesis, (2) increase triacylglycerol (TAG) synthesis from diacyl-glycerol and acyl-CoA, (3) optimize TAG storage and (4) minimize TAG hydrolysis in vegetative tissues. However, high level constitutive expression of lipogenic factors may also lead to a reduction in field performance, while developmentally regulated or stem specific promoters will slow the Design-Test-Build-Learn (DTBL) cycles, which are required for step changes in lipid accumulation. In contrast, inducible promoters will accelerate the selection of most promising gene variants and synergistic gene combinations for complex multi-gene engineering. Inducible promoters prevent potentially deleterious effects of hyperaccumulated target products on regeneration of transgenic plants from tissue culture or plant vigor and allow us

to explore synergistic effects of alternative gene combinations by expressing one factor under inducible promoter and the other factors under constitutive promoter.

Oilcane plants expressing all lipogenic factors under constitutive promoters were evaluated in replicated and randomized field plots at the University of Florida-IFAS Plant Science Research and Education Unit, near Citra, FL under USDA-APHIS authorization. Data will be presented detailing transgene expression, TAG content, total fatty acid content and biomass yield. Lipid accumulation varied depending on position and maturity of leaves and stems, and reached more than 10% of TAG per leaf dry weight at the time of biomass harvest, averaged over 15 biological replications. The oilcane plants grew vigorously, ratooned successfully and produced 52% of the biomass of the non-engineered sugarcane.

Inducible promoters may allow lipid production at will, at a time when tissue culture or critical stages of plant development are already completed. Since sugarcane tolerates elevated temperatures between 40° and 45°C for an extended period of time, we explored different heat shock promoters (HSP) and the combination of heat inducible and constitutively expressed lipogenic factors to accelerate DTBL cycles. When using HSP promoters to drive the most critical lipogenic factors, we noted an elevated production of transgenic events and rapid establishment of transgenic plants in soil with vigor comparable to non-transgenic sugarcane. A time course experiment evaluating heat induced transgene expression informed a treatment protocol comparing 2, 4 and 8 days of heat exposure with two 4h cycles at 40°C each day. The results showed that sugarcane leaves and stems can be induced for hyperaccumulation of TAG in both leaves and stems within 4 to 8 days of transgene expression. This strategy has great potential to accelerate DTBL cycles in sugarcane.

**References**


1. Parajuli, S, Kannan, B, Karan, R, Sanahuja, G., Liu , H., Garcia-Ruiz, E., Kumar, D., Singh, V., Zhao, H., Long, S., Shanklin, J. and Altpeter, F (2020). Towards oilcane: Engineering hyperaccumulation of triacylglycerol into sugarcane stems. *GCB Bioenergy*. 12: 476– 490. https://doi.org/10.1111/gcbb.12684
2. Zale, J., Jung, J.H., Kim, J.Y., Pathak, B., Karan, R., Liu, H., Chen, X., Wu, H., Candreva, J., Zhai, Z., Shanklin, J. and Altpeter, F. (2016), Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. Plant Biotechnol J, 14: 661-669. https://doi.org/10.1111/pbi.12411



**Funding Statement:**

This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

# High Throughput Screening of Mutant Libraries for Producing Medium Chain Fatty Acids Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Analysis of Microbial Colonies

Kisurb Choe[1], Michael Jindra[2], Pu (Mason) Xue[3], Huimin Zhao[3], Brian Pfleger[2], Jonathan Sweedler[4] (jsweedle@illinois.edu)

[1] Department of Biochemistry, University of Illinois Urbana Champaign, Urbana, IL, USA; [2] Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI, USA; [3] Department of Chemical and Biomolecular Engineering, University of Illinois Urbana Champaign, Urbana, IL, USA; [4] Department of Chemistry, University of Illinois Urbana Champaign, Urbana, IL, USA

**Project Goals:** The creation of a mass spectrometry workflow for high throughput screening of microbial colonies to produce medium chain fatty acids and medium chain fatty alcohols.

## Abstract

Fatty acids and fatty alcohols are value-added compounds used as industrial chemicals and biofuels. There is an unmet need for medium-chain variants of these compounds because of a lack of supply and their greater utility. One issue slowing biocatalyst research to produce enhanced organisms for their production is the lack of high throughput analytical screening methods. We developed a high throughput mass spectrometry (MS) workflow for screening up to 10,000 mutants per day. A custom software suite[1] and a unique sample preparation workflow[2] enable the high throughput workflow. The workflow tested microbial colonies grown on Petri dish instead of liquid culture, greatly simplifying and accelerating the chemical screen for the compounds. Random mutant library of several thioesterases were screened, revealing enzyme variants showing modified substrate specificity towards producing medium-chain fatty acids[3]. The workflow can be applied to screen projects for a wide variety of compounds that are detectable by matrix-assisted laser desorption / ionization MS, and the workflow is easily adaptable by other labs due to the simplicity of the required experimental tools.

## Publications

[1] Choe, K., Xue, P., Zhao, H., & Sweedler, J. V. (2021). maroMS: Image-guided analysis of random objects by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Journal of the American Society for Mass Spectrometry, 32(5), 1180–1188. https://doi.org/10.1021/jasms.1c00013

[2] Choe, K., Sweedler, J. V. Phenotyping by Testing Microbial Colonies Using Image Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. Bio-protocol (*in preparation*)

[3] Jindra, M., Choe, K., Chowdhury, R., Kong, R., Ghaffari, G., Sweedler, J.V., Maranas, C.D.,

Pfleger, B.F. Achieving exclusive chain-length acyl-ACP thioesterase selectivity with a novel colony screening technique. Metabolic engineering (*in preparation*)

**Funding statement**
This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

**Title:** Analyzing Lipid Synthesis and Turnover Using Stable Isotopes

**Authors:** Teresa J. Clark[1]* (tclark@bnl.gov) and **Jörg Schwender**[1,2]

**Institutions:**

[1]Biology Department, Brookhaven National Laboratory, Upton, New York;
[2]DOE Center for Advanced Bioenergy and Bioproducts Innovation

**Project Goals:**

To metabolically phenotype the leaves of wild-type sugarcane and oilcane line 1566 to discern how much assimilated $CO_2$ is diverted into triacylglycerol accumulation, the rate at which lipids are turned over, and the broader metabolic effects from engineering high-oil lines.

**Abstract Text:**

Plant oils for food, animal feed, and industrial applications have traditionally been extracted as triacylglycerols (TAG) from seeds, but there is growing interest in genetically engineering plants to accumulate oil in stems and other vegetative organs. To compare potential engineering strategies, we have been using isotope tracer studies to measure TAG synthesis and turnover in vegetative tissues. Here, we used pulse-chase labeling to compare wildtype and transgenic (line 1566) sugarcane. Young, fully expanded leaves were labeled with $^{13}CO_2$ for 6 hr. After 0, 15, 60, or 180 minutes of a chase period in ambient air, we analyzed abundance and $^{13}C$-labeling of fatty acids obtained from extracted TAG. This revealed that the genotypes had similarly high $^{13}C$-TAG enrichment but differed in which fatty acids were more enriched. The data suggest the rate of TAG synthesis is countered by degradation and that fatty acid content influences at least one of these processes. Next, we will repeat the 6 hr labeling with a longer (24 hr) chase period to better analyze the degradation kinetics.

**References:**

Parajuli et al. (2020). Towards oilcane: Engineering hyperaccumulation of triacylglycerol into sugarcane stems. GCB Bioenergy, 12(7), 476-490.

**Funding Statement:**

This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

**Title:** Automated high-throughput genome editing of TAG-related genes in plants

**Authors:** Jia Dong[1]* (dongjia1@illinois.edu), Deepti Pattabhi[2,] Rafael Anorga Lazo[3] and **Matthew Hudson**

**Institutions:** [1,2,3]University of Illinois at Champaign-Urbana, Urbana, IL

**Project Goals:** This project will closely link the Conversion and Feedstock themes by integrating techniques developed by the Conversion theme, namely high-throughput automation, with a major goal of the Feedstock theme: maximizing TAG output in plants using CRISPR genome editing.

**Abstract Text:** Developing transgenic or genome-edited plants is a time- and labor-intensive process. However, iBioFAB offers a potential solution through high-throughput robotic automation of plant genome editing. In our project, we are developing automated high-throughput processes for CRISPR genome editing of triacylglyceride (TAG) content in three plant protoplast systems: *Nicotiana benthamiana*, Maize B73, and Sorghum TX623. We first established protocols for isolation and transfection of protoplasts from these three plant systems with high efficiency. We then constructed CRISPR deletion vectors targeting two TAG-related and three photosynthesis genes to be used as proof of concepts for automated genome editing in plants. Furthermore, we have programmed the robotic protocols enabling automated protoplast transfection. In the coming months, we will develop robotic protocols for automated protoplast isolation and TAG quantification using fluorescent staining of lipids. Finally, we will employ the full high-throughput system to target a variety of TAG-related genes and determine their individual and combinatorial affects on TAG accumulation. Our work will enable rapid and accurate determination of the effects of specific genome edits in plants, which can later be applied to developing germline-edited plant varieties.

**Funding Statement:** *This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*

**Title:** Cell-type specific analysis reveals the spatial and temporal transcriptome of the Sorghum bicolor stem

**Authors:** Jie Fu[1, 2]* (jief3@illinois.edu), Lye Meng Markelie[3], John Mullet[4], Kankshita Swaminathan[1, 5], and **Amy Marshall-Colon**[1, 2]

**Institutions:** [1]DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI), University of Illinois, Urbana-Champaign, Urbana, [2]Department of Plant Biology, University of Illinois, Champaign-Urbana, Illinois; [3]Environmental Molecular Sciences Laboratory (EMSL), Richland, Washington; [4]Department of Biochemistry & Biophysics, Texas A&M University, College Station, Texas; [5]HudsonAlpha Biotechnology Institute, Huntsville, Alabama

**Project Goals:**

The specific aim of this project is to achieve a comprehensive understanding of sorghum stem biology by constructing molecular atlas on the scale of each specific stem cell type as well as bulk stem tissue by exploring their transcriptome, proteome, and metabolome data. Our goals of this project are:

1. Construct sorghum stem cell-type specific gene regulatory network and advance our knowledge of gene regulatory mechanism and stem function on the scale of individual cell types. To implement such study on the single cell type scale, we develop a customized sweet sorghum cell type separation protocol.

2. Exploit this information to facilitate sorghum molecular engineering to improve plant desirable properties, especially on the aspects of either reducing degradation difficulty of cellulosic biomass, or converting non-structural carbohydrate to more energy-intense substance, like triacylglycerol.

**Abstract Text:**

Sorghum bicolor is currently an important source of food, forage, and feedstock for bioenergy. Due to its advantage conferred by $C_4$ photosynthesis, it will play a more critical role in the foreseeable future where population growth and environmental pressure will demand higher grain production and more sustainable bioenergy. Many studies have explored sorghum genome information and there already exist abundant resources of different sorghum genotype genome sequence data. However, the majority of studies related to gene expression and regulation were conducted on the scale of plant tissue type, and there is no study investigating this information with a finer resolution, like single cell or single cell type. In this study, we focus on stem cell-type specific transcriptome data at vegetative and reproductive stage, which could advance our knowledge of sorghum stem biology on the finer scale over time and space.

Transcriptome data of different stem cell types were achieved by laser capture microdissection (LCM) followed by RNA-Seq. Differentially expressed genes (DEGs) between two growing

stages were identified by *edgeR* (Robinson et al., 2010) pipeline and cell-type specific genes at each growing stage were identified by combining Tau (τ) index (Kryuchkova-Mostacci et al., 2017) and Wilcoxon test. Enriched GO termed were pulled out by *clusterProfiler* (Yu et al., 2012) for genes in each category identified as either DEGs or cell-type specific genes to indicate their specific functions. Finally, gene regulatory network (GRN) for each cell type at each developmental stage was constructed by integrating co-expression strength of cell-type specific genes and their predictive regulatory relationship. The transcriptome data achieved by the protocol developed in this study indicates that different stem cell types had distinct gene expression profiles, which underlay their morphological and physiological differentiation. By comparing two growing stages, we find that most of genes were not differentially expressed over time. These genes may execute basic functions that are indispensable for all the cell types in plant life. Furthermore, enriched GO term for these DEGs indicates that different cell type had both common and unique functions at each growing stage. Tau, combined with Wilcoxon test, successfully identified cell-type specific genes at each growing stage. The enriched functions of these cell type specific genes were consistent with their characteristic physiological functions, which added another layer of confidence for the precision and feasibility of this single cell separation protocol. Gene regulatory network of pith parenchyma is of great interest among all the cell types since this is the stem location where plant stores most of non-structural carbohydrate after anthesis and it focuses most engineering attention from the perspective of either converting cellulosic biomass to bioproducts or transforming carbohydrate to triacylglycerol (TAG) that have more than twofold energy than that of carbohydrate. Transcription factor hubs, which were identified by node outgoing degree number, had a large control over GRN, indicating these genes are potentially most influential regulators in pith parenchyma.

**References/Publications**


1. Kryuchkova-Mostacci, N., & Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. Briefings in bioinformatics, 18(2), 205-214.
2. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26(1), 139-140.
3. Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology, 16(5), 284-287.


**Funding Statement:**


This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

**Title:** Impact of Sugarcane Cultivation on C and $H_2O$ Fluxes in Southeastern United States following Conversion from Grazed Pastures

**Authors:** Gomez-Casanovas, N.[1,2]* (ngomezca@illinois.edu), Blanc-Betes, E.[1,2], Moore, C.[1,2], Bernacchi, C.J.[1,2], Boughton, E.H.[1,3], Yang, W.[1,2], Finzi, A.[1,4], **DeLucia, E.H.**[1,2]

**Institutions:** [1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]ISEE, University of Illinois at Urbana-Champaign, Urbana, IL; [3]Buck Island Ranch, Archbold Biological Station, FL; [4]Boston University, MA

**Project Goals:**

The overarching objective of this research is to improve our understanding of the environmental costs and benefits of pasture conversion to sugarcane, and inform sustainable land use decisions for bioenergy crop cultivation.

**Abstract Text:**

The Southeastern US (SE US) has the capacity to produce almost a third of the 36 billion gallons target established by the Energy Independence and Security Act. The expansion of cane, a subtropical high yielding feedstock, will likely reshape the US bioenergy landscape, although the consequences of this land use change on the environment are highly uncertain particularly as it may displace grazed pastures, a typical landscape of the SE US. Here, we investigated how the conversion of pastures to sugarcane in subtropical Florida impacts carbon (C) and $H_2O$ fluxes using eddy covariance, chamber and biometric methods. Pastures included grazed improved and semi-native pastures, which make up for 53% of agricultural land. Sugarcane exhibited 3-fold and 11-fold higher aboveground productivity relative to improved and semi-native pastures. The land conversion drastically reduced root C allocation from 45% and >95% in improved and semi-native pastures to barely 5% in sugarcane. The conversion shifted belowground C inputs from a root-dominated to a litter-dominated system, and decreased the contribution of root respiration to soil C losses by 5-10%. Given differences in the chemical recalcitrance, aggregation and mineral associated organic matter formation of root- and litter- derived C inputs, establishing sugarcane in grazed pastureland could profoundly alter the formation and persistence of soil organic matter, and long-term soil C storage. Immediately following conversion from pasture, sugarcane was a stronger net source of $CO_2$ than grazed pastures but after first regrowth it became a stronger net $CO_2$ sink (i.e. Net Ecosystem $CO_2$ Exchange). After accounting for C removal (i.e. due to fire, harvest, consumed biomass by grazers), sugarcane was a net C source to the atmosphere (i.e. Net Ecosystem C Balance; +662 and +196 gC $m^{-2}$ $yr^{-1}$ following sugarcane conversion and after first regrowth), the semi-native pasture was a net sink, and the improved pasture was C neutral (-223 and +5 gC $m^{-2}$ $yr^{-1}$ for semi-native and improved pastures for the studied period). These results emphasize the importance of conversion disturbances (i.e. tillage, cultivation, planting) and management practices (i.e. fire) in shaping the C balance of ecosystems. Although sugarcane produced higher yield per water consumed, evapotranspiration was higher in cane and improved pasture than in semi-native pasture, suggesting that the regional water balance implications of land use conversion will depend on the proportion of improved versus semi-native pastures converted to sugarcane. Overall, our results suggest that feedstock optimization and the implementation of management strategies targeting carbon storage are critical to the development of a sustainable bioenergy landscape in SE US,

and highlight the role of bioenergy crops in potentially altering the water cycle to support bioenergy demand.

**Funding Statement:** Funding for this work was provided by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420).

# Leveraging Population Genomics Analysis of *Issatchenkia orientalis* for Engineering a Better Strain for the Production of Lignocellulosic-biomass-based Bioproducts

Ping-Hung Hsieh[1,2], Yusuke Sasaki[2], Jing Ke[2], Zong-Yen Wu[1,2], Zhiying Zhao[2] and **Yasuo Yoshikuni**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Lawrence Berkeley National Laboratory, Berkeley, California

## Project Goals:

*Issatchenkia orientalis* is an ideal chassis for the organic acid production from lignocellulosic biomass because of its stress tolerance to low-pH and lignocellulosic inhibitors. However, the mechanisms underlying its multi-stress tolerance remain elusive. This project aims to identify the genetic variation contributing to the stress tolerance of *I. orientalis* by population genomics analysis and use the knowledge gained to engineer a better strain for the production of lignocellulosic-biomass-based bioproducts.

## Abstract:

*Issatchenkia orientalis* is a non-model ascomycetes yeast with exceptional ability to tolerate extremely low pH, high concentrations of organic acids, and high concentrations of lignocellulosic inhibitors[1-5]. These unique characteristics make *I. orientalis* an attractive chassis for producing organic acids directly from lignocellulosic hydrolysates. Understanding how *I. orientalis* evolved to tolerate multiple stresses may allow engineering of a strain more suitable for industrial use than natural isolates are. We performed a population genomics study of 162 strains collected from various habitats and identified 305,435 single nucleotide polymorphism (SNPs), 16,177 insertions and deletions (InDels), and other genetic variations, including polyploidy and gene copy number variation. The genome-wide association study identified a putative membrane transporter gene associated with the tolerance to hydroxymethylfurfural (HMF) and phenolic inhibitors. We further engineered loss-of-function mutants of this transporter gene from the industrial SD108 strain, and found that the mutant strains are more susceptible to HMF when missing this transporter gene. Therefore, overexpression of this transporter gene is expected to enhance the HMF tolerance in SD108 strain. Moreover, we use machine-learning analysis to determine genetic variants associated with the unfavorable fluconazole resistance. Based on the machine learning analysis and experimental validation, we found that deleting an ABC transporter gene in the new benchmark strain IO21 can generate a desirable strain with decreased fluconazole resistance and increased HMF tolerance. In sum, our study reveals genes involved in HMF and fluconazole stress tolerance and provides ample genomics resources for the communities.

**References/Publications**


1. Xiao, H.; Shao, Z.; Jiang, Y.; Dole, S.; Zhao, H. Exploiting *Issatchenkia Orientalis* SD108 for Succinic Acid Production. *Microb. Cell Fact.* **2014**, *13* (1), 1–11.
2. Mukherjee, V.; Radecka, D.; Aerts, G.; Verstrepen, K. J.; Lievens, B.; Thevelein, J. M. Phenotypic Landscape of Non-Conventional Yeast Species for Different Stress Tolerance Traits Desirable in Bioethanol Fermentation. *Biotechnol. Biofuels* **2017**.
3. Seong, Y. J.; Lee, H. J.; Lee, J. E.; Kim, S.; Lee, D. Y.; Kim, K. H.; Park, Y. C. Physiological and Metabolomic Analysis of *Issatchenkia Orientalis* MTY1 With Multiple Tolerance for Cellulosic Bioethanol Production. Biotechnol. J. **2017.**
4. Park, H. J.; Ko, H.; Jeong, H.; Lee, S. H.; Ko, H.; Bae, J.; Sung, B. H.; Han, J.; Sohn, J. Draft Genome Sequence of a Multistress-Tolerant Yeast, *Pichia Kudriavzevii* NG7. *Genome Announc.* **2018**.
5. vanRijswijck, I. M. H.; Derks, M. F. L.; Abee, T.; deRidder, D.; Smid, E. J. Genome Sequences of *Cyberlindnera Fabianii* 65, *Pichia Kudriavzevii* 129, and *Saccharomyces Cerevisiae* 131 Isolated from Fermented Masau Fruits in Zimbabwe. *Genome Announc.* **2017**.



**Funding Statement:**

This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

# A New Bioenergy Model that Simulates the Impacts of Plant-Microbial Interactions, Soil Carbon Protection, and Mechanistic Tillage on Soil Carbon Cycling

Stephanie M. Juice[1,2*] (stephanie.juice@mail.wvu.edu), Christopher A. Walter[2], Kara E. Allen[3], Danielle M. Berardi[1,4], Tara W. Hudiburg[1,4], Benjamin N. Sulman[5], **Edward R. Brzostek**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Department of Biology, West Virginia University, Morgantown; [3]Manaaki Whenua Landcare Research, Lincoln, New Zealand; [4]Department of Forest, Rangeland and Fire Sciences, University of Idaho, Moscow; [5]Climate Change Science Institute and Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

https://cabbi.bio/cabbi-models/

**Project Goals:**

**We developed a new model of soil carbon (C) in bioenergy systems that incorporates plant-microbe interactions, microbial physiology, and mechanisms of stable soil C creation and loss to test hypotheses on the response of soil C to bioenergy feedstock traits and environmental conditions. We then used the model to address the following questions:**

**(1) How do commonly used model representations of tillage vary in soil C projections as compared to a novel tillage mechanism that increases microbial access to soil C?**

**(2) How do differences between annual and perennial bioenergy feedstocks in their integrated plant-microbial traits and tillage regimes impact model estimates of soil C?**

**(3) To what extent do these model differences in the representation of feedstock traits alter the response of soil C estimates to warming temperatures?**

**(4) Does the outcome of altering plant traits (i.e., increased rhizosphere C exudation) on model estimates of soil C vary across different feedstocks?**

Advancing our predictive understanding of bioenergy systems is critical to designing decision tools that can inform which feedstock to plant, where to plant it, and how to manage its production to provide both energy and ecosystem carbon (C) benefits. Here, we lay the foundation for that advancement by integrating recent developments in the science of belowground processes in shaping the C cycle into a new bioenergy model, FUN-BioCROP (Fixation and Uptake of Nitrogen-Bioenergy Carbon, Rhizosphere, Organisms, and Protection). We show that FUN-BioCROP can approximate the historical trajectory of soil C dynamics as natural ecosystems were successively converted into intensive agriculture and bioenergy systems. This ability relies in part on a novel tillage representation that mechanistically models tillage as a process that increases microbial access to C. Importantly, the impacts of tillage and feedstock choice extend into FUN-BioCROP simulations of warming responses with no-till

perennial feedstocks, miscanthus and switchgrass, having more C that is unprotected and susceptible to warming than tilled annual feedstocks like corn-corn-soybean (Fig. 1). However, this susceptibility to warming is balanced by a greater potential for increases in belowground C allocation to enhance soil C stocks in perennial systems. Collectively, our model results highlight the importance of belowground processes in evaluating the ecosystem C benefits of bioenergy production.

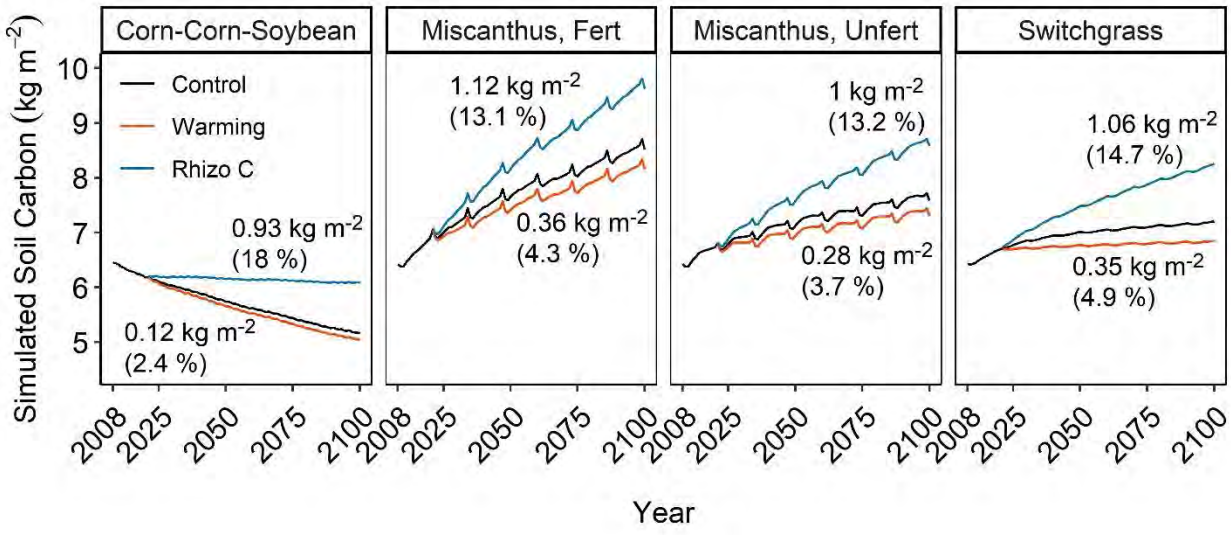



**Figure 1.** Total soil carbon (C) simulated by FUN-BioCROP at the University of Illinois at Urbana-Champaign Energy Farm under three sets of experimental conditions: control (black lines; same data as shown in Figure 3), 5 °C soil warming (red lines), and increased rhizosphere C flux ("Rhizo C;" blue lines; augmented mycorrhizal C flux and non-mycorrhizal root C exudates by 5% of average daily NPP). Values on the figure indicate the difference between each experimental (i.e., warming, or increased rhizosphere C) final C pool value as compared to control, with the percent difference in parentheses. E.g., increased rhizosphere C in corn-corn-soybean led to 18% higher SOC at the end of the simulation. Fert- fertilized, Unfert- unfertilized, FUN-BioCROP- Fixation and Uptake of Nitrogen-Bioenergy Carbon, Rhizosphere, Organisms, and Protection.

***Funding Statement:*** *This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*

**Title:** Investigation of xylose metabolism in *Rhodosporidium toruloides* using a modular cloning kit (RT-EZ)

**Authors:** Hyungi Koh[1,2*](hyungi@illinois.edu), Payman Tohidifar[2,3], Hyunjoon Oh[4], Christopher Rao[1,2,3], **Yong-su Jin**[1,2,4]

**Institutions:** [1] Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL; [2]DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, Urbana, IL; [3]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL; [4]Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, IL

**Website URL:** https://cabbi.bio/

**Project Goals:** Understanding xylose metabolism in *Rhodosporidium toruloides*, and enhancing xylose utilization efficiency using a modular cloning kit

**Abstract Text:**

*Rhodosporidium toruloides* is an oleaginous yeast strain, which has the ability to utilize diverse kinds of carbon sources including glucose, xylose, fructose, xylitol, arabitol, galactitol and etc. [1]. However, its preference is highly biased on glucose, showing slower consumption rate and growth rate (50% ~) under different carbon source such as xylose. In addition, significant amount of arabitol is generated as a byproduct when xylose is provided as a sole carbon source, which is re-consumed after depletion of xylose in the media. Hence, this study aimed to improve the xylose utilization efficiency of *R. toruloides*, which would allow better growth of the strain in hydrolysates containing xylose as well. First of all, in order to reduce the troubles often caused by high GC-contents (62.01%) of *R. toruloides* during vector construction, we developed a toolkit (RT-EZ) composed of genetic modules that allows hierarchical assembly based on Golden Gate cloning prior to actual cloning [2]. The toolkit contains uni-/bi-directional promoters, antibiotic markers, terminators, 2A linkers and etc. Using the toolkit, heterologous *Xyl1*, *Xyl2*, and *Xyl3* genes from *Pichia stipitis,* encoding xylose reductase (*p*XR), xylitol dehydrogenase (*p*XDH), and xylulose kinase (*p*XK), respectively, were expressed in *R. toruloides*. Although expression of *p*XR or *p*XDH did not show much difference compared to the parental strain, expression of *p*XK resulted in improved growth rate with doubled maximum growth rate. Surprisingly, the accumulation of arabitol, which reached up to 15 g/L in parental strain, was completely removed from the *p*XK expressing strain as well, implying improved overall sugar consumption rate. The low expression level of endogenous XK in *R. toruloides* with xylose [3] suggests malfunction, mis-annotation or lack of the XK activity, explaining the distinct effect of *p*XK expression. We believe that our findings will help understanding the xylose metabolism in *R. toruloides*, and also the developed toolkit to boost the research on this high-potential oleaginous yeast strain.

**References/Publications**


1. *Yaegashi* et al, *Biotechnol. Biofuels*, 2017, 10:241
2. *Weber* et al, *Plos one*, 2011, 18:6(2)
3. *Jagtap* et al, *Appl Microbiol Biotechnol*. 2021, 105(19)



**Funding Statement:** *This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*

# Metabolic Engineering of Triacylglycerols in Vegetative Tissues of Sorghum

Kiyoul Park, Truyen Quach, Hyojin Kim, Tara J. Nazarenus, Shirley Sato, Edgar B. Cahoon, and Tom E. Clemente

United States Department of Energy Center for Advanced Bioproducts and Bioenergy Innovation, University of Nebraska, Lincoln, Nebraska, USA

Triacylglycerols (TAG) provide for energy-dense carbon storage of fatty acids (FA) in plants. TAG (or “oil”) naturally accumulates to high levels in seeds to provide energy and carbon-skeletons needed for germination. By contrast, vegetative tissues typically only store TAG transiently, in part to protect cells from cytotoxic free FA. Here, we have redesigned the biomass crop sorghum (*Sorghum bicolor*) to accumulate TAG to high levels in vegetative tissues, including leaves and stalks. The goal of this work is to provide a TAG-derived co-product value stream in a Midwest-adapted, drought tolerant crop for advanced biofuels and bioproducts. Benefiting from our ability to transform sorghum with multi-transgene constructs, we introduced gene combinations to: (1) push photosynthetic carbon into FA biosynthesis using a Wrinkled1 transcription factor transgene; (2) pull of FA into TAG storage using a novel diacylglycerol acyltransferase (DGAT) transgene; and (3) protect TAG from turnover using an oleosin oil body coat protein transgene. After extensive primary transformants, we identified $T_0$ lead events that accumulate TAG in leaves to ≥2% DW, compared to ≤0.1% DW of TAG found in non-transformed plants in a Tx430 grain sorghum background. We advanced the $T_1$ generation of the lead events in the University of Nebraska-Lincoln Biotech Field Facility in Mead, NE. For field-grown Tx430 lines, TAG levels of ≥6.6% DW were detected in leaves after flowering (99 days after sowing) and 2.5% DW in stalks after harvest (127 days after sowing). No obvious differences in biomass of engineered lines were observed. We also generated events in the Ramada sweet sorghum background that has greater biomass and sucrose stem content compared to Tx430. These events accumulated ≥5.6% DW TAG in leaves under greenhouse

conditions versus 0.1% DW TAG in leaves of non-transformed plants. We have initiated collaborative fluxomics and root-soil metagenomics experiments. A field trial for summer of 2022 is also planned to more fully assess the traits. Our results to date show the feasibility of sorghum as a vegetative oil production platform. Information gained from our current lines will guide iterative development of the vegetative TAG trait to address biosynthetic and catabolic factors that limit further increases in oil production and the balance of photosynthetic carbon partitioning to support growth and storage.

**Title:** Using Cross-Scale Data to Constrain an Agro-Ecosystem Model to Produce Estimates of Miscanthus Production at a Field-Scale

**Authors:** Bryan Petersen[1*] (bryan20@iastate.edu), Chongya Jiang,[2] Kaiyu Guan,[3] Emily Heaton[4], and **Andy VanLoocke**[1]

**Institutions:** [1]Iowa State University, Ames, IA; [2]University of Illinois at Urbana Champaign, Agroecosystem Sustainability Center, Institute for Sustainability, Energy, and Environment, Urbana, IL; [3]University of Illinois at Urbana Champaign, College of Agricultural, Consumer and Environmental Sciences, Urbana, IL; and University of Illinois at Urbana Champaign, Urbana, IA[4]

**Website URL:** NA

**Project Goals:** Evaluate a research plot-calibrated agro-ecosystem model performance at predicting harvest yields at commercial miscanthus sites and determine what environmental and management factors most likely explain model error.

**Abstract Text:** There is evidence from both field experiments and ecosystem models that the perennial C4 bioenergy crop *Miscanthus x giganteus* (miscanthus) can provide more energy per sown area, reduce net carbon dioxide emissions, and lower nitrate leaching rates than corn ethanol in certain locations. Typically field experiments are at scales too small while ecosystem models predictions are often at scales too coarse to represent commercial operations. Because the primary goal of producers is to generate profit, it is important to identify locations where miscanthus can potentially compete economically with other management options while not compromising environmental goals. This requires resolving genotype, environment, and management interactions and representing those interactions in a framework capable of resolving variations at scales useful to miscanthus producers. Therefore, the objective of this project is to utilize plant- and field-scale measurements from several commercial and experimental plots of miscanthus in Iowa and Illinois to inform a mechanistic agro-ecosystem model (Agro-IBIS) and ultimately improve miscanthus yield predictions. Growing season leaf-level gas exchange, biomass, and leaf area data were collected from a miscanthus field experiment in central Iowa and were used to calibrate Agro-IBIS. The model was also calibrated and validated using SLOPE, a novel satellite-based product that estimates gross primary productivity, and observed biomass yields to calibrate and validate Agro-IBIS's predictions at the commercial miscanthus sites. Utilizing a length-mass allometric model for miscanthus, we corrected Agro-IBIS yield predictions based on the estimated cutting height of the chopper. Preliminary results indicate that using subfield-scale measurements in conjunction with field-scale measurements while also accounting for different harvest managements will improve Agro-IBIS's miscanthus yield predictions at a field-scale. This work will allow for more strategic placement of this bioenergy crop candidate that would optimize profit for the producer while reducing nitrate leaching and sequestering carbon.

**References/Publications**


1. Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: the potential of miscanthus. Global Change Biology, 14, 2000–2014.
2. Stampfl PF, Clifton-Brown JC, Jones MB (2007) European-wide GIS-based modeling system for quantifying the feedstock from Miscanthus and the potential contribution to renewable energy targets. Global Change Biology, 13, 2283–2295.
3. VanLoocke A, Bernacchi CJ, Twine TE (2010) The impacts of Miscanthus x giganteus production on the Midwest US hydrologic cycle. Global Change Biology – Bioenergy, 2, 180–191.
4. Zhu, X.-G., Long, S. P., & Ort, D. R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Current Opinion in Biotechnology*, *19*(2), 153–159. https://doi.org/10.1016/j.copbio.2008.02.004



**Funding Statement:** This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

**Title:** Optimization of Solvent-Free Enzymatic Esterification of Free Fatty Acids using Taguchi Design of Experiments

**Authors:** Ramkrishna Singh,[1]* (rs25888@illinois.edu), Bruce S. Dien,[2] and **Vijay Singh**[1]

**Institutions:** [1]University of Illinois at Urbana Champaign, Urbana-Champaign, Illinois; [2]USDA, Agricultural Research Service, National Center for Agricultural Utilization Research, Peoria, Illinois

**Project Goals:** Esterification of free fatty acids (oleic acid as model) with glycerol using immobilized lipase for deacidification of oil for improved biodiesel production.

**Abstract Text:** The preferred process of biodiesel production converts triglycerides into their ester derivatives (biodiesel). Free fatty acids affect the efficiency and yield of the process. In this work, immobilized lipase was used for esterification of free fatty acid into glycerides to decrease acidity of oil, making the oil more suitable for biodiesel production. Several process parameters and their levels were optimized using a Taguchi design of experiments to achieve maximum conversion. The reusability of the enzyme was evaluated to allow for reduction in cost of the enzyme in the process.

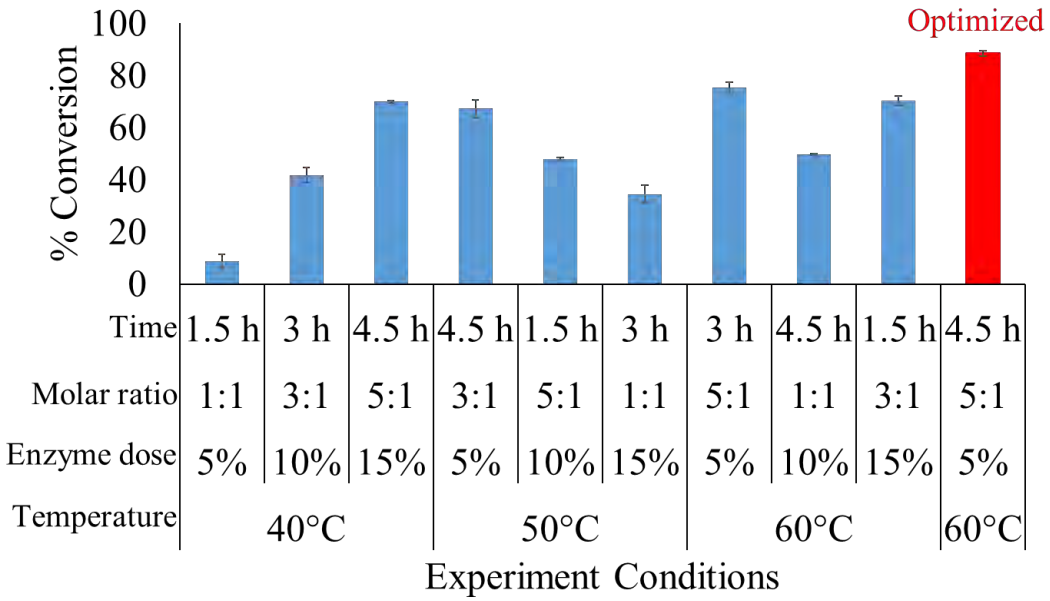
Biodiesel production using alkali catalyzed transesterification process is the preferred commercial process because it gives a high yield and requires lower capital and operating costs [1]. However, the process efficiency decreases in the presence of free fatty acids [2]. We have observed excess free fatty acids in oil recovered from engineered lipid producing crops developed by the Center for Advanced Bioenergy and Bioproducts Innovation (Urbana, IL). Hence, pre-processing of lipids for deacidification would be required for this new source of lipids. The conventional deacidification process requires high capital investment, generates waste, and causes loss of triglycerides and other phytoconstituents [3]. Therefore, a greener process of oil deacidification is desirable, which minimizes constituent loss and improves biodiesel yield.

In this work, immobilized lipase (*Candida antarctica* lipase b) was used for the esterification of oleic acid with glycerol. Oleic acid was chosen as it is the dominant fatty acid of lipids in vegetative tissue. Several parameters including temperature (40, 50 and 60°C), enzyme dose (5, 10, 15 wt%), molar ratio of oleic acid to glycerol (1:1, 1:3, 1:5), and reaction time (1.5, 3 and 4.5 h) were evaluated. The effect of process parameters and their levels were investigated using Taguchi design of experiments to maximize oleic acid conversion from a feasible number of experimental runs. Water produced *in-situ* during esterification reaction can decrease enzyme activity and thereby the product yield. Therefore, improvement of conversion by adding a molecular sieve for *in-situ* removal of water was tested. The reusability of the enzyme was evaluated under optimal conditions for maximum conversion.

The best experimental conditions (60°C, 5 wt%, 5:1 and 3 h) allowed 75.23 ± 2.19% conversion of oleic acid to esters. Using molecular sieves (type 3 Å) to remove produced water increased the

conversion to 86.73 ± 1.09% (15% increase). However, the Taguchi design predicted optimal parameters were 60°C, 5 wt%, 5:1 and 4.5 h. By applying these conditions, conversion of oleic acid was increased to 88.5 ± 1.11% and, furthermore, this was achieved without the added expense of a molecular sieves. The immobilized enzyme could be reused up to seven times with only a 10% decrease in conversion.

Thus, the developed enzymatic esterification process provides a greener alternative for oil deacidification. In addition, it will allow the valorization of glycerol which is produced abundantly in the biodiesel process.




**References/Publications**


[1] A.A. Apostolakou, I.K. Kookos, C. Marazioti, K.C. Angelopoulos, Techno-economic analysis of a biodiesel production process from vegetable oils, Fuel Process. Technol. 90 (2009) 1023–1031. https://doi.org/10.1016/J.FUPROC.2009.04.017.

[2] I.M. Rizwanul Fattah, H.C. Ong, T.M.I. Mahlia, M. Mofijur, A.S. Silitonga, S.M. Ashrafur Rahman, A. Ahmad, State of the Art of Catalysts for Biodiesel Production, Front. Energy Res. 8 (2020) 101. https://doi.org/10.3389/FENRG.2020.00101/BIBTEX.

[3] U.S. Pal, R.K. Patra, N.R. Sahoo, C.K. Bakhara, M.K. Panda, Effect of refining on quality and composition of sunflower oil, J. Food Sci. Technol. 52 (2015) 4613. https://doi.org/10.1007/S13197-014-1461-0.



**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018420.

Dr Bruce Dien is supported by the U.S. Department of Agriculture, Agricultural Research Service. USDA is an equal opportunity provider and employer. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Departments of Energy or Agriculture. USDA is an equal opportunity provider and employer.

**Metabolic Engineering of *Issatchenkia orientalis* for succinic acid production**

Vinh Tran, Yihui Shen, Joshua Rabinowitz, Huimin Zhao

The nonconventional yeast *Issatchenkia orientalis* is an industrially relevant platform microorganism for production of organic acids thanks to its unusual ability to grow in highly acidic conditions and tolerance to high concentrations of several organic acids. Previously, introduction of a reductive TCA pathway into the yeast enabled succinic acid production at titer of 11.6 g/L in shake flask fermentations. Here we describe further metabolic engineering efforts to enhance the production of succinic acid in *I. orientalis*. Further genetic optimizations, including deletion of by-product pathways and heterologous expression of a dicarboxylic acid transporter, allowed 24.1 g/L of succinic acid to be produced from 50 g/L of glucose in shake flask fermentation. Nevertheless, pyruvic acid accumulated at titer of 19.8 g/L because of the lack of cytosolic NADH, which hindered further production of succinic acid. To increase cytosolic NADH availability, glycerol was used as co-substrate. Fermentation using 50 g/L of glucose and 20 g/L of glycerol allowed conversion of the accumulated pyruvic acid to succinic acid and attained a titer of 38.6 g/L of succinic acid. Deletion of a dicarboxylic acid importer and overexpression of the glycerol consumption pathway further improved the titer of succinic acid to 42.1 g/L. In fed-batch fermentation in which pH was set at 3, the best strain could produce succinic acid at titer of 103 g/L, yield of 0.56 g/g glucose equivalent, and productivity of 0.43 g/L/h from pure glucose and glycerol and at titer of 120 g/L and productivity of 0.56 g/L/h from sugarcane juice and glycerol. This is the highest titer achieved in minimal media at low pH 3 to date from a recombinant organism. Overall, this work shows that *I. orientalis* can serve as a potential industrial platform for production of succinic acid.

**Metabolic Engineering *Yarrowia lipolytica* to Produce 3-acetyl-1,2-diacyl-sn-glycerol**

Qiang Yan[1,2*] (qyan32@wisc.edu), William Cordell,[2] and **Brian Pfleger**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI), Urbana, IL;
[2]Department of Chemical and Biological Engineering, University of Wisconsin at Madison, Madison

**Project Goals: The goal is to showcase a combination of metabolic engineering strategy to produce high-level 3-acetyl-1,2-diacyl-sn-glycerol. *Yarrowia lipolytica* QY06 strain devoid lipid droplet formation can serve as a platform strain to study lipid droplet bud-off mechanism and isolate high activity acTAG synthase.**

The triacylglycerols 3-acetyl-1,2-diacyl-sn-glycerol (acTAG) have many potential industrial applications such as engine lubricant oil, emulsifiers, food coatings, and plasticizers. As an oleaginous yeast, *Yarrowia lipolytica's* high flux toward native TAG (lcTAG) synthesis serves as an excellent chassis for production of acTAG. The acTAG are unusual triacylglycerols (TAG) with an acetyl group at the sn-3 position instead of the typical long-chain acyl group. Compared to regular TAG, the acetyl group of acTAG confers useful physical and chemical properties such as reduced kinematic viscosity and lower melting points.

The main challenge here is how to rewire *Y. lipolytica's* native metabolism from lcTAGs to acTAGs. To accomplish this, the presented work provides a two-fold strategy: removing the competing pathways of lcTAG synthesis and identifying highly active acTAG synthases. In the first strategy, we generated a multi-knockout *Y. lipolytica* strain QY06 that deleted three acyltransferases, two lipases, and one dehydrogenase. This engineered strain is incapable of producing lcTAG and forming lipid droplets due to removal of three TAG synthase genes. In the second strategy, we evaluated the activities of 30 bioprospected acTAG synthase homologs in an engineered *Y. lipolytica*. An enzyme homolog from *Euonymus alatus* showed higher activity of acTAG synthesis. When evaluating the acTAG titer as function of lipid droplet content, it was observed that acTAG titer is proportional to the lipid droplet content (e.g. lcTAG titer), indicating acTAG localization in lipid droplet may help alleviating acTAG toxicity to cells by pulling it away from the cytosol and storing in the lipid droplet. Finally, we reported an engineered strain is capable of producing 12 g/L acTAG from glucose in a fed-batch fermentation experiment, and 4 g/L acTAG from sugarcane juice in shake flask experiments. We showcase that combination of metabolic engineering strategy such as blocking competing pathway, pulling flux towards acTAG synthesis, and alleviating toxicity by compartmentalization, to achieve high level production of acTAG.

**References**

1. Durrett, T.P., McClosky, D.D., Tumaney, D.D., Elzinga, D.A., Ohlrogge, J., and Pollard, M. (2010) A distinct DGAT sn-3 acetyltransferase activity that synthesizes unusual, reduced-viscosity oils in Euonymus and transgenic seeds. *PNAS* 107(20): 9464-9

**Funding statement:** This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

# Responses of Total (DNA) and Metabolically Active (RNA) Microbial Communities in *Miscanthus* x *giganteus* Cultivated Soil

**Authors:** Jihoon Yang[1]* (jhyang@iastate.edu), Jaejin Lee,[1] Emily Heaton,[2] and **Adina Howe**[1]

**Institutions:** [1]Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa, [2]Center for Advanced Bioenergy and Bioproducts Innovation, Urbana, Illinois

**Project Goals: *Miscanthus* x *giganteus* is one of the most efficient bioenergy crops and is supported by an active soil microbiome. Previous studies of the role of this microbiome in supplying nitrogen to miscanthus have produced inconsistent results. In this study, we hypothesize that this may be due to methods used to identify the active membership of the soil microbial community. We compared RNA-based to DNA-based soil microbial community structure analysis to improve the identification of metabolically active microbes. We also identified the impacts of management, such as stand age and nitrogen fertilization, on microbial community membership. These efforts provide insight into the best methods to study plant-soil-microbial interactions and the role of microbes in miscanthus production.**

## Introduction

*Miscanthus x giganteus* is a promising high-yielding bioenergy crop to meet growing bioenergy demands with little fertilizer compared to other bioenergy crops. Although plant-soil-microbe interactions were known to affect the productivity of *M.* x *giganteus* at various fertilization rates, previously performed characterization of the microbial community has been limited to DNA-based analysis of potentially active microbial membership. Therefore, there were limitations to identifying metabolically active microbial membership that plays a role in the nitrogen cycle. In this study, we compare DNA and RNA approaches to expand our understanding of how soil microbiome can impact the sustainable production of *M.* x *giganteus*.

## Research approach

Two-, three-, and four-year-old *M.* x *giganteus* soil samples (n=271) from replicated blocks receiving 0, 224, and 448 kg $ha^{-1}$ N were collected in 2018. Paired DNA and RNA extractions were performed using MagAttract PowerMicrobiome DNA/RNA EP kit (Qiagen, USA). The 16S rRNA gene amplicon sequencing of both extracted DNA and RNA was performed on an Illumina Miseq platform (Argonne National Laboratory). The DADA2 package in R and the RDP classifier were used for taxonomic identification of observed ASV (amplicon sequence variants). Permutational multivariate analysis of variance (PERMANOVA) was used for statistical comparison based on Bray-Curtis dissimilarity matrix.

## Results

DNA and RNA microbial communities were significantly different, although similar numbers of microbial membership were detected. The profiles of the dominant microbial membership within

DNA and RNA showed significant differences in the proportions of *Actinobacteria*, *Firmicutes*, and *Proteobacteria*. There was a seasonal response to nitrogen fertilization only in the RNA microbial communities, and this difference was associated with nitrogen-cycling bacteria with a relative abundance 7-fold higher in RNA than in DNA. Among them, genes associated with denitrifying bacteria are significantly abundant in RNA, suggesting they can be underestimated with DNA-only approaches.

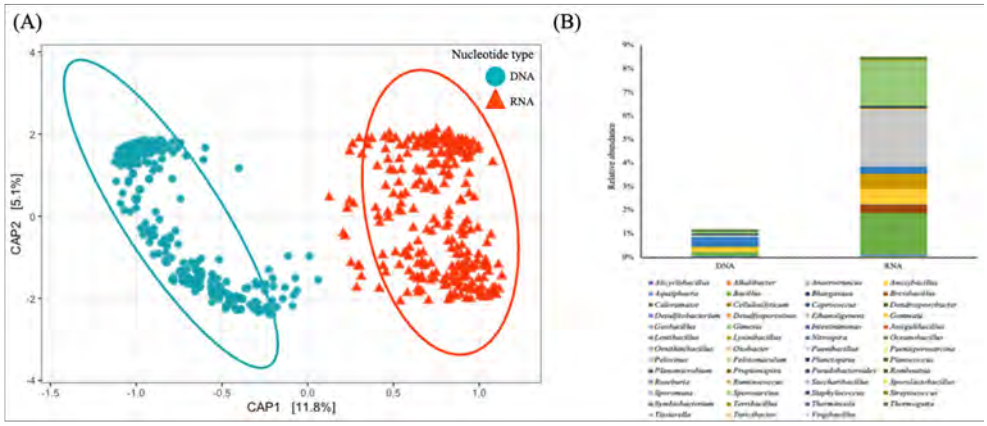



Figure 1. Difference between DNA and RNA microbial communities in *M.* x *giganteus* cultivated soil for (A) microbial membership and (B) abundance of nitrogen cycle associated bacteria.

**Summary**

We found that DNA and RNA-based methods for characterizing the microbial response to management changes showed different results. The RNA-based method appears to capture better the response of microbial membership known to be associated with nitrogen cycling. Increasing numbers of microbial ecology studies are identifying the environment or gradient for which the microbial community is changing. Future work needs to highlight which taxa or function is changing, and our results indicate that RNA-based SSU characterization can be a resource.

**References/Publications**

1. Heaton et al. Meeting US biofuel goals with less land: the potential of Miscanthus. Glob Chang Biol. 14, 2000-2014 (2008).
2. Ma et al. The impact of stand age and fertilization on the soil microbiome of Miscanthus x giganteus. Phytobiomes J. 689 (2020).
3. Yang et al. Response of total (DNA) and metabolically active (RNA) microbial communities in Miscanthus x giganteus cultivated soil to different nitrogen fertilization rates. Microbiol. Spectr. In revision

**Funding Statement:** *This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*

**Title:** Gene Duplication and Single Nucleotide Polymorphisms (SNPs) via Adaptive Laboratory Evolution (ALE) of Engineered *Yarrowia lipolytica* Enabled the Efficient Utilization of Sugars in Lignocellulosic Hydrolysate

**Authors:** Sangdo Yook[1,2], Deewan Anshu[1,3], Payman Tohidifar[1,3], Christopher V. Rao[1,3],and Yong-Su Jin[1,2,*]

**Institutions:** [1]Carl Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA; [2]Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA; [3]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

**Project Goals:**

1. To enable rapid and efficient utilization of xylose by *Y. lipolytica*
2. To understand the limiting factors in xylose metabolism by *Y. lipolytica* and to identify critical genes capable of alleviating the limiting factors.
3. To enable the ample production of cytosolic acetyl-CoA and achieve the production of acetyl-CoA derived products from lignocellulosic hydrolysates.

## Abstract

*Y. lipolytica* has received extensive attention for converting cellulosic hydrolysates due to its potential to produce acetyl-CoA-derived products, such as TAL, lipids, and polyketides [1]. However, *Y. lipolytica* cannot metabolize xylose, the second abundant sugar in cellulosic hydrolysates. Therefore, metabolic engineering to confer xylose metabolism into *Y. lipolytica* is necessary. While *Y. lipolytica* harbors the endogenous xylose utilizing enzymes, namely xylose reductase (ylXR), xylitol dehydrogenase (ylXDH), and xylulokinase (ylXK), the activities of those enzymes are not active enough to facilitate xylose metabolism in *Y. lipolytica* [2]. Therefore, previous researchers introduced heterologous oxidoreductase and isomerase pathways to enable xylose utilization in *Y. lipolytica* but conducted adaptive laboratory evolution to improve lipid accumulation [3]. As such, the causality of the identified mutations on xylose fermentation has not been proved yet. In this study, we introduced a heterologous XR, XDH, and XK from *Pichia stipitis* into *Y. lipolytica* PO1f, and the resulting transformant was evolved under xylose conditions. Evolved strains capable of consuming xylose rapidly were isolated, and their genome sequences were determined to identify genetic perturbations related to the improved xylose utilization phenotypes. Based on the read-depth analysis, RT-PCR, and enzymatic analysis, we confirmed that rapid xylose assimilation by the evolved strain was enabled by the amplification of heterologous XR, XDH, and XK. Several genetic variants in YALI0B12100g

coding for GTPase-activating protein, YALI0B18282g coding for cAMP-independent regulatory protein, YALI0F05346g coding for cutinase gene palindrome-binding protein, YALI0F32065g coding for CCR4 transcriptional subunit, YALI0E18117g coding for E3 ubiquitin-protein ligase, YALIE18700g coding for CTP_transf_like domain-containing protein, YALI0E23474g coding for cytochrome P450 were also identified. Among the mutations, we confirmed the positive effects of YALIB12100g S409F on xylose assimilation via Cas9-based genome editing. When lignocellulosic hydrolysates were used, the lipid titer of the evolved strains was 2.3-times higher (0.83 g/L → 1.94 g/L) than a control strain (before the ALE). Overall, this study provides a better understanding of xylose metabolism in *Y. lipolytica* and beneficial mutations to achieve economic conversion from cellulosic sugar to value-added products.

**References/Publications**


1. Lee, J. W., Yook, S., Koh, H., Rao, C. V., & Jin, Y. S. (2021). Engineering xylose metabolism in yeasts to produce biofuels and chemicals. Current Opinion in Biotechnology, 67, 15–25. https://doi.org/10.1016/j.copbio.2020.10.012
2. Ledesma-Amaro, R., Lazar, Z., Rakicka, M., Guo, Z., Fouchard, F., Coq, A. M. C. L., & Nicaud, J. M. (2016). Metabolic engineering of *Yarrowia lipolytica* to produce chemicals and fuels from xylose. Metabolic Engineering, 38, 115–124. https://doi.org/10.1016/j.ymben.2016.07.001
3. Rodriguez, G. M., Hussain, M. S., Gambill, L., Gao, D., Yaguchi, A., & Blenner, M. (2016). Engineering xylose utilization in *Yarrowia lipolytica* by understanding its cryptic xylose pathway. Biotechnology for Biofuels, 9(1). https://doi.org/10.1186/s13068-016-0562-6



**Funding Statement:** *This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*

**Title:** Engineering Cyclopropane Fatty Acid Accumulation in Plant Vegetative Tissues

**Authors:** Xiao-Hong Yu, [1,2,3] * (xhyu@bnl.gov), Yuanheng Cai,[2,3] Moni Qiande,[4] Fredy Altpeter,[1,4] and **John Shanklin**[1,2]

**Institutions:** [1]Center for Advanced Bioenergy and Bioproducts Innovation; [2]Biology Department, Brookhaven National Laboratory, Upton, NY; and [3]University of Stony Brook, Stony Brook, NY and [4]University of Florida - IFAS, Gainesville, FL

**Project Goals:** Our overall goal is to create and build DNA designs for the synthesis of specialty fatty acids (cyclopropane, hydroxy, and variant unsaturated fatty acids) in sugarcane.

**Abstract Text:** Cyclopropane fatty acids (CPAs) contain a three-carbon ring, and can be used as industrial feedstocks for coatings, lubricants, plastics, and cosmetics. Our previous research showed that expression of CPA synthase in plant seeds results in low levels of CPA accumulation and we identified metabolic bottlenecks in acyl transfer between the site of synthesis (phospholipids) and deposition in TAG. Additional acyltransferases from *Sterculia* seeds increased CPA accumulation in storage oil in transgenic seeds[1,2]. To build DNA designs for synthesis of CPAs in sugarcane, we first established a rapid screening system using transient expression in *Nicotiana benthamiana* leaves for evaluating the effects of additional genes on CPA accumulation in vegetative tissues. We succeeded in accumulating about 1% CPA in *N. benth* leaves upon transient coexpression of *Escherichia coli* cyclopropane fatty acid synthase (CPS) in combination with WRI1, Oleosin and diacylglycerol acyltransferases (DGAT)2. The use of either single or double 35S promoter had no effect on CPA yield, and addition of an intron to EcCPS reduced CPA accumulation in *benth* leaves. Using the *N. benth* transient expression system, we evaluated the effects of different acyltransferases on leaf CPA production, including *Sterculia foetid*a lysophosphatidic acid acyltransferase (LPAT), *Litchi chinensis* and *Mus musculus* DGAT2, and phosphatidylcholine:diacyl glycerol cholinephosphotransferase. Indeed, coexpression of acyltransferases from plants that naturally accumulate CPA increased CPA production in leaves, with SfLPAT showing the most significant effect. Proof-of-principle constructs that increased CPA in *N. benth* were transferred to sugarcane. The effective gene combination for CPA accumulation in *N. benth* gave the strongest effect in sugarcane yielding one transgenic line that accumulates approximately 1% CPA. This is the first demonstration of accumulation of CPA, a specialty fatty acid in vegetative tissues of oilcane.

**References/Publications**

1. Yu XH, Prakash RR, Sweet M, Shanklin J (2014) Coexpressing Escherichia coli cyclopropane synthase with Sterculia foetida Lysophosphatidic acid acyltransferase enhances cyclopropane fatty acid accumulation. Plant Physiol 164: 455-465

2. Yu XH, Cahoon RE, Horn PJ, Shi H, Prakash RR, Cai Y, Hearney M, Chapman KD, Cahoon EB, Schwender J, Shanklin J (2018) Identification of bottlenecks in the accumulation of cyclic fatty acids in camelina seed oil. Plant Biotechnol J 16: 926-938

**Funding Statement:** This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

**Title:** Determining Profit-Maximizing Dynamic Mix of Nitrogen Rate and Stand Age in Miscanthus and Switchgrass Production

**Authors:** Na Zhang[1,2]*(nazhang2@illinois.edu), Bijay P. Sharma,[1,3] and **Madhu Khanna**[1,2,3]

**Institutions:** [1]DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, Urbana, IL; [2]Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, Urbana, IL; [3]Institute for Sustainability, Energy, and Environment, University of Illinois at Urbana-Champaign, Urbana, IL

**Project Goals:** The purpose of our research is to determine the profit-maximizing N application rate at different stand ages and the optimal stand age for field replanting in miscanthus and switchgrass (Alamo, Blackwell and Cave-In-Rock) production in the rainfed region of the United States. Profit-maximizing N application rate over stand ages and optimal stand age for field replanting are critical management decisions faced by a landowner, but there have been limited pieces of evidence in the literature. A few previous studies have examined the profit-maximizing N rates using short-term field-level data. A major limitation of these studies is that they estimate the profit-maximizing N by calculating net returns of constant N treatment rates applied annually over the years in the field trials. However, recent studies show that N requirements change at different maturity stages as yield responsiveness to N changes (Sharma et al., 2021). We extend this literature with estimating the dynamic profit-optimal N application rate and the profit-optimal lifespan of producing these energy crops. Results of our research can guide the profitable production of these two perennial energy crops by improving N rate recommendations and optimal replanting times.

**Abstract Text:** The benefits of producing perennial energy crops depend on market prices, production costs, and crop yields which may rely on climatic conditions, soil productivity, and management practices, especially nitrogen (N) fertilizer. Profitable production of these perennial energy crops requires understanding the yield response to management practices, especially the application of N fertilizer and the interaction with stand age (Sharma et al., 2021). We firstly propose a dynamic optimization framework where the landowner's objective is to maximize the net present value (NPV) of a stream of future benefits from producing perennial energy crops instead of conventional crops by choosing the optimal N application rate for different stand ages and the optimal stand age for field replanting. We use a generalized yield response function including N input, stand age and other factors such as weather and soil productivity for both miscanthus and switchgrass. Theoretically, we derive the optimal N application rate for each stand age given market prices and a set of parameters of the yield response function. Furthermore, we derive the optimal stand age that maximizes the NPV for field replanting. Next, we conduct numerical simulations in 2287 counties in the rainfed region (east of the 100th meridian) of the United States under six assumption scenarios: (1) baseline scenario (biomass price \$70 $Mg^{-1}$ , discount rate 4%); (2) higher biomass price (\$90 $Mg^{-1}$), (3) higher biomass price (\$110 $Mg^{-1}$); (4) higher Discount Rate (10%); (5) 50% increase in N price; (6) 50%

increase in land rent. In the numerical model, we apply the yield response function parameters from Sharma et al. (2021) and use weather, soil productivity, and production costs data at county level.

We find that the profit-maximizing N application rate varies over stand ages for miscanthus. Under the baseline scenario, on average, the profit-maximizing N application rate for miscanthus increases continuously from 10.81 $kg^{-1}$ $ha^{-1}$ at stand age 2 to 165.76 $kg^{-1}$ $ha^{-1}$ at stand age 11, then slightly declines at older stand ages. In contrast, the profit-maximizing N rate for switchgrass maintains relatively stable (around 150 $kg^{-1}$ $ha^{-1}$ $yr^{-1}$ ) after establishment. There are only marginal differences in the profit-maximizing N application rate among these three cultivars(Alamo, Blackwell and Cave-In-Rock). Moreover, at younger stand ages, miscanthus requires a lower profit-maximizing N application rate than switchgrass, but the yield of miscanthus is higher than that of switchgrass. For middle-aged stands, switchgrass reaches almost the same high yield level as miscanthus while maintaining the same profit-maximizing N application rate as it is at younger stand ages. The profit-maximizing N application rate of these perennial energy crops over stand age is sensitive to the prices of biomass and N fertilizer, but it is not sensitive to discount rate and the land rent. Furthermore, on average, the optimal replanting stand age of miscanthus is 13 years, achieving the maximum NPV of \$2065 $ha^{-1}$ (equivalent ANPV is \$203 $ha^{-1}$ $yr^{-1}$). For switchgrass, on average, the optimal replanting stand age is 9 years for Alamo and Blackwell, and 8 years for Cave-In-Rock. The corresponding maximum NPV of producing Cave-in-Rock, Alamo, and Blackwell is \$1164 $ha^{-1}$ $yr^{-1}$, \$1560 $ha^{-1}$ $yr^{-1}$, and \$2303 $ha^{-1}$ $yr^{-1}$ ( equivalent ANPV is \$162 $ha^{-1}$ $yr^{-1}$, \$181 $ha^{-1}$ $yr^{-1}$, and \$274 $ha^{-1}$ $yr^{-1}$) averaged at county level, respectively. In biophysical models, the common assumptions are that the productive life span is 15-20 years for miscanthus and 10-15 years for switchgrass, which is longer than the optimal lifespan estimated from the economic perspective. Estimates of profit-maximizing N application rate, predicted yield, and ANPV of these two perennial energy crops also vary across geographic regions. These results imply that recommending the optimal N application rate over stand ages and across regions will help improve the economic benefits of producing perennial energy crops.

**References/Publications**


1. B.P. Sharma, N. Zhang, D.K. Lee, E.A. Heaton, E.H. DeLucia, E.J. Sacks, I.B. Kantola, N.N. Boersma, S.P. Long, T.B. Voigt, and M. Khanna (2021). Responsiveness of Miscanthus and Switchgrass Yields to Stand Age and Nitrogen Fertilization: A Meta-regression Analysis. *Working Paper*.



**Funding Statement:** This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

**Title:** Genomics of Winter-hardiness and Yield in Diverse Miscanthus Germplasm

**Authors:** Xuying Zheng[1]* (xuyingz2@illinois.edu), and **Erik J. Sacks**[1]

**Institutions:** [1]University of Illinois at Urbana-Champaign, Urbana-Champaign

**Project Goals:** Develop genomic selection models to improve the efficiency of breeding Miscanthus for adaptation to temperate environments and for biomass yield.

**Abstract Text:** The current major Miscanthus cultivar, M. ×giganteus 'Illinois', is not sufficiently winter-hardy in the Midwest U.S., which limits its productivity. Because M. ×giganteus is an interspecific hybrid, selecting winter-hardy accessions of its parental species, M. sinensis and M. sacchariflorus, will be essential for breeding new M. ×giganteus cultivars with greater hardiness. Miscanthus is most susceptible to winterkill and damage during the first winter after planting. In this study, M. sinensis and M. sacchariflorus accessions from diverse genetic and geographic backgrounds were evaluated for first-overwintering ability and yield performance. In an initial experiment, 330 half-sib families representing three M. sinensis genetic groups (North, Central, and South Japan; 55, 117, 158 half-sib families respectively) were evaluated in Urbana, IL for first winter overwintering ability in spring 2020 (planted in spring 2019), and yield data were collected in autumn 2020 and are currently being collected for the 2021 season. First-winter overwintering ability was relatively high for the M. sinensis North and Central Japan genetic groups (81% and 82%, respectively) but low for the South Japan group (49%). Substantial variation among families within genetic groups was observed for overwintering ability and yield, which along with sequence data from JGI will facilitate development of genomic selection models for this important trait. In a subsequent experiment planted in spring 2020 at Urbana, IL, we assessed first-winter overwintering ability for diversity panels consisting of 294 M. sacchariflorus accessions and 236 M. sinensis accessions, representing the geographic ranges and all known genetic groups of these species (Clark et al., 2014; Clark et al., 2019). First-year overwintering ability of both species varied among genetic groups, which will allow us to more efficiently target our Miscanthus breeding efforts to different geographic regions in the U.S.

**References**

1. Clark LV, Brummer JE, Głowacka K, Hall MC, Heo K, Peng J, Yamada T, Yoo JH, Yu CY, Zhao H, Long SP, Sacks EJ. A footprint of past climate change on the diversity and population structure of Miscanthus sinensis. Ann Bot. 2014 Jul;114(1):97-107. doi: 10.1093/aob/mcu084. Epub 2014 Jun 10. PMID: 24918203; PMCID: PMC4071102.

2. Clark LV, Jin X, Petersen KK, Anzoua KG, Bagmet L, Chebukin P, Deuter M, Dzyubenko E, Dzyubenko N, Heo K, Johnson DA, Jørgensen U, Kjeldsen JB, Nagano H, Peng J, Sabitov A, Yamada T, Yoo JH, Yu CY, Long SP, Sacks EJ, Population structure of Miscanthus sacchariflorus reveals two major polyploidization events, tetraploid-mediated unidirectional introgression from diploid M. sinensis, and diversity centred around the

Yellow Sea, Annals of Botany, Volume 124, Issue 4, 13 September 2019, Pages 731–748, https://doi.org/10.1093/aob/mcy161

**Funding Statement:** This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

**Title:** Developing, Understanding, and Harnessing Modular Carbon/Nitrogen-Fixing Tripartite Microbial Consortia for Versatile Production of Biofuel and Platform Chemicals

**Authors:** David Carruthers,[1] Andrew Allman,[1] Maciek Antoniewicz,[1] Sujit Datta,[2] Jagroop Pandhal,[3] and **Xiaoxia "Nina" Lin**[1]* (ninalin@umich.edu)

**Institutions:** [1]University of Michigan, Ann Arbor, MI; [2]Princeton University, Princeton, NJ; [3]University of Sheffield, UK

**Project Goals:** The overall goal of this project is to design, construct, analyze and optimize a synthetic microbial consortium system consisting of three closely interacting members: a $CO_2$-fixing photosynthetic specialist, a $N_2$-fixing specialist, and a third specialist that can convert organic carbon and nitrogen generated by the first two specialists to synthesize a desired product. By integrating complimentary expertise from multiple research labs at three institutions, we are pursuing three specific objectives: i) Develop tripartite microbial consortia for carbon/nitrogen fixation and production of bio-molecules with various nitrogen/carbon ratios; ii) Investigate molecular and cellular mechanisms governing the tripartite consortia via omics study and predictive modeling; and iii) Explore alternative spatial configurations and develop scalable design principles.

**Abstract:**

Microbial communities are ubiquitous in nature, exhibiting incredibly versatile metabolic capabilities and remarkable robustness. Inspired by these synergistic microbial ecosystems, rationally designed synthetic microbial consortia is emerging as a new paradigm for bioprocessing and offers tremendous potential for solving some of the biggest challenges our society faces. In this project, we focus on a tripartite consortium consisting of a $CO_2$-fixing photosynthetic specialist, a $N_2$-fixing specialist, and a third specialist that can convert organic carbon and nitrogen generated by the first two specialists to synthesize a desired product. In addition to $CO_2$ fixation, a noteworthy feature of this design is the elimination of the requirement for nitrogen fertilizer, which has been produced through ammonia synthesis using the Haber-Bosch process and accounts for an estimated 2% of global energy expenditure. We aim to develop a modular and flexible model system capable of producing diverse bio-molecules (varying C:N ratio) as advanced biofuel or platform chemicals, to dissect this complex ecosystem using a spectrum of cutting-edge systems approaches, and to ultimately derive scalable and broadly applicable design principles for maximizing the system performance.

Our first prototype tripartite consortium employs genetically modified strains of photosynthetic cyanobacterium *Synechococcus elongatus*and and nitrogen-fixing bacterium *Azotobacter vinelandii*, that secrete sucrose and ammonia respectively, to form a symbiotic foundation for supporting a third producer member. We demonstrated supported growth for a range of producer strain candidates, including a sucrose-metabolizing *Escherichia coli* K-12 derivative strain, *Corynebacterium glutamicum*, and *Bacillus subtilis*, using a multi-chamber bioreactor system

under continuous culture conditions. We have also developed an ODE-based mathematical model to capture essential molecular, cellular, ecological and process characteristics of the experimental system.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022136.

Ten *Pseudomonas* spp. from Svalbard Active Layer are using amino acids for energy

Raegan Paul[1], Katie Sipes[1], Aubrey K. Fine[1], Sean M. Schaeffer[1], Tatiana Vishnivetskaya[1], Karen G. Lloyd[1]

University of Tennessee, Knoxville, USA[1]

The goal of this project is to study the microbial decomposition within active layer permafrost in Ny Alesund, Svalbard.

The *in situ* population of microbes in the Ny Ålesund, Svalbard permafrost active layer are mostly uncharacterized. It is unknown how microbes survive in this nutrient limited seasonally thawed and frozen soil. Our goal was to isolate representatives of the microbial community to infer their environmental interactions. Complementary techniques were performed by culturing, sequencing, and testing isolates for temperature sensitivity of enzyme activities for carbon and nitrogen hydrolytic enzymes. Active layer soil was rendered on R2A agar and isolates were whole genome sequenced with Illumina MiSeq. The annotated genomes contain metabolic pathways for carbon and sulfur assimilation and $CO_2$ emission. Soil enzyme activities suggest that these organisms are accustomed to carbon and nitrogen starvation. Leucine aminopeptidase activities up to 200 nmol/hour signify that these isolates gather carbon and nitrogen compounds from amino acid degradation. The low temperature enzymatic activities for N-acetyl-β-D-glucosaminidase and xylosidase showed cold adaptation of *Pseudomonas spp.* These metabolic pathways combined with the low temperature enzymatic activities indicate that these organisms acclimate to harsh conditions in the active layer and show their potential for quick adaptation to increasing temperatures in the Arctic. These findings suggest that increasing temperatures associated with climate change could fundamentally shift microbial activity and substrate availability in these systems in short term. However, the microbial population may be able to adapt to these environmental changes long term.

US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program (DE-SC0020369)

Metagenome-assembled genomes from active layer in Ny Ålesund, Svalbard (79°N) show population dynamics related to seasonal thawing and soil depth

Katie Sipes, Michael Tomaino, Raegan Paul, Renxing Liang, Sarahi Garcia, Julia Boike, Tullis C. Onstott, Tatiana A. Vishnivetskaya, Karen G. Lloyd

**ABSTRACT:**

The active layer of permafrost in Ny Ålesund, Svalbard (79°N), is increasing in thickness at a rate of ~one vertical centimeter per year in most locations around the Bayelva River in the Leirhaugen glacier moraine. Microbes in the active layers may drive organic carbon degradation and greenhouse gas production, creating a positive feedback on climate change. However, the microbial metabolisms relating to the environmental geochemistry and population dynamics have not been well characterized. We tested if organismal abundance was related to site or stratigraphy by analyzing metagenome assembled genomes (MAGs) from 56 highly resolved intervals of five active layer permafrost sites. Two sites were collected in the winter time (April 2018) and three were collected in the summer (September 2019). The MAGs' read recruitment to metagenome libraries showed an *Actinobacteria* population that was dominant when the soils were thawed (September 2019) and when the surface soil were covered in snow (April 2018). Inversely, another group of *Actinobacteria* dominated the deeper depths while the soils were frozen and were nearly absent during the thawed/summer season. We show that microbial populations in the Svalbard active layer shift by depth and season, maintaining populations of continuously metabolically active microbes at all depths, year-round.

**Title:** High Enzyme Stability May Enable Slow Microbial Life in the Subsurface

**Authors: Andrew D. Steen**,[1,2*] (asteen1@utk.edu), Taylor M. Royalty[1], Jenna M Schmidt,[2] Karen G. Lloyd[1]

**Institutions:** [1]University of Tennessee, Knoxville, Department of Microbiology; [2]University of Tennessee, Knoxville, Department of Earth and Planetary Sciences

**Project Goals:** To test the hypothesis that the enzymes of subsurface microbes are unusually stable

**Abstract Text:** Subsurface microbes live extraordinarily slow lives, with timescales of biomass turnover on the order of decades or longer. Because living microbes must have functional enzymes, and enzyme biomass typically comprises the majority of microbial biomass, this implies that enzyme lifetimes must scale with biomass turnover times, and therefore that the enzymes of subsurface microbes are likely to be extraordinarily stable. We present measurements of extracellular enzyme activities from subsurface sediments near Denmark, combined with a simple theoretical model of the minimum feasible turnover time for extracellular enzymes, which suggest the minimum possible turnover time for extracellular peptidases in those sediments is at least 7.5 months. This is orders of magnitude slower than typical enzyme lifetimes. We also present preliminary analyses of the structures of those extracellular enzymes, which bear structural features that confer excess stability. In surface environments, there is little apparent selective pressure on enzymes to be more stable than required to maximize their activity, since biomass is turned over rapidly in those environments. In the subsurface, selective pressure to produce stabler enzymes may lead to novel biochemical mechanisms for enzyme stability that are not present in more frequently-studied, fast-growing microbes.

**Funding Statement:** *Example for grants* - This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), Genomic Science Program (DE-SC0020369).

Comparative metagenomics of Arctic landscapes.

Tatiana Vishnivetskaya[1,2]* (tvishniv@utk.edu), Xiaofen Wu[1], Wyatt A. Cyr[1], Katie Sipes[1], Andrew Steen[1], Andrey A. Abramov[2], Elizaveta Rivkina[2], **Karen G. Lloyd**[1]

[1]University of Tennessee, Knoxville Tennessee 37996 USA
[2]Institute of Physicochemical and Biological Problems of Soil Science, Pushchino, Russia

With increasing availability of metagenome datasets from different locations in Arctic, we have a unique opportunity to compare the diversity and functionality of microbial communities to better predict microbial responses to climate changes. Samples from the permafrost active layer were collected from three geographic locations, Bayelva, Svalbard (78.92062 N, 11.84158 E), Axel Heiberg Island (AHI), Canada (79.415 N, 90.75833 W), and Bykovsky Peninsula, Russia (71.783167 N, 129.41086 E) and were processed for the total genomic DNA extraction followed by metagenome sequencing. Samples from Bayelva (n=3) and AHI (n=12) were collected in early spring when ground was frozen using portable drill, while samples from Bykovsky (n=5) were collected in late summer from open pit. The amount of DNA was low for Bykovsky (10.7-14.5 ng/g soil) and Bayelva (13.3 ng/g soil), while DNA content was significantly higher for AHI and showed decrease from 8.5 µg/g in upper 5 cm layer to 2.6 µg/g at 35 cm depth to 0.9 µg/g at 65 cm. Bayelva and AHI represent mineral soil with a dominance of pebbles and coarse sands, while the soil in Bykovsky is loam with trace of humus. Analysis of metagenomes showed differences in microbial diversity depending on depth. Integrated analysis of metagenomes showed that upper 5-7 cm layer (n=5) is dominated by Alphaproteobacteria (23.9% vs 12.0%) and Acidobacteria (2.9% vs 0.9%), while Actinobacteria (27.5% vs 20.6%), Bacteroidetes (1.3% vs 0.5%), Clostridiales (4.1% vs 2.4%), Thermotogales (0.2% vs 0.4%) and methanogenic Archaea (1.7% vs 0.9%) were more abundant in deeper layers (n=16). Alphaproteobacteria were represented mostly by aerobic and microaerophilic methanotrophic, methylotrophic, and nitrogen-fixing bacteria. Bacteria identified in metagenomes of the upper active layer were able to metabolize a wide range of organic carbon sources, degrade complex aromatic compounds, obtain energy through photosynthetic reactions or form symbiotic relation with plants during nitrogen fixation. This analysis indicates similarities in microbial community between the low organic carbon mineral soils collected from geographically different locations.

US Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program (DE-SC0020369)

**Enhancing Vegetative Oil Content through Optimized Lipogenic Factors**

Sanket Anaokar[1]*(anaokars@bnl.gov), Yuanheng Cai[2], Yingqi Cai[1], John Shanklin[1, 2], and **Stephen P. Long[3]**

[1]Brookhaven National Laboratory, Upton, NY; [2]Stony Brook University, Stony Brook, NY; [3]University of Illinois at Urbana-Champaign, Urbana, IL

https://rogue.illinois.edu/

## Project Goals

The main objective of ROGUE (Renewable Oil Generated with Ultra-productive Energycane) project is to engineer the two most productive American crops—energycane and *Miscanthus*—to produce a sustainable supply of biodiesel, biojet fuel, and bioproducts.

1. Increasing oil accumulation and targeting this to the mature stem
2. Increasing photosynthetic efficiency to power oil synthesis
3. Multi-gene construct transformation of energycane and Miscanthus
4. Field testing, processing and techno-economic analysis

## Abstract

Most of the plant biomass is present in vegetative tissues and strategies to accumulate triacylglycerol (TAG, aka oil) in these tissues provides an attractive target for increasing oil yields relative to conventional oilseed crops. Generally, plants accumulate less than 0.1 % DW of oil in vegetative tissues, so, our goal is to use a multi-gene approach employing optimized components to increase oil biosynthesis and decrease oil degradation. Our strategies to enhance oil accumulation align with the proven 'push-pull-protect' model for fatty acid (FA) synthesis.

We targeted the Homomeric Acc1(HoAcc1) to the plastid to serve as a "push" factor. Accase represents the first committed step in FA synthesis. The native plastid-located heteromeric Accase is subject to multiple distinct forms of negative feedback regulation *in vivo*. Our previous transient expression results showed that targeting AtHoAcc1 to the plastids with an N-terminally fused plastid transit peptide and GFP tag can potentially circumvent this negative regulation, enhance FA synthesis, and promote TAG accumulation in tobacco leaves by increasing the malonyl-CoA supply. We stably transformed the AtHoAcc1 into Arabidopsis. Our preliminary analysis suggested that lines transformed with plastidially targeted AtHoAcc1 had larger seed size and higher seed weight without affecting the seed gemination and establishment. Moreover, they also showed higher vegetative TAG accumulation in 14-day old seedlings, demonstrating the effectiveness of using plastidial targeted HoAcc1 as a "push" factor in stably transformed plants to increase vegetative TAG accumulation.

The Pull factor used was diacylglycerol acyl transferase (DGAT), that catalyzes the final conversion of DAG to TAG in the ER. This can efficiently channel fatty acyl chains towards the synthesis of neutral lipid for storage. Overexpression of plant DGATs have been shown to increase TAG production. We identified a mammalian DGAT2 which was very efficient in producing TAG in tobacco transient overexpression assays. Tobacco leaves with DGAT2 overexpression driven by a senescence inducible promoter resulted in more than a 6-fold increase in TAG accumulation relative to controls.

The protect factor was an engineered Sesame oleosin (Ole1), that masks the oil bodies to protect the TAG from hydrolysis by lipases. Previous published research suggested that Ole1 could be degraded by ubiquitin conjugation to lysines, cysteines, serines, threonines, or the N-terminus of a target peptide. To inhibit ubiquitin ligation, we choose to replace all the lysine residues with arginine (KR) and delete the cysteine residues (CysDel) from a *Sesame* oleosin peptide to create an efficient Ole1 variant, Ole1_CysDel_KR. Transient expression of this variant in tobacco leaves indicated that CysDel_KR modifications resulted in an increase in TAG content compared to a Cys_Ole1 control. As the substitutions and deletions were applied to all lysine and cysteine residues, respectively, we predicted some of these changes could negatively impact the oil accumulation. Therefore, we performed a dropout experiment reverting each of these modifications to pinpoint those that contribute most to increased TAG accumulation. By combining the set of mutations that increased TAG accumulation we created Ole1_5 Mod. Our data from tobacco transient assays demonstrated that Ole1_5 Mod resulted in the most significant increase in TAG content. Based on these findings, we stably transformed the oleosin variants into Arabidopsis. Results from the stably transformed Arabidopsis lines are consistent with the data from transiently transformed tobacco in which Ole1_CysDel_KR and Ole1_5 Mod lines in combination with mammalian DGAT2 produced higher TAG and total FA content compared to CysOle1 with mammalian DGAT2.

In summary, our strategy was first to individually optimize the Push, Pull and Protect factors described above. We subsequently combined the optimized factors to test their ability to promote plant vegetative TAG accumulation. Our results show that the effects of each optimized factor are additive. When the three factors are transiently expressed in *N. benthamiana* leaves, TAG accumulation increased more than 18-fold relative to controls, reaching a final TAG content of ~4% (w/w) over a three-day period. The results from transient *N. benthamiana* expression were reproduced in stably transgenic Arabidopsis lines.

**Funding Statement**

The work of Sanket Anaokar, Yuanheng Cai and John Shanklin on Oleosin and ACCase was supported by the Office of Biological and Environmental Research (Award Number DE-SC0018254) and FWP 2021-BNL-BI-173. The work of Yingqi Cai and John Shanklin on the mammalian DGAT was supported by the Center for Advanced Bioenergy and Bioproducts Innovation (Award Number DE-SC0018420).

**Sustainable Improvement of C4 Photosynthesis in Bioenergy Grasses**

Kher Xing Chan[1]* (cindyckx@illinois.edu), Prasenjit Saha[1], Baskaran Kannan[2], Thaibinhduong Nguyen[2], Amit Kumar[1], Fredy Altpeter[2], and **Stephen P. Long[1]**

[1]University of Illinois at Urbana-Champaign, Urbana, IL; [2]University of Florida, Gainesville, FL

https://rogue.illinois.edu/

## Project Goals

The main objective of ROGUE (Renewable Oil Generated with Ultra-productive Energycane) project is to engineer the two most productive American crops—energycane and *Miscanthus*—to produce a sustainable supply of biodiesel, biojet fuel, and bioproducts.

1. Increasing oil accumulation and targeting this to the mature stem
2. Increasing photosynthetic efficiency to power oil synthesis
3. Multi-gene construct transformation of energycane and Miscanthus
4. Field testing, processing and techno-economic analysis

## Abstract

Bioenergy feedstock, such as energycane (*Saccharum* spp.) and miscanthus (*Miscanthus* × *giganteus*), are some of the NADP-ME type C4 perennial grasses adapted to grow in marginal croplands which could be engineered to improve their photosynthetic efficiency to increase yield (Mitchell et al. 2016). These grasses can then be grown in marginal croplands to avoid the need to utilize more land for agriculture (Wang et al. 2021, Mitchell et al. 2016). Regeneration of phospho*enol*pyruvate (PEP) in the NADP-ME type of C4 photosynthesis is limited by the activity of pyruvate orthophosphate dikinase (PPDK) and rubisco (Long et al., 2013, Wang et al., 2008, Naidu et al., 2003). In order reduce the bottleneck effect of PPDK, we hypothesized that increased expression of PPDK in bioenergy crops could improve C4 photosynthesis. Research also shows that photosynthetic efficiency could be improved under fluctuating light when a faster photoprotection response time is observed by overexpressing violaxanthin de-opoxidase (VDE), photosystem II subunit S (PsbS) and zeaxanthin epoxidase (ZEP) (note as VPZ hereafter) involved in non-photochemical quenching (NPQ) (Kromdijk et al. 2016).

Using the particle bombardment-mediated transformation approach, synthetic constructs of *MxgPPDK* and *SbPPDK* were transformed into energycane at the University of Florida and into miscanthus at the University of Illinois at Urbana-Champaign. A total of 3 transgenic energycane lines of overexpressed *MxgPPDK* gene

and its corresponding wild type were grown in the greenhouse and photosynthetic measurements of 8- to 12-week-old plants showed increased $V_{pmax}$ (the rate of PEP carboxylation). Field trial data of MxgPPDK transgenic lines are currently being analyzed. Preliminary measurements of the photosynthetic efficiency of 29 SbPPDK transgenic lines are being carried out in the greenhouse.

Calli induced from immature inflorescence of *Miscanthus* × *giganteus*, were transformed with synthetic constructs bearing *VPZ* or *PPDK* and several putatively transformed plants obtained. PCR genotyping of plants transformed with VPZ constructs using selection marker-specific primer pairs to detect the presence of BAR, NPTII or oHPT selection markers yielded 18, 9 and 29 positive lines out of 62, 114 and 46 putative transformants, respectively. Over 100 plants putatively transformed SbPPDK constructs in 2020 are currently being genotyped. Photosynthetic measurements of the miscanthus transgenic lines will be made in summer 2022.

**References/ Publications**

Mitchell RB, Schmer MR, Anderson WF, Jin V, Balkcom KS, Kiniry J, Coffin A, White P (2016) Dedicated Energy Crops and Crop Residues for Bioenergy Feedstocks in Central and Eastern USA. Bioenergy Res. 9: 384-398.

Long SP, Spence AK (2013) Toward Cool C4 Crops. Annu Rev Plant Biol 64(1): 701–722.

Wang D, Naidu SL, Portis AR, Moose SP, Long SP (2008) Can the cold tolerance of C4 photosynthesis in Miscanthus × giganteus relative to Zea mays be explained by differences in activities and thermal properties of Rubisco? J Exp Bot 59(7): 1779–1787.

Naidu SL, Moose SP, Al-Shoaibi AK, Raines CA, Long SP (2003) Cold Tolerance of C4 photosynthesis in Miscanthus × giganteus: Adaptation in Amounts and Sequence of C4 Photosynthetic Enzymes. Plant Physiol 132(July): 1688–1697.

Kromdijk J, et al. (2016) Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. Science 354(6314): 857–861.

Wang Y., Chan K.X. & Long S.P. (2021) Towards a dynamic photosynthesis model to guide yield improvement in C4 crops. Plant Journal, 107, 343-359.

**Funding Statement**

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018254.

# Improving Energycane by Metabolic Engineering for Hyperaccumulation of Lipids and RNAi Suppression of Flowering

Baskaran Kannan[1]* (kbaskaran@ufl.edu), Viet Cao[1], Shelby Korynta[1], Thaibinhduong Nguyen[1], Qasim Ali[1], Frico Situmeang[1], Hui Liu[2], John Shanklin[2], Fredy Altpeter[1], and **Stephen P. Long[3]**

[1]University of Florida, Gainesville, FL; [2]Brookhaven National Laboratory, Upton, NY; [3]University of Illinois at Urbana-Champaign, Urbana, IL

https://rogue.illinois.edu/

## Project Goals

The main objective of ROGUE (Renewable Oil Generated with Ultra-productive Energycane) project is to engineer the two most productive American crops—energycane and *Miscanthus*—to produce a sustainable supply of biodiesel, biojet fuel, and bioproducts.

1. Increasing oil accumulation and targeting this to the mature stem
2. Increasing photosynthetic efficiency to power oil synthesis
3. Multi-gene construct transformation of energycane and Miscanthus
4. Field testing, processing and techno-economic analysis

## Abstract

Metabolic engineering to divert carbon flux from sucrose to oil in a high biomass crop like energycane has been proposed as a strategy to boost both energy density of high biomass crops and lipid yields per acre for biodiesel production. Recently, we have succeeded with metabolic engineering to drastically increase triacyl glycerol (TAG) content in vegetative tissues of sugarcane by upregulating fatty acid synthesis, TAG synthesis and optimization of TAG storage (Zale et al. 2016, Parajuli et al. 2020). Energycane is like sugarcane an interspecific hybrid in the genus Saccharum. In contrast to sugarcane, energycane has a high proportion of the ancestral species *Saccharum spontaneum* in its genome which contributes to higher tiller number, biomass yield and persistence in addition to a reduced stem diameter and sugar content. Therefore, energycane is an ideal feedstock for this approach due to its superior biomass production and persistence. However, energycane is among the most recalcitrant crops in tissue culture and genetic transformation.

In this study, a multigene expression construct for lipogenic factors and selectable marker gene was generated by modular Golden Gate assembly with the goal to elevate the production of free fatty acids, catalyze their conversion into TAG and reduce TAG hydrolysis. Culture media optimizations included media supplementation to overcome browning and necrosis in tissue culture. Following biolistic gene transfer to embryogenic callus and selection of antibiotic resistant callus, transgenic plantlets were regenerated. The presence and expression of transgenes in the

regenerated plants were confirmed by PCR and qRT-PCR analysis, respectively. Transgenic plants were vegetatively propagated in the greenhouse and were transplanted to a field site at the University of Florida Plant Science and Education Center near Citra, FL in randomized and replicated plots under USDA-APHIS permit in March 2021. TAG and total fatty acid (TFA) content for different leaf positions, stem sections and juice as well as biomass fresh and dry weight were determined at the end of the growing season. The TAG and TFA content varied depending on tissue, maturity, stem and leaf position with the highest detected TAG and TFA content reaching 9.9% and 12.3% of leaf dry weight on average of three replications, respectively.

Lipid yield per land area from high biomass crops like energycane is determined by the lipid concentration in the biomass, the total biomass yield and the extractability of the lipids from the biomass. Flowering of energycane may affect oil yield and the extractability of oil. Upon flower induction vegetative growth ceases and sucrose/oil that has accumulated in the stalks is re-mobilized for use in reproductive development. Flowering often leads to dehydration of the stalk tissues, which negatively affects stalk density, and compromises sugar extraction in conventional energycane or lipid extraction in metabolically engineered lipid cane. Therefore, we generated transgenic energycane plants harboring a construct for RNAi mediated co-suppression of multiple flowering genes. Since energycane is vegetatively propagated for establishment of plantings, suppression of flowering will not require an altered agronomic practice while improving the biocontainment of the engineered crop. Transgenic plants were planted in replicated field plots at the University of Florida Plant Science and Education Center near Citra, FL in randomized and replicated plots, following a decision letter from USDA-APHIS that these transgenic plants are not-regulated by USDA-APHIS. Data will be presented comparing expression of the RNAi construct during photo inductive period with flowering response and biomass production.

**References/Publications**

Parajuli, S., Kannan, B., Karan, R., Sanahuja, G., Liu, H., Garcia-Ruiz, E., Kumar, D., Singh, V., Zhao, H., Long, S., Shanklin, J. and Altpeter, F (2020). Towards oilcane: Engineering hyperaccumulation of triacylglycerol into sugarcane stems. *GCB Bioenergy*. 12: 476– 490. https://doi.org/10.1111/gcbb.12684

Zale, J., Jung, J.H., Kim, J.Y., Pathak, B., Karan, R., Liu, H., Chen, X., Wu, H., Candreva, J., Zhai, Z., Shanklin, J. and Altpeter, F. (2016), Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. Plant Biotechnology Journal, 14: 661-669. https://doi.org/10.1111/pbi.12411

**Funding Statement**

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018254.

# Bridging the Gap between Academic and Commercial Biofuel Production: Pilot-scale Processing of Transgenic Energycane for Lipid and Sugar Recovery

Shraddha Maitra[1]* (smaitra@illinois.edu), Dang Viet Cao[2], Baskaran Kannan[2], Hui Liu[3], Fredy Altpeter[2], John Shanklin[3], Vijay Singh[1], and **Stephen P. Long[1]**

[1]University of Illinois at Urbana-Champaign, Urbana, IL; [2]University of Florida, Gainesville, FL; [3]Brookhaven National Laboratory, Upton, NY

https://rogue.illinois.edu/

## Project Goals

The main objective of ROGUE (Renewable Oil Generated with Ultra-productive Energycane) project is to engineer the two most productive American crops—energycane and *Miscanthus*—to produce a sustainable supply of biodiesel, biojet fuel, and bioproducts.

1. Increasing oil accumulation and targeting this to the mature stem
2. Increasing photosynthetic efficiency to power oil synthesis
3. Multi-gene construct transformation of energycane and Miscanthus
4. Field testing, processing and techno-economic analysis

## Abstract

Biofuels provide sustainable alternatives to petroleum-based fuels and lignocellulosic biomass is considered as the source of abundant renewable carbon. Cellulosic sugars and plant oils are particularly valued as they can be converted conveniently to bioethanol, biodiesel, and bio-jet. However, oilseeds that are also part of human food are still used predominantly for biodiesel production. To this end, transgenic bioenergy crops metabolically engineered to produce and accumulate high triacylglyceride (TAG) content have shown immense potential to replace oilseeds (Parajuli et al., 2020; Vanhercke et al., 2019; Zale et al., 2016). In ROGUE, bioenergy crop-energycane has been genetically modified to shift the carbon flux towards synthesis and accumulation of TAG molecules in their vegetative tissues. Transgenic energycane crops can be an excellent source of fermentable sugars and lipids/oil for biofuel production (Fouad et al., 2015). However, after establishing the basic characteristics of transgenic plants under controlled environmental conditions for genetic design and initial processing at bench-scale, it is critical to scale up the processes and test the proof-of-concept at pilot-scale for holistic analysis needed for potential commercialization. Bioprocessing at a pilot-scale provides industrially relevant data and helps identify research gaps for the successful commercialization of the project. Pilot-scale processing helps in getting better estimates for the techno-economical analysis and process simulation.

The transgenic energycane crop was cultivated under field conditions at the University of Florida-IFAS Plant Science Research and Education Unit, FL, USA. Upon harvesting, brown and green leaves were separated from the stem. A total of 223 kg (166 kg stem; 38.6 kg green leaves+tops; 18.3 kg brown leaves) were received at Integrated Bioprocess and Research Laboratory (IBRL), a translational research facility at the University of Illinois, IL, USA. Before harvest, the total TAG content in green leaves, stem, and juice were 3 to 10% (depending on leaf maturity and position), 0.6 to 1.3% (depending on stem maturity and position), and 1.4% per g dry weight, respectively. Transgenic energycane stems (97 kg) were processed in the pilot-scale facility. A combination of pilot-scale continuous hydrothermal pretreatment and disk milling, followed by hydrolysis with commercial cellulases, successfully recovered fermentable sugars and lipids. Chemical-free pretreatment at pilot-scale has shown no detrimental effect on oil recovery or the lipid profile from these transgenic energy crops. Thus, pilot-scale bioprocessing has been successfully established for transgenic energycane to produce fermentable sugars for bioethanol and lipids for biodiesel.

**References/Publications**


Parajuli, S., Kannan, B., Karan, R., Sanahuja, G., Liu, H., Garcia-Ruiz, E., Kumar, D., Singh, V., Zhao, H., Long, S., Shanklin, J., Altpeter, F., 2020. Towards oilcane: Engineering hyperaccumulation of triacylglycerol into sugarcane stems. GCB Bioenergy 12, 476–490. https://doi.org/10.1111/gcbb.12684

Vanhercke, T., Belide, S., Taylor, M.C., El Tahchy, A., Okada, S., Rolland, V., Liu, Q., Mitchell, M., Shrestha, P., Venables, I., Ma, L., Blundell, C., Mathew, A., Ziolkowski, L., Niesner, N., Hussain, D., Dong, B., Liu, G., Godwin, I.D., Lee, J., Rug, M., Zhou, X.-R., Singh, S.P., Petrie, J.R., 2019. Up-regulation of lipid biosynthesis increases the oil content in leaves of Sorghum bicolor. Plant Biotechnol. J. 17, 220–232. https://doi.org/10.1111/pbi.12959

Zale, J., Jung, J.H., Kim, J.Y., Pathak, B., Karan, R., Liu, H., Chen, X., Wu, H., Candreva, J., Zhai, Z., Shanklin, J., Altpeter, F., 2016. Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. Plant Biotechnol. J. 14, 661–669. https://doi.org/10.1111/pbi.12411

Fouad, W. M., Hao, W., Xiong, Y., Steeves,C., Sandhu,S. K., Altpeter, F., 2015. Generation of Transgenic Energy Cane Plants with Integration of Minimal Transgene Expression Cassette. Current Pharmaceutical Biotechnology 2015; 16(5). https://doi.org/10.2174/1389201016666150303151559


**Funding Statement**


This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018254.

**Title:** Germination and Seed Size Variability in Camelina

**Authors:** GunNam Na[1], Racheal Upton[1], Helen Tsai[2], Fernando Correr[1], Jennifer Lachowiec[1]* (jennifer.lachowiec@montana.edu), Luca Comai[2] and **Chaofu Lu[1]**

**Institutions:** [1]Montana State University, Bozeman; [2]University of California, Davis

**Website URL:** https://www.montana.edu/econproject/

**Project Goals:** Camelina is a Brassica oilseed crop that has great potential to become a sustainable source of bioenergy in the US. However, the low nitrogen use efficiency and the low seed and oil yield compared to other major oilseed crops hinder this potential. The goal of this project is to decipher the genetic and physiological mechanisms that determine the nitrogen use efficiency and oilseed yield during the most critical processes of the camelina life cycle: 1) how camelina, in partnership with soil microbes, maximizes its ability to absorb and assimilate nitrogen into vegetative biomass; and 2) upon the transition to reproductive growth, how nitrogen is efficiently remobilized from senescing tissues (leaves and silicles) into sinks (seeds) to optimize yield potential by increasing seed size and enhancing oil synthesis.

**Abstract Text:**

Predictable germination is critical to the life cycle of productive crop growth. Within an optimal environment, isogenic seed can demonstrate considerable germination variability. We examined sources and causes of variable germination in camelina, with a focus on seed size. The oilseed camelina generally reproduces through self-pollination, with outcrossing rates of 0.09-0.28%, resulting in largely homozygous lines; however, the seed batch produced by a single plant still varies in size. Indeed, across accessions and seed size mutants, we observed differences in not only the mean seed size—as measured by length, width, and area—but also the variance. For example, the accession Suneson produced seed batches more variable in seed size compared to other accessions grown in the same conditions. Seed size (Huang et al 2021) is positively correlated to time to 50% germination. We therefore hypothesize that variable seed size across batches result in more variable germination. Seed size batch variability was also impacted in TILLING mutant lines produced in the Licalla background, with a subset of mutant families having highly variable seed batches relative to others. Furthermore, non-Mendelian patterns of inheritance of seed size in the mutant lines suggest epigenetic factors influencing seed size. Epigenetic factors also may play an important role in seed germination in camelina given our observation that some accessions were sensitive to nitrogen fertility of the previous generation. Low nitrogen conditions also resulted in delayed and varied germination, which may represent an environmentally induced dormancy effect enabling camelina seed to overwinter. An important candidate regulator of seed size is miR167a (Na et al, 2019), which, when overexpressed specifically in camelina seeds, results in larger seed size. The larger seed size may arise through a delay in development, wherein the expression pattern of transgenic seed 10 and 12 days after fertilization are similar to wild-type seed at 8 days after fertilization. Examining the genetic

network tied to miR167a overexpression points to several candidate regulators, with one of the most highly connected genes, *DELAY OF GERMINATION1* (*DOG1*) found in the module of genes with delayed expression in the overexpression line. Further work will tie these observations, examining the role of *DOG1* in camelina seed size variation and impacts on germination variability in light of epigenetic factors.

**References/Publications**


1. Na, G., Mu, X., Grabowski, P., Schmutz, J. and Lu, C., 2019. Enhancing microRNA 167A expression in seed decreases the α-linolenic acid content and increases seed size in Camelina sativa. *The Plant Journal*, *98*(2), pp.346-358.
2. Li, H., Hu, X., Lovell, J. T., Grabowski, P. P., Mamidi, S., Chen, C., Amirebrahimi, M., Kahanda, I., Mumey, B., Barry, K., Kudrna, D., Schmutz, J., Lachowiec, J., & Lu, C. 2021. Genetic dissection of natural variation in oilseed traits of camelina by whole-genome resequencing and QTL mapping. *The Plant Genome, 14* (2), e20110.



**Funding Statement:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021369.

**Title:** Understanding Nitrogen Use Efficiency and Oilseed Traits in Camelina by High Resolution Genome Sequencing and Whole-genome Resequencing

**Authors:** Charlemagne Lim[1*](charlemagnealexalim@montana.edu), Samuel Decker[1*](samuel.decker@student.montana.edu), Huang Li[1], John Lovell[2], Paul Grabowski[2], Jeremy Schmutz[2], Jennifer Lachowiec[1], Chengci Chen[1], **Chaofu Lu**[1]

**Institutions:** [1]Montana State University, Bozeman; [2]HudsonAlpha Institute for Biotechnology, Huntsville, AL.

**Website URL**: https://www.montana.edu/econproject/index.html

**Project Goal**s: Camelina is a Brassica oilseed crop that has great potential to become a sustainable source of bioenergy in the US. However, the low nitrogen use efficiency and the low seed and oil yield compared to other major oilseed crops hinder this potential. The goal of this project is to decipher the genetic and physiological mechanisms that determine the nitrogen use efficiency and oilseed yield during the most critical processes of the camelina life cycle: 1) how camelina, in partnership with soil microbes, maximizes its ability to absorb and assimilate nitrogen into vegetative biomass; and 2) upon the transition to reproductive growth, how nitrogen is efficiently remobilized from senescing tissues (leaves and silicles) into sinks (seeds) to optimize yield potential by increasing seed size and enhancing oil synthesis.

**Abstract Text:**

Camelina is an emerging oilseed crop that has great potential as a renewable, high quality, and sustainable source of bioenergy in the US. However, this great potential is offset by camelina's low nitrogen use efficiency and low seed and oil yield compared to other oilseed crops. Development of camelina varieties with superior seed and oil yield characteristics requires careful examination and the decryption of the underlying genetic and physiological mechanisms critical in the camelina life cycle that govern the foraging and uptake of nutrients from the soil, source-sink relationships and resource allocation, reproduction, seed formation, and oil synthesis. This can be facilitated by new technological advancements in biotechnology in conjunction with modern breeding techniques. To understand the natural variation within camelina species, we determined the population structure of a worldwide collection of 222 *Camelina sativa* accessions using 161,301 single nucleotide polymorphisms (SNPs) derived from whole-genome resequencing. Additionally, when coupled with genome-wide association studies (GWAS) and linkage mapping using a recombinant inbred line population, QTL and candidate genes that control field-evaluated agronomic traits related to oilseed production such as seed size, oil

content, fatty acid composition, and flowering time were identified. To further locate these QTLs and candidate genes with high precision, we have assembled a high-quality camelina genome through Pac-bio HiFi genome sequencing coupled with Hi-C genome scaffolding. The sequenced camelina line (Suneson) has been used in our mapping studies and will provide more accurate genetic markers. Field trials under low and high nitrogen soils were conducted in North East and Central Montana, and in Eastern Washington, US in 2021 (and to be repeated in 2022) to generate a new set of phenotypic data from the accessions, which were used to determine the relationships among morpho-agronomic traits and for the GWAS to identify QTLs and candidate genes for increased nitrogen use efficiency and oilseed traits. Identification of QTL and genes together with their genetic markers will greatly facilitate breeding efforts to develop camelina varieties with high nitrogen use efficiency, high seed yield and oil quality, and are adaptable to the US northwest environments.

**References/Publications:**

1. Li H, Hu X, Lovell JT, Grabowski PP, Mamidi S, Chen C, Amirebrahimi M, Kahanda I, Mumey B, Barry K, Kudrna D, Schmutz J, Lachowiec J, & Lu C. Genetic dissection of natural variation in oilseed traits of camelina by whole-genome resequencing and QTL mapping. Plant Genome. 2021;14:e20110. https://doi.org/10.1002/tpg2.20110

**Funding Statement:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021369.

**Title:** Genetic Engineering of Camelina to Improve Seed Oil Yield

**Authors:** Xiao-Hong Yu, [1,2] * (xhyu@bnl.gov), Sanket Anaokar,[1] Yuanheng Cai,[1,2] and **John Shanklin**[1,2]

**Institutions:** [1]Biology Department, Brookhaven National Laboratory, Upton, NY; and [2]University of Stony Brook, Stony Brook

**Website URL:** https://www.montana.edu/econproject/index.html

**Project Goals:** This work is part of the ECON, i.e., Enhancing Camelina Oilseed Production with Minimum Nitrogen Fertilization in Sustainable Cropping Systems project. Camelina is a Brassica oilseed crop that has great potential to become a sustainable source of bioenergy in the US. However, its low nitrogen use efficiency, seed and oil yield as compared to other major oilseed crops limit its potential. The goal of this project is to decipher the genetic and physiological mechanisms that determine the nitrogen use efficiency and oilseed yield during critical processes of the camelina life cycle: 1) how camelina, in partnership with soil microbes, maximizes its ability to absorb and assimilate nitrogen into vegetative biomass; and 2) upon the transition to reproductive growth, how nitrogen is efficiently remobilized from senescing tissues (leaves and silicles) into sinks (seeds) to optimize yield potential by increasing seed size and enhancing oil synthesis.

**Abstract Text:** The synthesis of fatty acids (FA), the building blocks of TAGs, is highly energy demanding and is therefore under tight multi-level homeostatic control. WRI1 is the master transcriptional activator that targets many FA synthesis genes[1], overexpression of which increase FA accumulation in seeds. Trehalose 6-phosphate (T6P) reduces the phosphorylation WRI by SnRK1, to preventing its degradation[2]. Overexpression of the *E. coli* T6P synthase, OtsA led to increased FA accumulation in Arabidopsis[3]. To improve WRI stability in Camelina, T6P synthase, under the control of the seed specific Phaseolin promoter was transformed into the wild type camelina Suneson to boost T6P level in its seeds. OstsA transgenics showed poor germination with only 3 out of 8 transgenic of the transgenic seeds germinating. Additional transformation yielded more than 300 lines of which 36 germinated. Seeds from these lines were collected for analysis of lipid composition and fatty acid content.

The first committed and rate limiting step in FA synthesis is acetyl-CoA carboxylase (ACCase), which converts acetyl-CoA to malonyl-CoA in the plastid. Excess FA inhibits the carboxytransferase (CT) ACCase half-reaction and triggers irreversible ACCase inhibition. This inhibition that is mediated by biotin attachment domain-containing (BADC) proteins, homologs of biotin carboxytransferase proteins (BCCPs)[4]. BADCs lack a critical lysine necessary for the covalently attachment of the biotin cofactor required for carboxylation in BCCPs[5]. BADCs are conditional inhibitors of FA synthesis that displace BCCP in the ACCase complex when intracellular FA accumulate. Knocking-out two of the Arabidopsis *BADC* genes, i.e., badc1badc3 resulted in an increase in TAG content of 25% with respect to that of wild type[4]. We are targeting the downregulation of BADC genes in camelina as a primary approach to increasing seed oil yield. CRISPR vectors targeting single BADC genes i.e., *BADC1, BADC2, and BADC3*

and the combination of *BADC1 and BADC3* have been constructed and transformed into camelina Suneson. 5 CRISPR-CsBADC3 lines were genotyped. One line was edited in 1 allele of each of the *BADC3* isoforms. We also obtained 8 CRISPR-CsBADC1 lines and 36 CRISPR-CsBADC1,3 lines for which genotyping is underway. Increases in FA synthesis rates are expected to result in increased levels of total FA accumulation, but we anticipate additional factors will be required to convert excess FA into TAG and to prevent its degradation. We have previously used an Arabidopsis or nasturtium diacylglycerol acyltransferase, DGAT1 and an oil droplet protecting protein, sesame OLEOSIN (cysOLE1) to increase oil accumulation[6]. We recently found that a mammalian DGAT2 (mDGAT2) is more effective than Arabidopsis DGAT1 in converting FA into TAG. We have also created variants of the sesame OLEOSIN that significantly increases its ability to protect TAG relative to cysOLE1. These "pull" and "protect" factors including *Sesame* oleosin, and several variants engineered to increase its stability, and mouse DGAT2 were employed to boost oil accumulation in camelina. Constructs containing the single gene for oleosins and/or mouse DGAT2 were transformed into camelina *suneson*. Currently we obtained 15, 13, 5, 14, 26, and 28 transformants of CysOle1, Ole1_CysDel_KR , Ole1_5_Mod, mDGAT, CysOle1+ mDGAT, Ole1_CysDel_KR + mDGAT and Ole1_5_Mod + mDGAT, respectively.

Future efforts will focus on identifying the most successful BADC suppression lines and the most effective additional gene combinations described above. The optimized approaches will be combined to create camelina seeds with significantly elevated TAG content beyond those discovered by genetic variation. Subsequently we will combine optimized oilseed accumulation factors in lines optimized for nitrogen use efficiency.

**References/Publications**


1. Cernac A, Benning C (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. Plant J 40: 575-585
2. Zhai Z, Keereetaweep J, Liu H, Feil R, Lunn JE, Shanklin J (2018) Trehalose 6-Phosphate Positively Regulates Fatty Acid Synthesis by Stabilizing WRINKLED1. Plant Cell 30: 2616-2627
3. Zhai Z, Keereetaweep J, Liu H, Feil R, Lunn JE, Shanklin J (2021) Expression of a BacteriaTrehalose-6-phosphate Synthase otsA Increases Oil Accumulation in Plant Seeds and Vegetative Tissues. Front Plant Sci 12: 656962
4. Keereetaweep J, Liu H, Zhai Z, Shanklin J (2018) Biotin Attachment Domain-Containing Proteins Irreversibly Inhibit Acetyl CoA Carboxylase. Plant Physiol 177: 208-215
5. Salie MJ, Zhang N, Lancikova V, Xu D, Thelen JJ (2016) A Family of Negative Regulators Targets the Committed Step of de Novo Fatty Acid Biosynthesis. Plant Cell 28: 2312-2325
6. Zale J, Jung JH, Kim JY, Pathak B, Karan R, Liu H, Chen X, Wu H, Candreva J, Zhai Z, Shanklin J, Altpeter F (2016) Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. Plant Biotechnol J 14: 661-669



**Funding Statement:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021369.

# Camelina growth impacted by bacteria and nitrogen stress

Qing Zheng[1,] * (qingzheng@lbl.gov), Chuntao Yin[2], Peter Andeer[1], Cody Willmore[3], **Chaofu Lu[4]**, Trent R. Northen[1,5], Timothy Paulitz[3], Susannah Tringe[1,5]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]North Central Agricultural Research Laboratory, USDA-ARS, Brookings, SD; [3]Wheat Health, Genetics, and Quality Research Unit, USDA-ARS, Pullman, WA; [4]Montana State University, Bozeman, MT; and [5]Joint Genome Institute, Berkeley, CA.

**Website URL:** https://www.montana.edu/econproject/index.html

**Project Goals: Camelina is a Brassica oilseed crop that has great potential to become a sustainable source of bioenergy in the US. However, the low nitrogen use efficiency and the low seed and oil yield compared to other major oilseed crops hinder this potential. The goal of this project is to decipher the genetic and physiological mechanisms that determine the nitrogen use efficiency and oilseed yield during the most critical processes of the camelina life cycle. Specifically, we look to address: 1) how camelina, in partnership with soil microbes, maximizes its ability to absorb and assimilate nitrogen into vegetative biomass; and 2) upon the transition to reproductive growth, how nitrogen is efficiently remobilized from senescing tissues (leaves and silicles) into sinks (seeds) to optimize yield potential by increasing seed size and enhancing oil synthesis.**

Camelina is a promising non-food oilseed crop with great potential for various applications in the US. It is normally grown as a rotation crop in the wheat-camelina systems, providing a biofuel feedstock and potentially improving cereal-based cropping systems and boosting rural economies. However, high nitrogen inputs and low oilseed yields hinder its great potential. To enhance camelina oilseed production with minimum nitrogen fertilization, it is important to understand how camelina plants interact with beneficial microbes that may enhance nitrogen use efficiency and boost oil yield.

To address how soil microbes impact camelina growth and nitrogen use efficiency, bacteria were isolated and identified from 33 locations across wheat cropping zones of eastern Washington. Soils were collected from 33 locations and camelina cv. Suneson were grown in collected soils in the

greenhouse for four weeks. Bacteria from camelina rhizosphere grown in 24 of the 33 location soils were cultured in Tryptic Soy Agar and R2A media. Over 3000 bacterial colonies were isolated and 920 of them were identified by 16S rRNA sequencing. A total of 51 unique bacterial genera, belonging to 31 families, were identified. *Pseudomonas, Bacillus, Flavobacterium*, and *Arthrobacter* were the most abundant genera in our bacterial collection. Interestingly, genera *Massilia, Pedobacter, Variovorax,* and *Caulobacter, which are* common in wheat rhizospheres, were also found. In addition, we found that culture media had a slight but significant effect on bacterial isolation, while no significant influence of locations was observed.

To understand how camelina uses root exudates to recruit beneficial microbes under nitrogen stress, we profiled root exudate compositions during the early growth period. We first evaluated camelina growth in axenic fabricated ecosystems (EcoFABs) designed for root exudate collection. Camelina was cultivated under limited (70 ppm N) and replete nitrogen (210 ppm N) conditions in EcoFAB 2.0 with weekly collection of metabolites. We successfully cultivated camelina in EcoFABs for five weeks and the plants were harvested for root and shoot biomass measurements. The results showed that the overall camelina biomass was lower under limited nitrogen conditions, but only root biomass was significantly lower in the limited nitrogen group compared to the replete nitrogen group. We will profile exudate composition during N stress using LC-MS/MS metabolomics and conduct camelina plant-microbe co-culture experiments using isolated bacteria and synthetic communities (SynCom) to further explore how camelina uses exudates to recruit beneficial microbes that may help increase nitrogen use efficiency and oil yield.

**Funding Statement:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021369.

**Title: Development and usage of atomic mapping for estimation of nitrogen fluxes in plant metabolic networks**

**Authors:** Sebastian Huß,[1,2] Presenting Author* (shuss@uni-potsdam.de), Rika Judd,[3] Kaan Koper,[3] Hiroshi A. Maeda,[3] and **Zoran Nikoloski**,[1,2]

**Institutions:** [1]Bioinformatics, Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany; [2]Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany; and [3]Department of Botany, University of Wisconsin-Madison, Madison, WI, USA

**Website URL:** https://nfluxmap.github.io/

**Project Goals:** To construct plant N flux maps (NFMs) from plant genomes and to determine both biochemical and system-level functionality of plant N metabolic network.

**Abstract: Nitrogen (N)** is an essential element of organic molecules, such as amino acids and proteins, but is often limiting for important traits, like biomass, growth, and yield. In addition, the supply of N in the soil is often limited, and the distribution of N throughout plant metabolic pathways is elusive. Understanding the impact that N has on the overall (re)distribution of cellular resources can provide insights useful for improving N use efficiency of model plants and crops such as determining rate limiting pathways that hinder N flow and therefore plant growth. One way towards achieving this goal is to assemble the **N flux map (NFM)**, embedded in large-scale models of plant (Arnold & Nikoloski 2014, Küken & Nikoloski 2019), and use it to estimate reaction fluxes by integration of isotope labeling patterns under different N availabilities. However, the available metabolic network and data resources in plants cannot address this challenge since they are focused on simulation and integration of carbon labeling patterns.

Estimation of intracellular fluxes based on isotope labeling patterns of metabolites in a metabolic network relies on approaches from metabolic flux analysis (MFA) (Wiechert, 2001, Basler, Fernie & Nikoloski 2018). Application of MFA requires a stoichiometric model with **atomic mappings** that are currently not available for most large-scale metabolic network models, particularly in plants. Here we established an automated workflow to apply reliable automated atom mapping (AAM) approaches, such as Reaction Decoder Toolkit (RDT) (Rahman, et al. 2016), on large-scale metabolic models of *Arabidopsis thaliana* to facilitate estimation of fluxes By using atom maps from the MetaCyc database, that cannot be readily used in flux estimation, we show that the atom maps created by RDT are accurate. The established workflow is made available at https://github.com/sebahu/UniversalRDT/ and the resulting atomic mappings will be integrated in KBase.

We further demonstrate the utility of our automated workflow by simulating $^{15}$N-isotope enrichment and identifying metabolites which show enrichment patterns that are informative

regarding one or several fluxes and thus need to be measured in $^{15}$N-MFA studies of *A. thaliana*. To experimentally examine isotope enrichment throughout plant metabolism, 21-day-old *A. thaliana* Col-0 plants were fed with $^{15}$N-labeled nitrate and ammonia over a course of 3 days using a hydroponics (Conn et al. 2013) and the resulting $^{15}$N incorporation was analyzed by both GC-MS and LC-MS/MS. Starting with amino acids, we demonstrate $^{15}$N label incorporation over 24 hours for the labeled amino acids detected (e.g., phenylalanine, alanine) while some other compounds (e.g. serine, methionine) showed much slower labeling kinetics. These $^{15}$N kinetic labeling data for amino acids, as well as other N containing compounds, will be incorporated into the developed atomic mappings to determine reaction fluxes of N in the large-scale models.

In addition to increasing plant biomass to address high demands for important crops, characterizing N flow and building the NFM will also allow the development of metabolic engineering of N pathways towards increase the production of compounds of interest and crop growth. The developed atom mapping of Arabidopsis metabolic network will be further expanded by incorporating additional connections of N flow based on novel aminotransferase (AT) activities identified by high-throughput AT substrate screening within this same project (as presented in a separate poster, Koper *et al.*). Beyond *A. thaliana*, the proposed workflow that we developed so far in *A. thaliana* will be expanded to other plant species, such as sorghum, for which metabolic models have been assembled and will facilitate MFA at the levels of large-scale metabolic networks. The comparison of NFMs from the model dicot and bioenergy monocot plants (*A. thaliana* vs. sorghum) will allow us to identify conserved and unique features of N metabolic network among different plants.

**References/Publications**


1. Arnold, A., & Nikoloski, Z. (2014, July). Bottom-up Metabolic Reconstruction of Arabidopsis and Its Applition to Determining the Metabolic Costs of Enzyme Production. Plant Physiology, 165(3), 1380-1391. doi:10.1104/pp.114.235358
2. Conn, S. J. *et al.* (2013). Protocol: optimising hydroponic growth systems for nutritional and physiological analysis of Arabidopsis thaliana and other plants. *Plant Methods* 9, 4, doi:10.1186/1746-4811-9-4
3. Basler, G., Fernie, A., & Nikoloski, Z. (2018, Nov 23). Advances in metabolic flux analysis toward genome-scale profiling of higher organisms. Biosci Rep., 38(6), p. BSR20170224. doi:10.1042/BSR20170224
4. Küken, A., & Nikoloski, Z. (2019). Computational Approaches to Design and Test Plant Synthetic Metabolic Pathways. Plant Physiology, 179(3), pp. 894–906. doi:10.1104/pp.18.01273
5. Rahman, S. A., Torrance, G., Baldacci, L., Cuesta, S. M., Fenninger, F., Gopal, N., . . . Thornton, J. M. (2016). Reaction Decoder Tool (RDT): extracting features from chemical reactions. Bioinformatics, 32(13), 2065–2066. doi:10.1093/bioinformatics/btw096
6. Wiechert, W. (2001, Jul). 13C metabolic flux analysis. Metab Eng., 3(3), pp. 195-206. doi:10.1006/mben.2001.0187



**Funding statement**: This research was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0020390.

# Comprehensive Characterization of Multi-Substrate Specificity of Aminotransferase Family Enzymes

**Authors:** Kaan Koper[1*](koper@wisc.edu), Marcus De Raad[2], Shogo Hataya[3], Sang-Woo Han[2], Ramani Kothadia[2], Hugh Salamon[2], Yasuo Yoshikuni[2], Taichi Takasuka[3], Trent Northen[2], **Hiroshi A. Maeda[1]**

[1] University of Wisconsin–Madison, Madison, Wisconsin, USA.
[2] Lawrence Berkeley National Laboratory, Berkeley, California, USA
[3] Hokkaido University, Sapporo, Japan

**Website URL:** https://nfluxmap.github.io/project/

**Project goals:** To comprehensively characterize multi-substrate specificities of all AT enzymes of *Arabidopsis* and *sorghum* through high-throughput mass spectrometry-based assays, and to integrate novel AT reactions to construct nitrogen flux map (NFM) of plants.

**Abstract: Aminotransferases (ATs)** are critical enzymes of plant **nitrogen (N)** metabolism that facilitate the transfer of amino N between biomolecules, namely amino acid donors and keto acid acceptors. Biochemical properties of many plant ATs have been tested with a certain substrate(s); however, the full spectra of **AT multi-substrate specificity** remain largely unexplored due to the tedious nature of AT characterization and poor sequence-function relationship. As a result, many cross-pathway N transfer reactions remain uncharacterized and unaccounted, making our understanding of plant N metabolic network far from complete.

To address these major knowledge gaps here we first generated a comprehensive phylogeny of AT enzyme families by utilizing fully-sequenced high-quality genomes of 15 species from each of six kingdoms, which identified a number of AT groups (i.e., well supported monophyletic clades, **Fig. 1A**). To further examine AT functionality, we successfully expressed and purified all 49 *Arabidopsis* AT candidates using *E. coli* and/or cell-free wheat germ expression systems (**Fig. 1B**) and validated previously-reported enzymatic activity for at least ~50% of them so far. To comprehensively examine their substrate specificity, we established high-throughput substrate screening methods using nanostructure-initiator mass spectrometry (NIMS) and matrix-assisted laser desorption/ionization MS (MALDI-MS) and determined substrate utilization profiles of eight ATs against 31 amino donors with 4 different amino acceptors (124 reactions for each). Even for ATs with well-established functions (e.g. AlaAT), we uncovered that many of them exhibit previously unknown reactions and multi-substrate specificities (**Fig. 1C**). Notably, we identified an unusual pattern of multi-substrate specificity among cytosolic aromatic ATs that allows them to utilize Phe, Tyr, Trp, Met, His, and Leu, which were further validated and quantified by LC-$MS^2$-based assays. These findings revealed the presence of a novel N network linking cytosolic pools of these six amino acids, which could be important for recycling N and rebalancing Phe and Met levels, such as during lignin biosynthesis (**Fig. 1D**). As we are completing the multi-substrate specificity screening for ATs from *Arabidopsis* and then *sorghum*, the obtained biochemical data are being integrated into the **nitrogen flux map (NFM)** that we are also developing in this project (see the other poster by Sebastian Huss *et al.*). The joint effort will discovery previously unknown AT-mediated N metabolic connections and thereby generate accurate NFMs of plants, which will guide us to identify engineering targets of N metabolic network to improve nitrogen use efficiency in crops.

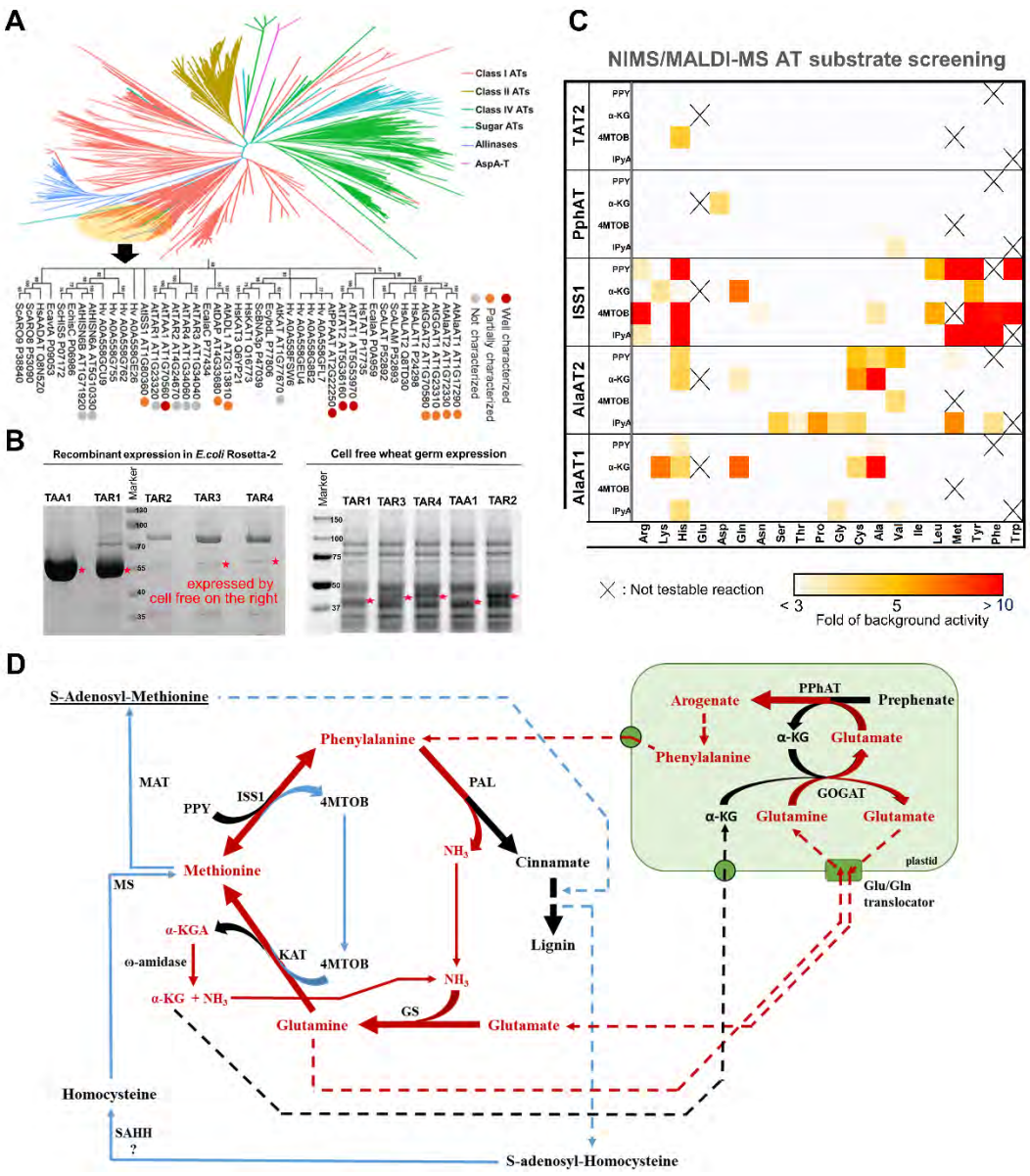



**Figure 1: Functional characterization of Arabidopsis AT family enzymes. A.** Phylogeny of ATs from 90 species of plant, animal, fungi, protist, eubacteria, and archaebacteria kingdoms, with the rectangular subtree highlighting a subset of AT clades from the representative model organisms – *Arabidopsis thaliana* (At), *Saccharomyces cerevisiae* (Sc), *Homo sapiens* (Hs), *Escherichia coli* (Ec) and *Halobacterium volcanii* (Hv). **B.** AT enzymes expressed recombinantly (left) and using the cell free wheat germ system (right). **C.** MALDI screening of multi-substrate specificities of five AT enzymes tested with phenylpyruvate (PPY), α-ketoglutarate (KG), 4-methyl-2-oxopentanoate (4MTOB) or indole-3-pyruvate (IPyA) as keto acid acceptors and twenty proteinogenic amino acids as amino donors. **D.** A new network of N rebalancing around phenylalanine biosynthesis and metabolism along with methionine and other amino acids (red). The *S*-adenosylmethionine (SAM) cycle is in light blue.

*This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0020390 and the Joint Genome Institute (JGI) Community Science Program (CSP) award no. 503757.*

**Biological Design of *Lemnaceae* Aquatic Plants for Biodiesel Production**

Evan Ernst*,[1,2] James Birchler,[3] Eric Lam,[4] Jorg Schwender,[5] John Shanklin,[5] and **Robert A. Martienssen (martiens@cshl.edu)**[1,2]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [2]Howard Hughes Medical Institute, Cold Spring Harbor, NY; [3]University of Missouri, Columbia, MO; [4]Rutgers University, New Brunswick, NJ; and [5]Brookhaven National Laboratory, Brookhaven, NY

**Project Goals:**

1. **Leveraging our transformation methods, we will develop a comprehensive toolset for genetic manipulation of *Lemnaceae*. We will establish CRISPR/Cas9 genome editing to complement our artificial miRNA silencing methods. We will construct artificial chromosomes in *Lemna minor* to potentiate whole pathway engineering.**
2. **Resting and over-wintering fronds have higher starch content than corn kernels, but the energy density of oil is more than twice that of starch. We will use regulatory network and metabolic flux modeling to re-engineer the carbon allocation pathways to optimize triacylglyceride (TAG).**
3. **We will use comparative genomics of multiple *Lemnaceae* genome sequences, an extensive living collection of global accessions, and systems network analysis to characterize gene expression networks underpinning developmental and environmental responses to maximize bioenergy products while preserving rapid biomass accumulation. Nutrient deprivation and $CO_2$ irrigation will be used to enhance yield.**

*Lemnaceae* species (commonly called duckweeds) are the world's smallest aquatic flowering plants. Under optimal conditions, their rapid clonal growth rate can double the number of fronds in 30 hours and produce 64 grams of biomass per gram starting weight in a week, which is far beyond that of terrestrial crops such as corn (2.3 g/g /week), and unencumbered by secondary products such as lignin. *Lemnaceae* offer an attractive alternative to algae as biofuel feedstocks because of their robust growth in open ponds and the relative ease of harvesting dry material. Convenient metabolic labeling in culture makes Lemna a good system for pathway modeling and engineering, as nutrients are taken up from liquid growth media, and non-responsive stomata can utilize very high levels of atmospheric $CO_2$. Our goal is to divert substantial accumulated carbon from starch to oil metabolism in *Lemnaceae*, using resting fronds as the storage tissue.

The Martienssen and Lam labs have produced eight new reference quality *Lemnaceae* genome assemblies of *W. australiana*, *L. gibba*, *L. minor*, *L. turionifera*, and three allotriploid hybrid *L. japonica* (*L. minor* x *L. turionifera*) clones using Oxford Nanopore long reads and Hi-C contact maps. Comparisons of the chromosome-scale assemblies demonstrate a high degree of synteny across all 21 chromosomes within the *Lemna* genus, with only a single translocation consistently identified specifically in *L. turionifera*, while *W. australiana* has 20 chromosomes like *S. polyrhiza*, yet with significant architectural differences. In the Birchler Lab, antibodies against centromeric histone H3 were raised for four species to determine centromere organization and

identify centromeric repeat sequences across the duckweeds via CUT&RUN. The sequenced species encode between 14,000 and 20,000 genes, significantly fewer than terrestrial monocots such as rice and *Brachypodium*, and comparable to the algae *Chlamydomonas reinhardtii*. Methylome and small RNA sequencing revealed dramatic differences between the three genera consistent with known pathways of RNA directed DNA methylation. Analysis of missing and diverged orthogroups across the *Lemnaceae*, 10 other monocots and 7 non-monocots revealed variations that likely account for these and other traits including: reduced morphology, aquatic habitat, clonal reproduction, dormancy, high photosynthetic rate, and lipid production.

Key experiments in the Lam Lab have confirmed that natural genetic variation in *S. polyrhiza* leads to variable turion production. RNA-sequencing of two genotypes at the extremes of turion yield have identified turion-specific genes associated with dormancy, starch biosynthesis, and putative transcription factors that may be involved in the developmental transition. In addition, turion-specific expression of genes involved in lipid metabolism and oil biosynthesis were found in both *S. polyrhiza* as well as *L. turionifera*. Current work comparing transcript induction kinetics between different genotypes is underway to filter out the most promising candidate genes for functional validation. In collaboration with the Shanklin lab, a four- to six-fold increase in total TAG levels were found in turions of both duckweed species, consistent with predictions from RNA-seq, providing novel leads to target genes for directed modification of lipid content.

In previous work from the Shanklin and Schwender Labs, we expressed an Arabidopsis WRINKLED1, (WRI1) the master transcriptional activator of fatty acid synthesis in *Lemna japonica* line Lj8627. This resulted in <1% of DW of TAG along with large reduction in growth rate along and significant developmental abnormalities. Next, we constructed a CFP-N terminally tagged version of the Arabidopsis WRI1 under the control of inducible estradiol inducible XVE promoter, co-expressed with a sesame Oleosin 1 gene variant, (ROGUE Biosystems Design) in which its degradation signals had been minimized to optimize its TAG protective function, along with a very strong mammalian DGAT2 (CABBI Energy Center funding) to create Lj8627-33 (ODW) transgenics. Growth of OWD transgenics cultured in the presence of 100uM estradiol for four days resulted in the accumulation of 16.4% total fatty acid by DW compared to 5.2% in the parental line and 8.7% TAG per DW compared to 0.07 (124-fold increase). Subsequent detailed analyses confirmed strong TAG production and accumulation of very long chain fatty acids in TAG, but revealed that estradiol induction of WRI1 in ODW lines results in an allocation tradeoff of starch to TAG and a reduction in growth rate. Work is underway to evaluate the extent of lipid futile cycling and to model biomass synthesis using physiological, biochemical, and transcriptomic data to understand the growth rate reduction and inform strategies for mitigation.

*Funding for this project is provided by the DOE Office of Biological & Environmental Research (DE-SC0018244).*

**Title:** Converting Methoxy Groups on Lignin-Derived Aromatics from a Toxic Hurdle to a Useful Resource: A Systems-Driven Approach

**Authors:** Alexander B. Alleman,[1] Nkrumah A. Grant,[1] Alekhya Govindaraju,[2] Akorede L. Seriki,[1] Isaiah D. Jordan,[3] Jeremy A Draghi,[3] N. Cecilia Martinez-Gomez,[2] Andreas E. Vasdekis,[1] and **Christopher J. Marx[1]* (cmarx@uidaho.edu)**

**Institutions:** [1]University of Idaho, Moscow, ID; [2]University of California, Berkeley, CA; [3]Virginia Institute of Technology, Blacksburg, VA

**Project Goals:** In order to convert methoxylated aromatics that originate from lignin hydrolysis into bioproducts it is critical to manage the toxicity of formaldehyde that is generated internally from these methoxy groups. We have discovered both novel genetic factors in *Methylobacterium extorquens* involved in managing formaldehyde toxicity and a key role for cell-to-cell variability in key phenotypes. As such, our goals are to (1) identify physiological hurdles to growth/viability and production, (2) assess the environmental influence upon single cell heterogeneity in these traits, and then (3) generate genetic combinations from targets uncovered in *1* and use a model-driven approach to synthesize knowledge regarding heterogeneous phenotypes uncovered in *2* to generate strains with improved butanol production and optimized combinations of environmental variations.

**Abstract Text:** Lignin-derived compounds from plant biomass are amongst the most recalcitrant for microbial conversion. Hydrolysates contain a wide variety of aromatic molecules, which are often toxic. Furthermore, a particular issue with these molecules is that many of them are methoxylated: these methoxy groups are released as formaldehyde during degradation, which creates a second source of toxicity that can challenge standard heterotrophs. In an earlier DOE project, we discovered that some *Methylobacterium* strains grow exceptionally well on aromatics, and do not release formaldehyde into the medium from the methoxy groups present, unlike classic systems for aromatic degradation (*e.g.*, *Pseudomonas putida*). One reason methylotrophic bacteria may be well suited for utilizing methoxylated aromatics is their high capacity to produce and consume formaldehyde internally during growth on single-carbon compounds like methanol. Beyond this, however, both Marx and Martinez-Gomez have recently discovered that methylotrophs also have a complex formaldehyde stress response involving multiple sensors and response proteins, as well as the unexpected involvement of lanthanide-dependent dehydrogenases.

We have been developing *Methylorubrum* (formerly *Methylobacterium*) *extorquens* into a model system for degradation of methoxylated aromatics. Recently we have uncovered that both formaldehyde stress response sensors and the lanthanide-dependent systems are needed for effective vanillic acid utilization. In our current project, we discovered that the aromatic backbone converts acetyl-CoA and β-ketoadipate into acetoacetate and succinyl-CoA, this allows carbon to flow efficiently into the species' high-flux glyoxylate-regeneration pathway from

which our target product, butanol, is produced. An advantage of targeting butanol is that the same glyoxylate-regeneration pathway is naturally used to generate the internally-accumulated compound, poly-β-hydroxybutyrate (PHB). As such, we leverage PHB as a reporter, a visualizable single-cell proxy for production capacity through this pathway that can guide our design process.

We have found that a fundamental challenge to developing *M. extorquens* into a catalyst for conversion of lignin-derived methoxylated aromatics into butanol is that populations exhibit tremendous cell-to-cell variability in key phenotypes. We have discovered significant heterogeneity in growth, viability, PHB production, and even internal accumulation of lanthanide granules. We hypothesize that the physiological thresholds that lead to cell-to-cell variability will also exist for mutations or environmental variables that affect the same processes. Indeed, we have observed examples of both genetic interactions between beneficial mutations and with environmental parameters such as substrate concentration consistent with this hypothesis. This variability amongst cells and the fundamental role of stress responses indicate that stability of growth and production, and not just catalytic capacity, is paramount to develop effective growth and production from these difficult feedstocks.

Our project combines genome-scale approaches that will broadly identify the physiological hurdles cells face in growth and production from vanillic acid with a suite of single-cell approaches to examine phenotypic heterogeneity. In particular, a new optical method developed by Vasdekis to assay PHB in single, live cells will permit us to examine the correlations between growth, stress response, and production in an unprecedented manner. Using a model of heterogeneity as a guide, we can then combine our learnings from genetic underpinnings and their impact on single-cell physiology to develop improved strains of *M. extorquens* for conversion of vanillic acid to butanol. In the process, we will develop and demonstrate a novel approach that embraces phenotypic heterogeneity as a major source for future innovation in DOE-relevant biosystems design.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022318.

**Systems metabolic engineering of *Novosphingobium aromaticivorans* for lignin valorization**

Marco N. Allemann (allemannmn@ornl.gov),[1] Christopher C. Azubuike (azubuikecc@ornl.gov),[1] Gerald N. Presley,[1] Leah H. Burdick,[1] Richard J. Giannone,[1] David C. Garcia,[1] and **Joshua K. Michener**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

**Project Goals:** To engineer a non-model bacterium, *Novosphingobium aromaticivorans*, for valorization of depolymerized lignin to value-added bioproducts. The project involves (1) discovery and optimization of pathways for assimilation of lignin-derived aromatic compounds, (2) engineering conversion pathways that match the stoichiometry of aromatic catabolism, and (3) development of genome-scale mapping techniques to identify new engineering targets in non-model bacteria.

Lignin is one of the abundant renewable materials found in nature. This heterogeneous aromatic polymer is composed of a variety of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) monomers that are connected by diverse chemical linkages. Lignin valorization would improve biofuel economics, potentially through bacterial conversion of thermochemically depolymerized lignin into valuable bioproducts. *Novosphingobium aromaticivorans* F199 is an Alphaproteobacterium capable of degrading G, S, and H monomers and, due to its genetic tractability, is an emerging model organism for conversion of lignin-derived aromatic compounds. However, F199 cannot natively catabolize every component of depolymerized lignin, which limits conversion yields.[1]

We are identifying new aromatic degradation pathways to increase the catabolic potential of F199, using a combination of barcoded transposon insertion sequencing, proteomics, and *in vitro* biochemistry. We demonstrated this approach with the aromatic monomer syringate,[2] the β-1 linked dimer 1,2-diguaiacylpropane-1,3-diol (DGPD),[3] and, more recently, the monomer guaiacol. However, there are multiple aromatic compounds for which F199 lacks the necessary catabolic pathway. We have previously isolated additional Sphingomonads that metabolize several of these compounds and are currently identifying and characterizing the relevant pathways for transfer to F199. These new strains include a *Novosphingobium* isolate that can assimilate the β-β linked dimer pinoresinol and a *Sphingobium* isolate that is remarkably similar at the genetic level to *Sphingobium* sp. SYK-6 and can likewise assimilate the 5-5 linked dimer dehydrodivanillic acid (DDVA).

In addition to introducing new pathways, we are also optimizing native assimilation pathways in F199 to efficiently channel more lignin-derived carbon into central metabolism intermediates. Simultaneously, we are converting the resulting intermediates into value-added products, such as building blocks for bio-derived polymers. Finally, to better understand the effect of host genetic variation on pathway function, we are adapting a novel technique, bacterial quantitative trait locus (QTL) mapping, to F199. We have demonstrated intraspecific recombination between strain of *N. aromaticivorans* and are currently studying and optimizing this process. By combining novel pathway discovery, heterologous expression, and optimization, we are engineering *N. aromaticivorans* F199 to be more efficient at valorizing lignin-derived compounds.

*This work was supported by the U.S. Department of Energy Office of Science, through the Office of Biological and Environmental Research (BER) Early Career Research Program. Preliminary data was collected through the BioEnergy Science Center and Center for Bioenergy Innovation, Bioenergy Research Centers funded by BER.*

1. Azubuike, C. C., Allemann, M. N. & Michener, J. K. Microbial assimilation of lignin-derived aromatic compounds and conversion to value-added products. *Curr. Opin. Microbiol.* **65**, 64–72 (2022).

2. Cecil, J. H., Garcia, D. C., Giannone, R. J. & Michener, J. K. Rapid, Parallel Identification of Catabolism Pathways of Lignin-Derived Aromatic Compounds in Novosphingobium aromaticivorans. *Appl. Environ. Microbiol.* **84**, AEM.01185-18 (2018).

3. Presley, G. N. *et al.* Pathway discovery and engineering for cleavage of a β-1 lignin-derived biaryl compound. *Metab. Eng.* **65**, 1–10 (2021).

**Title:** Scalable Computational Tools For Inference Of Protein Annotation And Metabolic Models In Microbial Communities

**Authors:** Janaka N. Edirisinghe[1*] (janakae@anl.gov), Mikayla A. Borton[2*] (mikayla.borton@pnnl.gov), Michael Shaffer[3], Zahmeeth Sakkaff[1], Filipe Liu[1], Rory M. Flynn[3], Saeedeh Davoudi[4], Lucia S. Guatney[4,5], Derick Singleton[4], Tobby Lie[4], James C. Stegen[2], Byron C. Crump[6], Jim Davis[1], Farnoush Banaei-Kashani[4], Kelly C. Wrighton[3], Christopher Henry[1], and **Christopher S. Miller**[4]

**Institutions:** [1] Argonne National Laboratory, Argonne, IL; [2] Pacific Northwest National Laboratory, Richland, WA; [3] Colorado State University, Ft. Collins, CO; [4] University of Colorado Denver, Denver, CO; [5] University of Colorado Anschutz Medical Campus, Aurora, CO; and [6] Oregon State University, Corvallis, Oregon.

**Website URLs:** https://www.kbase.us ; https://github.com/WrightonLabCSU/DRAM

**Project Goals:** High-throughput omics technologies have made the assembly of microbial genomes recovered from the environment routine. Computational inference of the protein products encoded by these genomes, and the associated biochemical functions, should allow for the accurate prediction and modeling of microbial metabolism, organismal interactions, and ecosystem processes. However, a lack of scalable, probabilistic protein annotation tools limits the full potential of metabolic modeling. Our approach to inference of improved models relies on developing new computational tools in three main areas: 1) improved protein annotations, 2) iterative cycles of gap-filling metabolic models with improved protein annotations and informing probabilistic protein annotations based on metabolic models, and 3) integrating improved protein annotations with community-level flux balance metabolic models. We aim to make these tools broadly accessible via the DOE Systems Biology Knowledgebase (KBase)[1].

**Abstract Text:** To improve the inference of protein annotations, we have made improvements over the past year to the comprehensive annotation pipeline, Distilled and Refined Annotation of Metabolism (DRAM)[2]. We have significantly improved DRAM to incorporate new functionality, bug fixes, and speed gains, to improve the speed and sensitivity of viral annotations, to allow for custom HMM sets, and to expand the metabolic repertoire (e.g. polymeric carbons / polyphenols, carbohydrates, organic nitrogen / methylamines, and bile salts). We have also improved annotation precision by including gene-customized cutoffs for specific functional genes that are problematically annotated using homology, as well as visual validation with pre-constructed phylogenetic trees. In summary, DRAM now provides faster, phylogenetically informed community profiles and genome annotations.

We have also improved the model reconstruction pipelines in KBase to directly utilize improved annotations from the DRAM KBase app[3]. Metabolic models are now capable of representing many ecologically important metabolisms, including methanogenesis, nitrogen cycling, and sulfur reduction. DRAM annotations of viral genomes are now in the KBase infrastructure and integrated to automatically ingest outputs from other virus-specific applications on KBase (e.g.VirSorter). Ultimately, our improvements to DRAM alongside tight KBase integration will lead to more accurate community metabolic models of microbiome systems.

We are also evaluating these improved annotations and model reconstructions with experimental data. We computationally predicted growth of 350 genomes on 60 carbon sources, comparing the predicted phenotypes with experimentally observed data from Biolog studies. Although DRAM and RAST individually performed with similar accuracy, RAST displayed more false negatives while DRAM displayed more false positives in comparison with experimental data. However, a gap-filling approach *combining* RAST and DRAM annotations boosted accuracy dramatically, while also identifying gene candidates for many gapfilled reactions. This gapfilling approach requires *a priori* knowledge of phenotypes to work; to provide this knowledge for a much broader set of genomes (including MAGs), we also developed machine learning classifiers to predict phenotypes based on annotated gene functions genome-wide.

Now that we have an established cyberinfrastructure with integration of high-quality annotations and models, we can next evaluate the scalability and performance using more complex datasets. As a use case, we demonstrate these improvements to genome annotation and modeling using Metagenome Assembled Genomes (MAGs) from the Genome Resolved Open Watersheds database (GROWdb). This growing resource samples a large number (currently 250) of river microbiomes worldwide, and offers a metadata-rich, complex, real-world dataset with which to evaluate the utility and scalability of our annotation and modeling efforts in KBase. We have ingested a snapshot of 163 samples from US surface waters into KBase, including genome-resolved metagenomics, extensive geospatial metadata, metatranscriptomics, and metabolomics (FT-ICR). To inform metabolic modeling, we used DRAM to genomically inventory the 2,093 nonredundant surface-water-derived MAGs for complete energy biosynthesis systems and analyze these results organized by phylogeny and geospatial metadata.

Finally, over the last year, we have focused on developing improved representations for biological sequence data by adapting universal language models from the Natural Language Processing (NLP) community for two specific applications: protein annotation and taxonomic binning. We are currently extending the ProteinBERT model[4] to 1) use the primary protein sequences to derive representations for the sequences in an unsupervised manner, and 2) utilize Enzyme Commission (E.C.) annotations for supervised training of the model to capture global representations of function. These models can be extended with other global features, such as E.C. numbers of adjacent genes, or inferred 3D protein structure classes. Encouraged by the improvements to metabolic modeling we have shown above when using multiple annotation sources, we will add an additional, orthogonal NLP-based annotation using these approaches.

**References:**


1. Arkin AP, Cottingham RW, Henry CS, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol*. 2018;36:566–9.
2. Shaffer M, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res*. 2020;48:8883–900.
3. Shaffer M, et al. kb_DRAM: Distilled genome annotations and metabolic modeling in KBase. *submitted*
4. Brandes N, et al. ProteinBERT: A universal deep-learning model of protein sequence and function. *bioRxiv.* 2021; 2021.05.24.445464.



**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021350, and the DOE Joint Genome Institute Community Science Program. KBase was supported by Award Numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

# HypoNPAtlas: an Atlas of hypothetical natural product for mass spectrometry database search

Yi-Yuan Lee[1], Mustafa Guler[1], Neel Mittal[1], Cameron Miller[1], Benjamin Krummenacher[1], Haodong Liu[1], Liu Cao[1*], Aditya Kannan[1], Keshav Narayan[1], Samuel T Slocum[3], Bryan L Roth[3], Alexey Gurevich[4], Roland Kersten[2], Bahar Behsaz[1], and **Hosein Mohimani**[1]

[1]Computational Biology Department, Carnegie Mellon University, Pittsburgh, USA
[2]College of Pharmacy, University of Michigan, Ann Arbor, USA
[3]Department of Pharmacology, University of North Carolina, Chapel Hill, USA
[4]Center for Algorithmic Biotechnology, Institute of Translational Biomedicine, St. Petersburg State University, St. Petersburg, Russia

http://mohimanilab.cbd.cmu.edu/

## Project Goals

Recent analysis of hundreds of thousands of public microbial genomes has resulted in the discovery of over a million biosynthetic gene clusters (BGCs). However, the connection of these BGCs to their molecular products has not kept pace with the speed of microbial genome sequencing. Currently, the molecular products for the majority of BGCs remain unknown. Global natural product social (GNPS) molecular networking infrastructure harbors billions of mass spectra of natural products with unknown structures and BGCs. In order to bridge the gap between large scale genome mining and mass spectral datasets for natural product discovery, we developed hypoNPAtlas, an Atlas of hypothetical natural product structures, which can be readily used for *in silico* database search of tandem mass spectra. HypoNPAtlas is constructed by mining the genomes of 22,671 microbial strains from the RefSeq database using seq2ripp, a novel machine learning tool for prediction of ribosomally synthesized and post-translationally modified peptides (RiPPs). Searching the hypothetical molecules from our Atlas against 46 mass spectral datasets from GNPS resulted in the discovery of numerous RiPPs, including two novel lassopeptides and one lanthipeptide from *Streptomyces* sp. NRRL B-2660, WC-3904 and WC-3560. Moreover, seq2ripp discovered two plant RiPPs from *Oryza sativa* and *Elaeagnus pungens* with a novel post-translational modification (PTM) [1].

## References

[1] Yi-Yuan Lee, Mustafa Guler, Neel Mittal, Cameron Miller, Benjamin Krummenacher, Haodong Liu, Liu Cao, Aditya Kannan, Keshav Narayan, Samuel T Slocum, Bryan L Roth, Alexey Gurevich, Roland Kersten, Bahar Behsaz, and Hosein Mohimani “HypoNPAtlas: an Atlas of hypothetical natural product for mass spectrometry database search”. Under review.

**Title:** Engineering novel microbes for upcycling waste plastic and solving climate crisis

**Authors:** Jinjin Diao, Yifeng Hu, and **Tae Seok Moon*** (tsmoon@wustl.edu)

**Institutions:** Washington University in St. Louis

**Website URL:** https://sites.wustl.edu/moonlab/

**Project Goals:** The goal is to develop a consolidated biological process to upcycle waste polyethylene terephthalate.

**Abstract Text:**

The use of petroleum-based products has negative impacts on our planet, leading to climate crisis. Polyethylene terephthalate (PET) represents 8% of global solid waste. PET chemical recycling has been an option to solve this global problem, but it suffers from its relatively high process cost and the extremely low price of virgin PET. One solution to address this issue is to upcycle waste PET rather than recycle it to generate the same PET typically with low quality. PET upcycling can be achieved by depolymerizing PET into terephthalic acid (TPA) and ethylene glycol (EG) and biologically converting these monomers into value-added products. However, there are only a handful of reports demonstrating microbes capable of growing on both TPA and EG generated from PET as sole carbon sources. To overcome this limitation, we have performed strain screening to discover a *Rhodococcus* strain RPET that can grow well on the alkaline hydrolysis products of PET as the sole carbon source without any purification step. Notably, this strain can grow on a mixture of TPA and EG at extremely high concentrations (up to 0.6M) and high osmolarity resulting from alkaline hydrolysis and pH neutralization. The resultant media supported RPET's growth without any purification and sterilization step except for their dilution to make up to 0.6M of monomer concentrations. In addition, many synthetic biology tools, developed for a related species *Rhodococcus opacus* (1,2), were functional in RPET, facilitating its engineering. We will discuss our effort to develop this novel chassis for waste PET valorization with PET conversion into carotenoids and muconate as two demonstration products (3).

1. DeLorenzo, Rottinghaus, Henson, Moon. 2018. ACS Synthetic Biology 7:727-38

2. DeLorenzo, Diao, Carr, Hu, Moon. 2021. An improved CRISPR interference tool to engineer *Rhodococcus opacus*. ACS Synthetic Biology 10:786-98

3. Moon et al. 2021. U.S. Provisional Patent Application

**Key words:** climate crisis; polyethylene terephthalate; plastic upcycling; synthetic biology; metabolic engineering

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022003.

**Spatial plasticity in plant-microbe interactions in response to applied nutrient heterogeneity in soil**

**James Moran**[1,*] (james.moran@pnnl.gov), Vivian Lin[1], Elias Zegeye[1], Jamie Nuñez[1], Sam Purvine[1], Nikola Tolic[1], Kim Hixson[1], Montana Smith[1], Nicholas Sconzo[1], Nicholas Huggett[1], Young-Mo Kim[1], Mary Lipton[1]

[1]Pacific Northwest National Laboratory, Richland, Washington
https://science.pnl.gov/staff/staff_info.asp?staff_num=8559

**Project Goals: This project seeks to elucidate key microbiological and geochemical controls on nutrient exchange within the rhizosphere and the role that spatial organization within the root-rhizosphere-soil continuum plays in directing nutrient acquisition by the host plant. Spatially resolved understanding of nutrient exchange through this dynamic zone will identify key variables that may form part of an effective rhizosphere management program targeting enhanced plant productivity. Our aims are directed towards identifying the microbial and geochemical factors that stimulate enhanced plant investment (in the form of root exudation) into specific regions of the rhizosphere and assessing the implications of this carbon input on the microbial and geochemical response.**

While small in physical stature, rhizosphere embodies the dynamic interface between plants and soil and can impart profound impacts on overall plant performance. Rhizosphere development is spurred by carbon delivery into the subsurface through a variety of plant processes collectively termed rhizodeposition. In this work, we are probing the ability of plants to spatial focus their subsurface carbon delivery in response to the heterogenous distribution of nutrients typical of soils. Rhizodeposition can be viewed as an investment by the plant, used to establish beneficial microbial and geochemical relationships to enable more effective nutrient extraction from soil and to help maintain defenses against pathogens and environmental stresses. The intense spatial heterogeneity in nutrient distributions in soil exemplifies the potential advantage conferred to a plant by spatially focusing its carbon investment at specific location conducive to highest nutrient extraction/return. Previous studies demonstrated a phenomenon of root proliferation in even small zones of increased nutrient availability within soil. Here, we are applying a multi-faceted approach to attempt capturing the functional plasticity of roots themselves, the resulting implications on the rhizosphere microbiome, and the propagation of these effects on plant growth. As a test platform, we cultured switchgrass (*Panicum virgatum*, var. Cave-in-rock) under marginal conditions in rhizoboxes containing a mixture of sand (as a low-nutrient growth medium to establish marginal growth conditions) and soil harvested from the Kellogg Biological Station (Hickory Corners, MI; to serve as a microbial inoculum and nutrient source). Within these rhizoboxes, we established a thin (approximately 1 cm in vertical thickness) horizon of soil amended with various nutrient resources included organic/inorganic

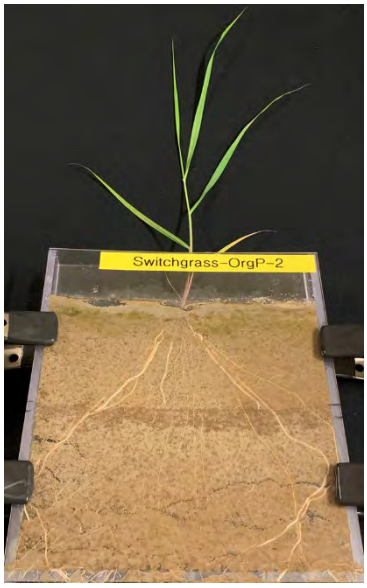



*Figure 1: Sample set up utilized a series of rhizoboxes having a horizontal resource amendment midway through the rooting zone as seen above in the darkened soil band.*

phosphorus, chitin, or a combination of organic phosphorus and chitin. We applied a a stable isotope tracer ($^{13}CO_2$) in concert with bulk isotope analysis, leveraged proteomic analysis to identify microbial uptake of plant-derived carbon, applied a chitinase enzyme mapping technique to spatially map enzyme activity, and tracked carbon transformations. Combined, we attempted to quantify the resulting plant-microbe adaptation to the applied nutrient heterogeneity.

We observed increased plant biomass in the experimental replicates containing a thin horizon of chitin-amended soil as recorded through larger aboveground biomass production. While we leveraged our $^{13}CO_2$ tracer to observe root exudation in all systems containing a plant, we observed the highest rates of root exudation specifically in the chitin-amended samples. Most importantly, in agreement with our hypothesis related to the ability of plants to spatially focus their carbon deposition, the increased root exudation was spatially focused within the chitin amendment zone and dropped off above and below this layer. The spatial focusing of root exudation was not observed in our controls nor in systems amended with various forms of phosphorus, suggesting this phenomenon was motivated by the presence of chitin, likely as a source of nitrogen.

To leverage increased carbon supply toward improved nutrient extraction from the amended chitin, it needs to first be broken down to release its organic nitrogen. We developed a fluorescent-based mapping approach for tracking chitinase activity in these systems. Briefly, we removed the rhizobox side panel and performed a membrane blotting technique to extract chitinases from the soil. We synthesized and applied a molecular chitin surrogate containing a fluorescent probe which becomes activated upon reaction with chitinase. Importantly, this approach more closely tracks chitinase potential activity versus quantifying in situ rates. Still, we observed a statistical increase in chitinase activity spatially focused within the amended band. We are following up on these observations by performing a more thorough proteomic analysis with specific attention to the exchange of plant derived carbon (here containing a $^{13}C$ tracer) into proteins associated with various microbial taxa to assess the impact of the observed changes on the rhizosphere microbiome.

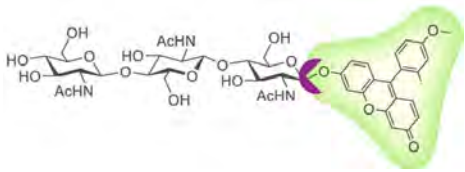



*Figure 2: We developed a molecular chitin surrogate that, when cleaved by a chitinase enzyme, activates a fluorescent tag for spatially tracking chitinase activity.*

Taken together, we sought to test the hypothesis that plant roots can spatially regulate their interactions with soil to target regions with higher potential for resource extraction. When grown in simulated marginal growth media, we observed spatial focusing of switchgrass root exudates and associated chitinase activity within a thin band of soil amended with chitin and that these plants outperformed experimental control cohorts. We are furthering this study to explore selective shifts in plant-microbe carbon transfers accompanying these spatially applied nutrient amendments to gain insights to the mechanisms supporting spatially driven subsurface responses. Understanding the adaptability and drivers that guide subsurface carbon flow can help contribute to better plant performance in marginal soils and inform future efforts to direct carbon deposition in soil systems to specific locations or horizons.

*This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER). This contribution originates from an Early Career Research Award granted at the Pacific Northwest National Laboratory (PNNL).*

**Environmental and Engineered Factors Influence Membrane Features of Shale Taxa**

Chika Ugwuodo[1*] (chikajude.ugwuodo@unh.edu), Yuxiang Shen,[2] Fabrizio Colosimo,[3] Appala Raju Badireddy,[2] and **Paula J. Mouser[1]**

[1]University of New Hampshire, Durham; [2]University of Vermont, Burlington; and [3]New England Biolabs, Ipswich, Massachusetts

**Project Goals:**

**The overall aim of this project is to advance our understanding of the microscopic complexity of non-sterile engineered hydrocarbon systems as they evolve from completion into mature use. Here, we characterize the effects of salinity and hydraulic retention time (HRT) on membrane features of persistent shale taxa.**

Deep subsurface shale reservoirs are increasingly being engineered in the United States and globally using unconventional techniques including horizontal drilling and hydraulic fracturing, to meet rising demands for natural gas which is a cleaner alternative energy source to other fossil fuels. Halotolerant anaerobic microbial communities, dominated by *Halanaerobium*, colonize fractured shale and adapt to extreme subsurface conditions such as anoxia, elevated pressure, brine-level salinities and high temperatures. Despite the role these microorganisms play in altering biogeochemical reactions, effective energy capture, and well longevity, *in situ* growth kinetics and activities, including interactions with fluids and shale matrices are poorly understood. The microbial membrane protects the cell from external stressors and mediates critical physiologies, including transport, metabolism, aggregation and cell-surface interactions. Membrane functions are associated with the activities of peripheral and integral proteins, which in turn depend on biophysical properties such as bilayer symmetry, viscosity, curvature, elasticity and thickness. These properties are collectively dictated, to a large extent, by the membrane lipidome comprised of polar head groups and hydrophobic fatty acid tails.

For *Halanaerobium* and other persistent microbial taxa of fractured shale, salinity and hydraulic retention time (HRT) are important perturbants of cell membrane structure. Hence, we used a suite of analytical techniques including gas chromatography-mass spectrometry (GC-MS), dark-field hyperspectral imaging, dynamic light scattering and atomic force microscopy to investigate the effects of salinity (7%, 13% and 20% NaCl) and HRT (19.2 h, 24 h and 48 h) on membrane fatty acid composition and mechanics of *Halanaerobium congolense* WG10 and mixed enrichment cultures from hydraulically fractured wells in West Virginia. For these experiments, cultures were grown in chemostat vessels operated in continuous flow mode under strict anoxia and constant stirring. Our findings show that salinity and HRT induce changes in membrane fatty acid chemistry and biophysical properties of *H. congolense* WG10 in distinct and complementary ways. Notably, under suboptimal salt concentrations (7% and 20% NaCl), *H. congolense* WG10 elevates the proportion of polyunsaturated fatty acids (PUFAs) in its membrane, and this results in an apparent increase in fluidity (homeoviscous adaptation principle) and an experimentally-determined decrease in bilayer elasticity (Figure 1). Taken together, these results provide insights into our understanding of how environmental and engineered factors might disrupt the physical and biogeochemical equilibria of fractured shale by inducing physiologically relevant changes in the membrane of persistent microbial taxa.

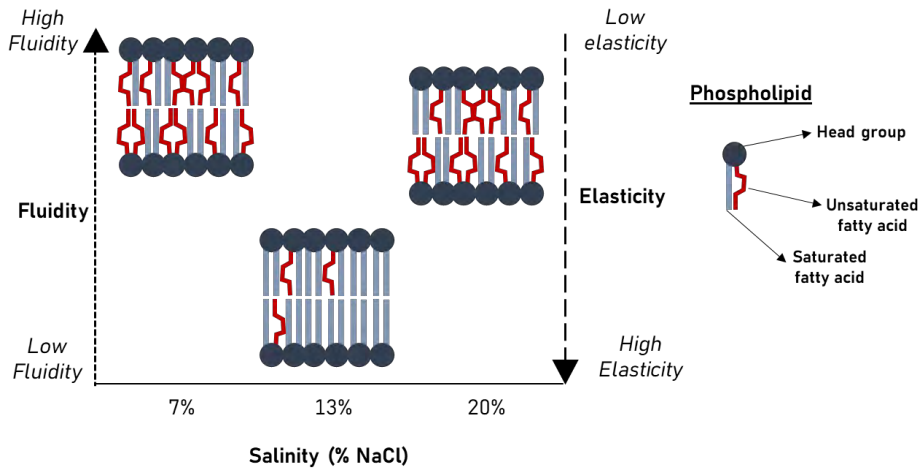



**Figure 1.** Salinity induces changes in membrane fluidity and elasticity of *Halanaerobium congolense* WG10.

**Funding Statement:** This research was funded by the U.S. Department of Energy, EPSCoR, Office of Science, Office of Biological and Environmental Research (BER), grant no. 181078 UDCEPM 710D51 020.

**Title:** Design, Detect, Evolve: Engineering Syringol Degradation in ADP1

**Authors:** Alyssa Baugh[1]* (alyssabaugh@uga.edu), Ramesh K. Jha,[2] and **Ellen L. Neidle[1]**

**Institutions:** [1]University of Georgia, Athens; and [2]Los Alamos National Laboratory, Los Alamos, New Mexico

**Project Goals:** Our long-term objective is to exploit bacterial aromatic compound metabolism for biotechnology. Applications range from lignin valorization to the degradation of environmental pollutants. A critical first step is to expand natural pathways using foreign and modified genes to create synthetic pathways with high conversion efficiency. A soil bacterium, *Acinetobacter baylyi* ADP1, is being used in this project to create a novel route for the catabolism of syringol, an aromatic component derived from lignin pyrolysis. To enable syringol consumption, for which there is no known metabolic route, we are combining parts of characterized pathways. Biosensors are being developed to facilitate the modification of enzyme substrate specificity, and laboratory evolution is being used for growth-based selection. The genetic malleability and catabolic versatility of this strain make it an ideal host for synthetic pathways. New catabolic functions can be used in ADP1 and/or ported for use in other organisms.

**Abstract Text:** Lignin is a vastly underutilized renewable resource. Initial processing yields a heterogenous mixture of compounds, including many aromatics. One approach to produce commercially valuable compounds from a lignin-derived mixture is to funnel as much carbon as possible through bacterial central metabolism and then into a desired product. To increase the consumption of lignin-derived mixtures by *A. baylyi* ADP1, our goal is to expand its repertoire for using lignin-derived aromatics. There is no known route for biological degradation of syringol. The development of a novel route for syringol degradation would contribute to lignin valorization and to a generalizable method to create catabolic modules. A three-part pathway was designed with candidate enzymes (indicated for each step, Fig. 1).

Multiple strains were constructed with foreign/synthetic DNA integrated in the ADP1 chromosome. To obtain a functional pathway, we are using two approaches: (1) a method for growth-based selection and evolution and (2) a method for improvement based on the ability of biosensors to detect pathway intermediates independently from growth.

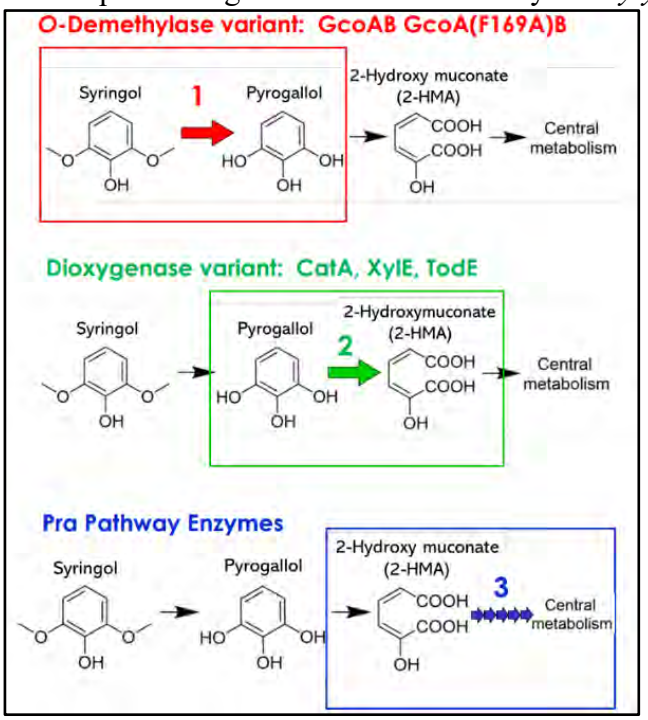



**Fig. 1.** Synthetic pathway for syringol catabolism

The first approach involves selection for growth on syringol during adaptive evolution with a method called EASy (Evolution by Amplification and Synthetic Biology) [1]. This method specifically amplifies a chromosomal region of ADP1, leading to a tandem array of that region. Multiple gene copies increase enzyme expression, thereby enhancing catalytic activity of a non-optimal enzyme. In addition, gene duplication provides a larger segment of DNA for the occurrence of advantageous mutations. In short, EASy is a way to offer an advantage to engineered strains prior to selective growth. Unlike in previous applications, I sought to amplify two different chromosomal regions of the same strain. The goal of this novel use of the method is to increase the number of copies of genes encoding separate parts of the pathway independently to prevent the accumulation of toxic intermediates. I successfully accomplished the simultaneous and independent EASy amplification of two chromosomal regions. Numerous amplified strains were constructed, and qPCR confirmed a wide range of the number of genes and different ratios of gene dosage in the two regions. These strains provide the starting cultures for laboratory evolution.

Since growth is an imperfect metric for enzyme activity, a parallel approach is being used to create this pathway: biosensor-based detection of pathway intermediates. Strains were constructed to express an O-demethylase variant GcoA(F169)B that demethylates syringol [2] (Fig. 1). To optimize this activity in our pathway, a biosensor is being developed to detect the reaction product, pyrogallol. For this purpose, we successfully respecified a biosensor that detects a similar compound [3] to detect pyrogallol in *E. coli*. An approximately two-fold response to pyrogallol was observed with no induction by its native inducer. Current work will improve the range of fluorescent response to pyrogallol, and this biosensor will be introduced into ADP1.

Initial successes with both approaches provide a foundation for continuation. Expansion of the "traditional" EASy method allows us to select the optimum ratio of initial gene expression in a pathway. Strains with various gene dosages are being used in experiments to select for growth on syringol. To complement this approach, we will improve biosensors for the detection of pyrogallol, allowing us to perform directed evolution of strains based on pathway flux rather than growth.

**References/Publications**


1. Tumen-Velasquez M, et al. 2018. *Proc Natl Acad Sci USA* 115: 7105-7110.
2. Machovina MM, et al. 2019. *Proc Natl Acad Sci* 116, 13970.
3. Jha RK, et al. 2015. *Proteins: Structure, Function, and Bioinformatics* 83, 1327-40.



**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE- SC0022220.
Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). This program is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) under FWP LANLF32A.
A.B. received support for some preliminary work from the DOE Office of Science Graduate Student Research (SCGSR) program.

**Title**: Snekmer – A Tool for Protein Classification Using Amino Acid Recoding and Kmer Analysis

**Authors**: Jason McDermott, [1] Christine Chang, [1] Abigail Jerger, [1] Robert Egbert, [1] and **William C. Nelson**[1*] (william.nelson@pnnl.gov)

**Institutions**: [1]Pacific Northwest National Laboratory, Richland, WA

**Website URL:** https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

**Project Goals:** The Pacific Northwest National Laboratory Persistence Control Scientific Focus Area is focused on developing fundamental understanding of factors governing the persistence of engineered microbial functions in rhizosphere environments. From this understanding, we will establish design principles to control the environmental niche of native rhizosphere microbes. In our first funding period, we are examining the efficacy of genome reduction and metabolic addiction to plant root exudates in environmental isolates as persistence control strategies using the bioenergy crop sorghum and defined microbial communities as a model ecosystem. Effective persistence control will lead to secure plant–microbe biosystems that promote stress-tolerant and highly productive biomass crops. Our supplemental KBase collaboration project to develop a protein functional annotation tool and reference database will support robust evaluation and modeling of SFA project systems and be broadly applicable to general microbiome research.

**Abstract Text:** Advances in sampling and sequencing technologies have led to rapid expansion of the size of microbial sequence databases and the diversity they represent. Existing functional annotation tools are not built to scale with current or anticipated large datasets and are increasingly hampered in their ability to perform analyses due to the time or hardware requirements. Annotation resources based on protein families can accelerate searches but require curation during construction and maintenance as additional sequence diversity is discovered. Should funding to support these tasks be lost, these resources will not keep pace with sequence information. We propose that a process compressing protein sequence space and constructing universal protein families would be a scalable and computationally efficient solution to this grand challenge in computational biology, capable of accurately predicting membership of unobserved sequence diversity.

Here we present Snekmer, a Python tool which constructs protein family models based on short subsequences of length *k* (kmer) profiles of protein sequences recoded into reduced-complexity character sets. The program is configurable by recoding scheme and kmer length. Snekmer has two modes of operation: 1) modeling a user-provided protein family and searching for members of the family from a collection of sequences; 2) unsupervised clustering of an input dataset and construction of protein family models for each identified cluster. In each mode, the sequences provided are first recoded into a user-specified reduced character alphabet. Several alphabets of varying complexity are available. The alphabets group amino acids with similar chemical properties and/or high observed frequency of substitution. The proteins are then divided into kmers – subsequences of length *k* – and represented as vectors of kmer composition (simple binary presence/absence). Snekmer machine learning algorithms input the vectors for supervised classification model construction or unsupervised clustering. We will deploy the code under an

open-source license on github and implement Snekmer as an app in the DOE Systems Biology Knowledgebase (KBase).

We are exploring the utility of Snekmer for the Persistence Control SFA by modeling proteins hypothesized to be important for catabolizing sorgoleone, a compound exuded by sorghum plants which is an attractive substrate to demonstrate persistence control through metabolic addiction. The models are being used to search bulk soil and rhizosphere microbiome metagenomic data sets to determine distribution of these genes in communities under different selective conditions.

**Funding Statement**: This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP) and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design Science Focus Area "Persistence Control of Engineered Functions in Complex Soil Microbiomes". A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by BER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.

# Metabolite Excretion and Metabolic Flux Analysis in *Chromochloris zofingiensis*, an Emerging Model Green Alga for Sustainable Fuel Production

Yuntao Hu[1*](yuntaohu@lbl.gov), Michelle F. Meagher[2], Melissa S. Roth[3], Alex J. Metcalf[2], Trent R. Northen[1,4], Nanette R. Boyle[2], and **Krishna Niyogi[3]**

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Colorado School of Mines, Golden CO; [3]University of California, Berkeley, CA; [4]Joint Genome Institute, Berkeley, CA; [4]

**https://sites.google.com/view/czofingiensis/home**

**Project Goals: Our overarching research goal is to engineer the green alga *Chromochloris zofingiensis* for the production of biofuels. Our strategy involves large-scale multi-omics systems analysis to understand the genomic basis for energy metabolism partitioning as a consequence of carbon source. Enabled by cutting-edge synthetic biology and genome-editing tools, we will integrate the systems data in a predictive model that will guide the redesign and engineering of metabolism in *C. zofingiensis*.**

*C. zofingiensis* is an emerging model system for the production of biofuels and bioproducts. It is an especially attractive system because it produces astaxanthin along with a large amount of lipids. Astaxanthin is a high value product (~$7,000 per kilogram) with uses in the pharmaceutical, nutraceutical, and cosmetic industries. It also demonstrates high levels of triacylglycerol accumulation and low photosynthetic productivity when additional organic carbon sources are provided[1], making it ideal for metabolic or genetic engineering focused on increasing algal lipid production.

First, we created a genome-scale metabolic model of this organism, iCre1925[2], using a new computational algorithm, Rapid Annotation of Photosynthetic Systems[3] (RAPS). The results of flux balance analysis (FBA) studies conducted using this model predicted the excretion of lactate, among other products when *C. zofingiensis* is grown on glucose. This prediction was later affirmed with experimental data on extracellular metabolites in *C. zofingiensis* cultures.

To investigate intracellular flux distributions during *C. zofingiensis* growth on glucose, an isotopically assisted metabolic flux analysis (MFA) experiment was performed. A model of central metabolism was created by pulling 140 reactions from iCre1925. Atom transitions were assigned for each reaction, and this network was then simplified by combining series of linear reactions to ease computational load in running simulations. A combination of $^{13}C$ labeled glucose tracers was chosen for this experiment based on isotopic tracer simulations performed on central metabolism. *C. zofingiensis* cultures were grown in the presence of this labeled substrate until isotopic steady state was reached, at which point the culture was harvested and samples quenched for metabolite analysis. Amino acids, organic acids, and sugars were derivatized from labeled biomass and analyzed via GC-MS to determine mass isotopomer distributions (MID) for relevant fragments. This MID data was used along with experimentally determined uptake and excretion fluxes and the central metabolic network model to calculate intracellular fluxes using the isotopomer network compartmental analysis (INCA) software package. Results of this analysis indicate that glucose is transported to the plastid after uptake and is either directed into starch biosynthesis or

utilized by the pentose phosphate pathway (PPP). This high flux through the PPP seems to indicate a need for reducing equivalents in the plastid, which is consistent with the high energy requirements for astaxanthin synthesis. Future MFA studies will be needed to continue probing intracellular flux distributions under various perturbations.

To generate additional phenotypic data for model refinement, we conducted a *C. zofingiensis* time-course experiment, by analyzing its metabolome and measuring alterations in media composition resulting from algal growth. During the heterotrophic growth of *C. zofingiensis*, we observed a decreased concentration of intermediates of the tricarboxylic acid cycle and amino acids, indicating a possible redirection of energy flux toward triglyceride synthesis. Our results showed that *C. zofingiensis* secreted more diverse exometabolites during its heterotrophic growth stages when supplemented with glucose. We find that *C. zofingiensis* has a clear nutrient (carbon/nitrogen source) preference order and recaptures secreted metabolites when the exogenous glucose is limited. To examine this observed diauxic growth in more detail, we screened 15 carbohydrates, including pentose, hexose, disaccharide, trisaccharide, to investigate their abilities to suppress algae photosynthesis and support heterotrophic growth. From this found that four hexoses: glucose, fructose, galactose, and mannose are able to serve as organic carbon sources for heterotrophic growth yet not all of these sugars were observed to inhibit algal photosynthesis.

The metabolic model that has been developed, used in combination with this extensive data set, has great potential to provide insights into this organism's metabolism and elucidate dramatic metabolic shifts within the organism. This will enable informed strain engineering strategies to maximize lipid productivity in this organism.

**References**

1. Roth, M. S., Gallaher, S. D., Westcott, D. J., Iwai, M., Louie, K. B., Mueller, M., ... & Niyogi, K. K. (2019). Regulation of oxygenic photosynthesis during trophic transitions in the green alga Chromochloris zofingiensis. The Plant Cell, 31(3), 579-601.

2. Meagher, M., Metcalf, A., Ramsey, S. A., Prentice, W., & Boyle, N. R. (2021). Genome-Scale Metabolic Model Accurately Predicts Fermentation of Glucose by Chromochloris zofingiensis. bioRxiv, 2021.2006.2022.449518. doi:10.1101/2021.06.22.449518

3. Metcalf, A. J., Nagygyor, A., & Boyle, N. R. (2020). Rapid Annotation of Photosynthetic Systems (RAPS): automated algorithm to generate genome-scale metabolic networks from algal genomes. Algal Research, 50, 101967.

*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018301*

Mineral Nutrition and Energy Pathways Regulate Lipid Accumulation and Photosynthesis

Tim L. Jeffers*[1] (tim_jeffers@berkeley.edu), Shivani Upadhyaya*[1] (shivani.u@berkeley.edu), Kim Hixson,[2] Sam Purvine,[2] Carrie Nicora,[2] Ryan McCombs,[1] Mary Lipton,[2] **Krishna K. Niyogi**,[1,3,4] Melissa S. Roth[1]

[1]Department of Plant and Microbial Biology, University of California, Berkeley; [2]Pacific Northwest National Laboratory; [3]Lawrence Berkeley National Laboratory; [4]Howard Hughes Medical Institute, University of California, Berkeley

**Project Website:** https://sites.google.com/view/czofingiensis/home

**Project Goals: Our overarching research goal is to design and engineer high-level production of biofuel precursors in photoautotrophic cells of the unicellular green alga *Chromochloris zofingiensis*. Our strategy involves using large-scale multi-'omics systems analysis to understand and model the genomic basis for how the energy metabolism of the cell is redirected based on the carbon source. Here, we focus on elucidating the effects of nutrients and the master regulator target of rapamycin (TOR) kinase on regulation of photosynthesis and metabolism to ultimately improve production of biofuels and bioproducts.**

Microalgae have the potential to become a major source of biofuels and bioproducts without exacerbating environmental problems. Photosynthetic microbes can utilize solar energy, grow quickly, consume $CO_2$, and be cultivated on non-arable land. However, there are presently considerable practical limitations in the photosynthetic production of biofuels from microalgae, resulting in low productivity and high costs. Insight into regulation of photosynthesis and metabolism will accelerate bioengineering of microalgae to maximize production of biofuels and bioproducts.

Photosynthetic organisms respond to multiple signals to produce, use, and store energy in diverse environments. We have exploited proteomics in the oleaginous green alga *Chromochloris zofingiensis* to understand the influence of essential nutrients and master regulators of energy pathways on photosynthesis and the accumulation of biofuel precursors. Previously, we showed that glucose (Glc) sensing and signaling via hexokinase (HXK1) increases growth, enhances triacylglycerol accumulation (TAGs), and completely switches off photosynthesis (1, 2). However, here, we found that the switching off of photosynthesis only occurs under iron deficiency. Multifactorial proteomics revealed that *C. zofingiensis* switches off photosynthesis to prioritize iron resources for respiration and *de novo* fatty acid synthesis. Our combinatorial iron and Glc experimental design effectively identified novel candidates involved in photosynthetic regulation, lipid accumulation, and iron partitioning across algae and plants.

Furthermore, we investigated the role of TOR kinase in Glc perception in *C. zofingiensis*. In the plant *Arabidopsis thaliana*, Glc sensing and signaling pathways are regulated by TOR kinase, and together they are important for the regulation of growth and metabolism. In algae, in contrast, the role of TOR kinase in glucose signaling and regulation of photosynthesis is unknown. Using the TOR kinase domain inhibitor AZD8055, we found that TOR inhibition causes inhibition of photosynthesis and accumulation of TAG and ketocarotenoids in *C. zofingiensis*, a phenotype similar to the Glc-mediated switch. However, TOR inhibition in the presence of glucose prevents biomass accumulation, whereas glucose on its own promotes accumulation of biomass. Using a phosphoproteomic approach, we identified novel targets of TOR kinase, whose phosphorylation changes in response to TOR kinase inhibition by AZD. Intriguingly, many proteins involved in TAG degradation appear to be targets of phosphorylation by TOR kinase. We also found that inhibition of TOR kinase resulted in mobilization of major membrane lipids into TAGs and not *de novo* TAG biosynthesis as was observed upon Glc addition. Altogether, these studies show the power of using *C. zofingiensis* for gene discovery and pathway elucidation related to nutrient physiology, photosynthesis, and carbon flow regulation, which will enable metabolic redesign and bioengineering for sustainable biofuels.

**References**


1. Roth MS, Gallaher SD, Westcott DJ, Iwai M, Louie KB, Mueller M, Walter A, Foflonker F, Bowen BP, Ataii NN, Song J, Chen J-H, Blaby-Haas CE, Larabell C, Auer M, Northen TR, Merchant SS, Niyogi KK (2019) Regulation of oxygenic photosynthesis during trophic transitions in the green alga *Chromochloris zofingiensis*. *Plant Cell* 31: 579-601.
2. Roth MS, Westcott DJ, Iwai M, Niyogi KK (2019) Hexokinase is necessary for glucose-mediated photosynthesis repression and lipid accumulation in a green alga. *Communications Biology* 2: 347.



*This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under Award Number DE-SC0018301.*

# Synthetic Biology in Oleaginous Green Algae

Jeffrey L. Moseley[1]* (jlmoseley@berkeley.edu), Dimitrios Camacho[1], Sean D. Gallaher[1] and **Sabeeha S. Merchant**[1]

[1]Quantitative Biosciences Institute, Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720

**Project Goals:** The fundamental research goal of this project is to design and engineer photoautotrophic green algae for scalable production of biofuel precursors and higher value bioproducts. Results from a series of "multi-omic"systems analyses of the model Chlorophyte, *Chlamydomonas reinhardtii*, and an oleaginous relative, *Chromochloris zofingiensis*, have improved our understanding of the genomic basis for remodeling of energy metabolism as a consequence of carbon source. Additional transcriptomic and proteomic analyses of Chromochloris under macro- and micro-nutrient stress conditions will provide information about how the organism maintains photosynthesis. Leveraging discoveries from green algal genomics, transcriptomics and proteomics, we are now building and applying new synthetic biology tools to modify metabolism in the genetically tractable, oleaginous Trebouxiophyte, *Auxenochlorella protothecoides*.

**Rationale:** Continued combustion of fossil fuels and production of greenhouse gases is incompatible with maintaining a stable climate regime. Various biological systems are being explored as platforms for the sustainable production of replacement fuel and chemical precursors that are commonly derived from petroleum. Crop plants provide readily scalable systems for synthetic biology, but raise concerns around diversion of arable land away from food production. Engineered prokaryotes and fungi have long been used for synthesis of alcohols, carboxylic acids, enzymes, therapeutic proteins and pharmaceuticals, but these organisms require a reduced carbon feedstock for growth. Facilities for large-scale heterotrophic fermentation are expensive and energy-intensive, and it can be difficult to reduce production costs to a sufficiently low level to justify making fuel or industrial biochemicals. Utilization of photosynthetic microalgae as host organisms for bioproduction could bridge some of the problems with crop plant and heterotrophic microorganisms. While it is not possible to achieve the high cell densities of heterotrophic fermenters, many algae have substantially higher photosynthetic productivity per unit area than plants. They can also be grown in marginal areas that are inappropriate for agriculture, and some can tolerate brackish or salt water and so do not increase pressure on potable water sources. While not as inexpensive as planting seeds, capital outlays to build algae ponds and bioreactors are substantially less than for industrial fermentors. Problems with large-scale outdoor algae production include reduced productivity due to self-shading, maintenance of monoculture, predation, and obtaining a positive energy balance when considering mixing, harvesting, de-watering and extraction of energy-rich compounds such as lipids.

**Progress:** Over the course of this project, systems analyses of the oleaginous freshwater Chlorophyte, *Chromochloris zofingiensis*, during trophic transitions between photoautotrophic growth and heterotrophic growth in the presence of glucose, have revealed the genomic basis for remodeling of cellular metabolism in the presence of different carbon sources[1–3]. Of particular interest to our group is how the organism responds to trace metal starvation, a situation faced by algae in nature and one that limits photosynthetic productivity. We have performed initial physiological characterization of photoautotrophic Chromochloris cells in order to define concentrations of copper, iron and zinc that are limiting, deficient and replete for growth, and we have used ICP-MS measurements to understand how starvation affects cellular quotas of macro- and micronutrients. Transcriptomic and proteomic analyses of how Chromochloris responds to trace metal starvation are in progress.

Chromochloris is a robust producer of triacylglycerides when supplied with glucose or another reduced carbon source. However, the Chromochloris cell wall has hindered efforts to

transform and engineer this species. Therefore, we have developed *Auxenchlorella protothecoides*, an oleaginous Trebouxiophyte that undergoes analogous trophic transitions, as a potential host strain for synthetic biology. We have developed methods for Auxenochlorella transformation, nuclear gene targeting and high-level transgene expression by means of homologous recombination, and we demonstrate the utility of this species as both a reference organism for basic research and as a potential biotechnology platform. To develop the system, we took advantage of public genome sequences for *A. protothecoides* Cp0710/UTEX 25 and UTEX 2341. These haploid assemblies were all highly fragmented, but we were able leverage these sequences to resolve polymorphic diploid alleles at multiple loci for *A. protothecoides* UTEX 250. Public transcriptome datasets were used to improve models for genes involved in central carbon metabolism, fatty acid and lipid biosynthesis, chlorophyll biosynthesis, light-harvesting and photosynthetic electron transfer.

We chose to demonstrate application of the system for reverse genetics and transgene expression by disrupting chlorophyll biosynthesis, since as the primary light-harvesting and photochemical pigment, chlorophyll is fundamental to bioenergy, and chlorophyll biosynthetic mutants have readily scorable, visual phenotypes. The *CHL27* gene encodes the di-iron component of Mg-protoporphyrin IX monomethylester (Mg-PIXMME) cyclase, a key enzyme that catalyzes formation of the chlorophyll isocyclic E ring. Sequential ablation of both *CHL27* alleles resulted in non-photosynthetic mutants that were completely chlorophyll-deficient and accumulated a red intermediate pigment with an absorption spectrum that was consistent with Mg-PIXMME. Chlorophyll biosynthesis and photosynthetic growth were restored to *chl27* double knockouts by knock-in of a synthetic, codon-optimized *CHL27* gene. Knock-ins could be targeted to a neutral locus, *DAO1*, encoding a non-essential D-amino acid oxidase, and expression of the *CHL27* transgene could be driven by any of three strong promoters that were expressed in photoautotrophic cells. Alternatively, knock-ins could be targeted to one of the mutated *chl27* alleles, and *in situ* expression of the transgene was driven by the native regulatory elements at *CHL27*.

We recently reported discovery of evolutionarily conserved polycistronic transcripts in several species of green algae[4]. Auxenochlorella strains transformed with synthetic bicistronic constructs co-expressed both a *SUC2* selectable marker gene, enabling growth on sucrose, and a *BKT1* reporter gene, encoding beta-carotene ketolase and catalyzing synthesis of red keto-carotenoids. *In vivo* comparison of SUC2 and BKT1 activity from polycistronic transcripts driven by two different promoters suggested an inverse relationship between the amount of translation from the upstream and downstream ORFs. This would be consistent with the hypothesis, suggested by *in vitro* results, that the abundance of proteins produced by polycistronic transcripts is determined by the relative strength of the Kozak sequences associated with each ORF. Collectively, these results demonstrate our ability to precisely and reversibly target Auxenochlorella nuclear genome sequences, and to control transgene expression, which will facilitate future metabolic engineering of strains for photosynthetic production of biofuels and value-added bioproducts.


1. Roth, M. S. *et al.* Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production. *Proc. Natl. Acad. Sci.* **114**, E4296–E4305 (2017).
2. Roth, M. S., Westcott, D. J., Iwai, M. & Niyogi, K. K. Hexokinase is necessary for glucose-mediated photosynthesis repression and lipid accumulation in a green alga. *Commun. Biol.* **2**, 1–10 (2019).
3. Roth, M. S. *et al.* Regulation of Oxygenic Photosynthesis during Trophic Transitions in the Green Alga Chromochloris zofingiensis. *Plant Cell* **31**, 579–601 (2019).
4. Gallaher, S. D. *et al.* Widespread polycistronic gene expression in green algae. *Proc. Natl. Acad. Sci.* **118**, (2021).



*This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), award number DE-SC0018301.*

**Metabolic Modeling and Engineering of Enhanced Anaerobic Microbial Ethylene Synthesis**

William R. Cannon[1]* (william.cannon@pnnl.gov), Ethan King[1], Adrienne B. Narrowe[2], Sarah J. Young[3], Kelly C. Wrighton[2] and **Justin A. North[3]**

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO; [3]Department of Microbiology, The Ohio State University, Columbus, OH

**Project Goals:** To develop robust and optimized anaerobic ethylene pathways in photosynthetic and lignocellulosic bacteria for high-yield conversion of renewable $CO_2$ and lignocellulose into bioethylene. This will be accomplished by:

1: First, bioinformatically mining and experimentally screening methylthio-alkane reductase homologs from cultivated and uncultivated organisms for functional enzymes that enhance ethylene yields.

2: Next, constructing and employing predictive systems-level models of ethylene production. Using a physics-based *R. rubrum* model, we predict enzymes that participate in competing or supporting pathways and are thus targets for knockdown or selection studies to increase ethylene yields. This will direct engineering strategies of top-performing genes (Aim 3) to maximize ethylene yield and minimize trade-off costs.

3: Finally, we metabolically engineer bacteria for enhanced, sustained ethylene production from CO2 and lignocellulose. We assemble the best-performing genes under control of optimized active transcription elements on a modular DNA fragment in a combinatorial manner with guidance from predictive models (Aim 2).

**Abstract Text:** Ethylene is the highest production industrial platform chemical on earth and constitutes a $300 billion global industry. Because of its versatility, ethylene is used to manufacture most plastics including polyethylene films, polyester fabrics, polystyrene packaging, and PVC piping. Plants and select fungi and bacteria naturally produce ethylene in a regulated manner as a signaling hormone. Known aerobic pathways employ one of two oxygen-dependent enzymes that specifically catalyze the oxidation of substrate to ethylene. However, both systems have faced scale-up challenges due to oxygen-ethylene combustion hazards and formation of cytotoxic products (hydrogen cyanide and guanidine). We uncovered a novel and previously uncharacterized methylthio-alkane reductase in *Rhodospirillum rubrum* that produces ethylene in the absence of oxygen (Fig. 1, rxn 4) [1]. The goal of this project is to optimize this anaerobic ethylene production pathway, as outlined above. This report describes predictions from a physics-based model for gene/enzyme engineering and the experimental outcomes.

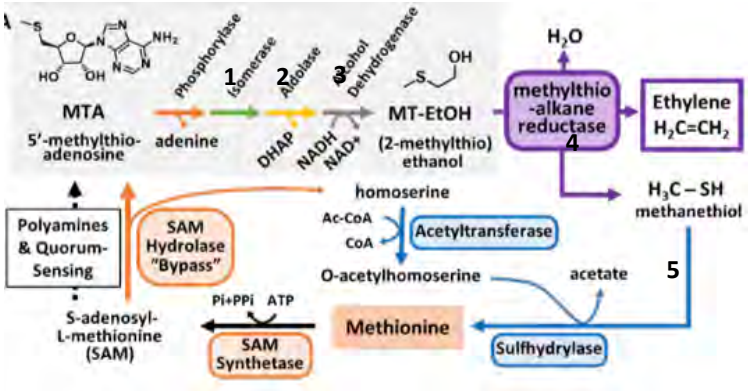



**Fig. 1 Anaerobic Ethylene Cycle for microbial synthesis of ethylene.** Reactions probed by thermo-kinetic modeling for their effect on pathway flux and ethylene yields are numbered. DHAP, dihydroxyacetone phosphate; CoA, coenzyme-A.

Biology and Physics: The physics-based model is based on recent developments in statistical

thermodynamics of non-equilibrium processes that rigorously describe time-dependent thermodynamic processes, in which classical thermodynamics and kinetics are united in the formulation of entropy production rates. Rate parameters needed for solving the differential equations describing metabolism are obtained by considering that natural selection is a thermodynamic process in which the fittest organisms seek to maximize their entropy production rates, which includes the process of replication, relative to their neighbors. Knowledge of these maximal rates allows for back-calculation of rate parameters. In addition, machine learning/mathematical optimization is used to predict regulation of pathways needed to keep metabolite concentrations in the feasible range and to optimize the activity of the metabolic pathways to be consistent with physiology [2]. Metabolic control analysis, a form of control theory specific for metabolism and systems of coupled reactions, is used to predict which enzymes and reactions influence the production of ethylene the most, either positively or negatively.

Metabolic Engineering: The proposed targets for metabolic engineering to enhance ethylene production were five reactions directly on the ethylene biosynthetic pathway (Fig. 1, rxn. 1-5) and two reactions that act to drain the ethylene pathway of precursors. The proposed on-pathway targets were the (1) isomerase, (2) aldolase, (3) alcohol dehydrogenase, (4) methyl-thioalkane reductase that directly produces ethylene, and (5) the sulfhydrylase that recycles products of the methyl-thioalkane reductase reaction. The proposed off-pathway targets are two enzymes that drain the aldolase substrate, 5-methylthio-ribulose-1-phosphate, from the ethylene biosynthetic pathway: (6) a RubisCO-like isomerase that converts 5-methylthio-ribulose-1-phosphate to 1-methylthio-ribulose-5-phosphate, and (7) a reductase that converts the latter product to deoxyxylulose-5-phosphate for isoprenoid synthesis.

The model suggested that down regulating the off-pathway RubisCO-like isomerase reaction (reaction 6, above) would change flux to ethylene by 2-10 fold, which has experimentally verified (9-fold increase ethylene yield). The model also predicted that the on-pathway isomerase (Fig. 1 rxn 1) would have the least impact on ethylene production, which has so far been the case in on-going experimental studies (max 2-fold increase in ethylene yields). The on-pathway aldolase and alcohol dehydrogenase activities (Fig. 1; rnx. 2, 3) are predicted to have intermediate impacts on ethylene production, and engineering of the aldolase reaction through overexpression or use of more catalytically active homologs has increased ethylene yields up to 4.8-fold, consistent with predictions. Modeling indicated the ethylene and sulfhydrylase reactions (Fig. 1; rxn. 4, 5) would have the largest impact on ethylene production. Indeed, overexpression of the native methylthio-alkane reductase increased ethylene yields directly from substrate, 2-methylthioethanol, by ~7-fold, and the sulfhydrylase reaction is being investigated experimentally now.

**References:**


[1] J. A. North *et al.*, "A nitrogenase-like enzyme system catalyzes methionine, ethylene, and methane biogenesis," *Science,* vol. 369, no. 6507, pp. 1094-1098, Aug 28 2020, doi: 10.1126/science.abb6310.

[2] S. Britton, M. Alber, and W. R. Cannon, "Enzyme activities predicted by metabolite concentrations and solvent capacity in the cell," *J R Soc Interface,* vol. 17, no. 171, p. 20200656, Oct 2020, doi: 10.1098/rsif.2020.0656.



**Funding:** *This project is supported by the U.S. Department of Energy's Office of Biological and Environmental Research under contracts* DE-SC0022091 and DE-SC0019338 (Ohio State University) and 78266 (PNNL).

# High Throughput Screening of Enzymes that Bolster Anaerobic Ethylene Synthesis

Adrienne Narrowe[2], Sarah J. Young[1*] (young.1517@osu.edu), Hannah Lucero[2], Maider E. Medrano[2], F. Robert Tabita[1] (Former PI, deceased), William R. Cannon[3], Kelly C. Wrighton[2], and **Justin A. North[1]**

[1]The Ohio State University, Columbus, OH; [2]Colorado State University, Fort Collins, CO; [3]Pacific Northwest National Laboratory, Richland, WA.

**Project Goals:** The overall long-term objective is to develop an industrially compatible microbial process to synthesize ethylene in high yields from $CO_2$ and lignocellulose. To optimize the activity of the recently discovered dihydroxyacetone phosphate (DHAP) ethylene pathway for increased ethylene yields, the following specific goals are:

1. Discover effective and active DHAP ethylene pathway enzymes encoded in cultured and uncultured organisms from anoxic environments. (Wrighton and North)
2. Construct a modular set of optimized genes on a DNA fragment containing specific regulatory elements that will allow high level gene expression in model organisms. (North and Cannon)

**Abstract Text:** Our previous work identified a novel anaerobic microbial pathway (DHAP-Ethylene Pathway) [1] that converted 5'-methylthioadenosine (MTA) to stoichiometric amounts of ethylene. MTA is a metabolic byproduct of methionine utilization in a multitude of cellular processes. The initial steps of the DHAP-ethylene pathway sequentially convert MTA to dihydroxyacetone phosphate (DHAP) and ethylene precursor (2-methylthio)ethanol (Fig. 1; gray), followed by reduction to the methionine precursor, methanethiol, and ethylene via a novel nitrogenase-like methylthioalkane reductase (Fig. 1; red box) [2]. Kinetic characterization of the initial DHAP-ethylene pathway enzymes from the photosynthetic bacterium, *Rhodospirillum rubrum*, revealed that while the phosphorylase had relatively high specific activity (MtnP = $3.2 \times 10^5$ $M^{-1}$ $s^{-1}$; Fig. 1, blue arrow), the subsequent MTR-1P isomerase and MTRu-1P aldolase had relatively lower specific activity (MtnA = $0.75 \times 10^3$ $M^{-1}$ $s^{-1}$; Ald2 = $1.7 \times 10^3$ $M^{-1}$ $s^{-1}$; Fig. 1, green and yellow arrows, respectively) [3]. This suggested that introduction of more catalytically efficient isomerase and aldolase homologs could increase ethylene yields by increase the DHAP-ethylene pathway flux.

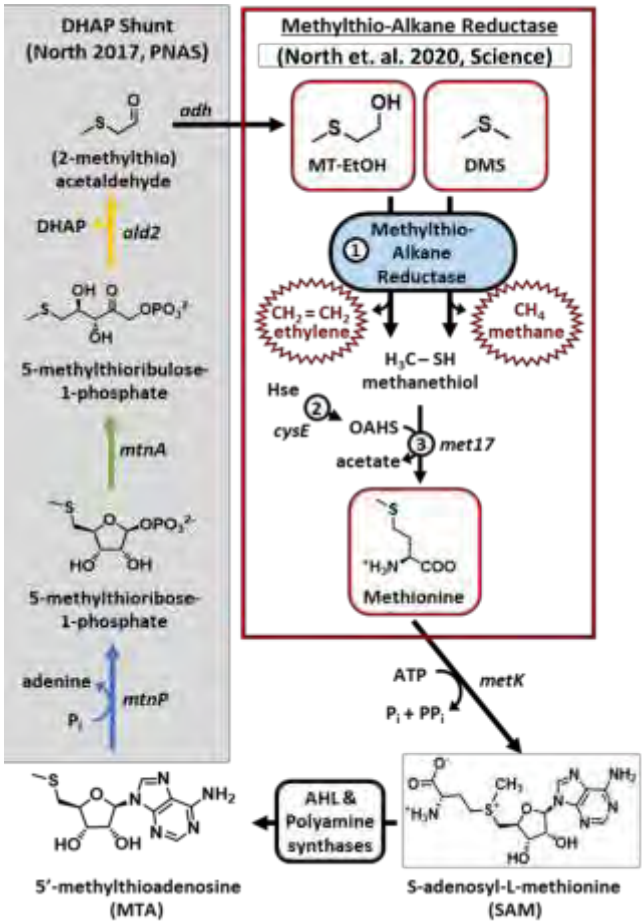



**Fig. 1. Anaerobic ethylene cycle (DHAP-ethylene pathway),** which is composed of the DHAP shunt (gray box) and methylthio-alkane reductase methionine synthesis pathway (red box). (1) Methylthio-alkane reductase (*marHDK*), (2) homoserine acetyltransferase (*cysE)*, (3) acetylhomoserine Sulfhydrylase (*met17*). MT-EtOH, (2-methylthio)ethanol; DMS, dimethylsulfide.

Indeed, introduction of more catalytically active aldolases from *E. coli* increased ethylene yields by 2-fold in initial trials [1,3].

***High-throughput screening of isomerases and aldolases from genome sequences of isolated and environmental bacteria*** – To identify and isolate enzymes with robust catalytic properties that enhance ethylene production, we adopted a high-throughput cell lysate activity screening approach for enzymes heterologously produced in *E. coli*. Mining of JGI IMG/M genome and metagenome sequence databases for candidate orthologs to the MTR-1P isomerase (*mtnA*) and the MTRu-1P aldolase (*ald2*) genes yielded 1,371,813 and 96,049 candidate genes for *mtnA* and *ald2*, respectively. We selected two hundred syntenous *mtnA / ald2* gene pairs that represented the phylogenetic breath of sequences and considered proximity to active orthologs. Gene product activity in *E. coli* was quantified by colorimetric detection of phenylhydrazine adducts with (2-methylthio)acetaldehyde produced from MTA via coupled enzyme assay in 96-well plate format. ***The isomerase library screen***: recovered 42 soluble and active isomerases out of 250 sequences with *in extracto* specific activity ranging from 3.3-8800 nmol/min/mg total protein. Introduction of a subset of these isomerases into the *R. rubrum* isomerase deletion strain via a plasmid resulted in restored conversion of MTA into ethylene. Select sequences increased ethylene yields by 30% and 60% compared to when the native *R. rubrum* isomerase was expressed from the plasmid or chromosome, respectively. ***The aldolase library screen*****:** recovered ~120 soluble and active aldolases out of 250 sequences with *in extracto* specific activity ranging from 2.0-7996 nmol/min/mg total protein. Introduction of a subset of these aldolases into the *R. rubrum* aldolase deletion strain via a plasmid similarly resulted restored conversion of MTA into ethylene. Select sequences increased ethylene by 35% and 5-fold compared to the *R. rubrum* aldolase expressed from the plasmid or chromosome, respectively.

***Conclusions and future directions:*** Orthologs of DHAP-ethylene pathway isomerases and aldolases with potentially increased catalytic properties were recoverable from bacteria isolates and metagenomes spanning firmicutes, alpha, beta, gamma, and deltaproteobacteria. An increase in DHAP-ethylene pathway isomerase activity, either through overexpression and/or introduction of more active orthologs in *R. ruburm* increased ethylene yields by only 60%, indicating a potential limitation in pathway flux enhancement that can be achieved through modifying the isomerase alone. In contrast, increasing aldolase activity clearly increased pathway flux, resulting in a 5-fold increase in ethylene production. Future studies will quantify the combined effect of efficient isomerases and aldolases.

**References:**


1. North JA, Miller AR, Wildenthal JA, Young SJ, Tabita FR. Microbial pathway for anaerobic 5'-methylthioadenosine metabolism coupled to ethylene formation. *Proc. Natl. Acad. Sci. U S A*. **2017**. 114: E10455-E10464
2. North JA, Narrowe AB, Xiong W, Byerly KM, Zhao G, Young SJ, Murali S, Wildenthal JA, Cannon WR, Wrighton KC, Hettich RL, Tabita, FR. A nitrogenase-like enzyme system catalyzes methionine, ethylene, and methane biogenesis. *Science.* **2020.** 369(6507):1094-1098
3. North, JA, Wildenthal, JA, Erb, TJ, et al. A bifunctional salvage pathway for two distinct S-adenosylmethionine by-products that is widespread in bacteria, including pathogenic *Es coli*. *Mol Microbiol.* **2020**. 113: 923– 937



**Funding Statement:** The U.S. DOE, Office of Science, Office of Biological and Environmental Research, Genomic Science Program supported the high throughput screening and ethylene production in photosynthetic bacteria work (DE-SC0019338 to F.R.T, K.C.W., W.R.C., and J.A.N.). DNA synthesis for functional diversity characterization was supported by the DOE JGI DNA Synthesis Sciences Program (505283 to K.C.W, A.B.N., J.A.N. and F.R.T.)

# The Twin Ecosystems Project: A New Capability for Field and Laboratory Ecosystems Coupled by Sensor Networks and Autonomous Controls

**Authors:** Javier A. Ceja-Navarro (jcnavarro@lbl.gov),[1,*] Sheryl Bell[2]*, James Sethian,[1] Nancy Hess,[2,3] Kirsten Hofmockel,[2] Susannah G. Tringe,[1,4] and **Trent Northen**[1,4]

**Institutions:** [1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Pacific Northwest National Laboratory, Richland, WA; [3]Environmental Molecular Sciences Laboratory, Richland, WA; and [4]DOE Joint Genome Institute, Berkeley, CA

**Website URL:** http://ecofab-twins.lbl.gov

**Project Goals: This project will develop 'twin' lab and field ecosystems to create a new capability that scales-down field observations inside fabricated ecosystems to uncover plant-microbial system responses to drought. The 'twins' will be integrated using sensor networks and interrogated through the development and application of novel in situ sensors, imaging, Omics analysis, and autonomous controls.**

Discovering the fundamental biology that regulates ecosystem responses to changing climate is vital for predicting and managing ecosystem outcomes. To accurately identify biological drivers of ecosystem responses, technical approaches are needed to scale-down field observations to determine causal mechanisms. The goal of the TWIN ecosystem project ('TWINS') is to pilot laboratory and field 'twin' ecosystems that use sensors and autonomous controls to test the hypothesis that *compositional changes in root exudates during drought stress select for beneficial rhizosphere microbes.* The field 'twin' will define climate conditions and hyperspectral signatures of drought stress enabling the lab 'twin' to characterize the composition, localization, and dynamics of microbes and exudates. As such, the lab 'twin' will also provide powerful environmental controls and measurements which are not possible in the field. More specifically, TWINS is investigating molecular interactions in the rhizosphere to gain insights into the types of rhizosphere microbial communities that tall wheatgrass (*Thinopyrum ponticum*) exudates select in response to drought. Tall wheatgrass is a widely distributed species adapted to dry northern latitudes that is known to develop soil "resource islands" or "hot spots" that may impose heterogeneous spatial distribution of important plant exudates impacting the soil microbiome. These effects on the soil microbiome may be more prominent in response to drought when plants may differentially allocate photosynthates to roots.

To investigate microbiome selection and the formation of hot spots of microbial activity by wheatgrass we will integrate existing soil and environmental sensor technology at the laboratory (EcoFABs and EcoPOD) and the field irrigation trial of the PNNL soil SFA. Integrated technologies include multispectral imaging for the measurement of above-ground plant response, minirhizotrons for root imaging, and novel microbe-based biosensors for surveying below-ground microbial activity.

We have successfully excavated and relocated soil monoliths containing whole tall wheatgrass plants from PNNL's field site at Washington State University to the Berkeley Lab for incubation in the EcoPOD. Minirhizotrons were installed to enable imaging of the root system within the EcoPOD and the plants incubated under controlled conditions of photosynthetically available radiation (PAR), humidity, and temperature (Fig. 1). Hyperspectral cameras are being added into the EcoPOD to analyze plant responses to environmental manipulations, such as drought and changes in temperature. Next, and as part of the optimization of environmental sensors (PAR, T, atmospheric and soil humidity) we will simulate normal and drought conditions (50% vs 20% soil WHC) to analyze the differential recruitment and microbial network formation in the rhizosphere of the tall wheatgrass plants.

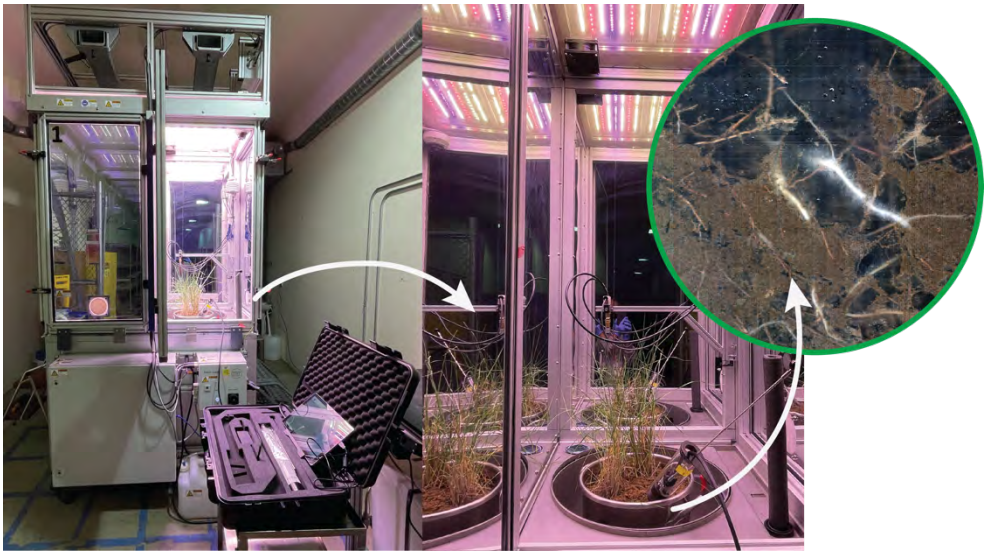


Fig. 1. Tall wheatgrass monoliths were excavated and relocated to the Berkeley Lab for incubation in the EcoPOD. The image shows the EcoPOD housing the plants with connected minirhizotrons. The upper inset image shows an image of roots photographed with the minirhizotron camera.

**Funding Statement:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under contract number DE-AC02-05CH11231. Work at PNNL is under FWP 78814. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

# Rhizosphere Biogeography of *Brachypodium distachyon* and Microbial Plant-Growth Promoting Traits

Romy Chakraborty[1], Mon O. Yee[1], Kristine G. Cabugao[1*], Spencer Diamond[2] , Shwetha Acharya[1], Peter F. Andeer[1], Nameera F. Baig[3], Jillian F. Banfield[2], Adam M. Deutschbauer[1], and **Trent R. Northen[1] (TRNorthen@LBL.gov)**

[1]Lawrence Berkeley National Laboratory, Berkeley; [2] University of California, Berkeley; [3]University of Minnesota Twin Cities, Minneapolis

http://mCAFEs.lbl.gov

**Project Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design.** To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

**Abstract:**

The rhizosphere is a critical interface between plants and microbial communities that mediates ecosystem processes such as nutrient cycling and carbon stabilization. Plants exude a variety of nutrient-rich compounds (exudates) that maintain a selection of microbes from the surrounding soil environment. The rhizosphere microbial community, in turn, can directly influence plant growth and development through plant-growth promoting and nutrient scavenging functions. We assessed the distribution of microbial communities in the rhizosphere (biogeography) of microbes colonizing the roots of *Brachypodium distachyon* in fabricated ecosystem devices (EcoFABs) as a result of spatially-distinct root exudation. In addition, we also examined key rhizobial isolates for their plant-growth promoting traits.

Distinct parts of the root can vary in exudation patterns. Specifically, fine roots, occurring at the tips of the root system, have higher exudation rates than coarser roots that occur near the base. However, the influence of these spatial patterns on the biogeography of rhizosphere microbial communities is not well understood. We analyzed the microbial community of *Brachypodium distachyon* at two distinct areas: roots collected at the tips of the primary root and roots collected at the base. Furthermore, we grew plants in unamended natural soil and compared results among plants grown in standardized fabricated ecosystems known as EcoFABs, as well as

in traditional pots and test tubes. Our results showed similar microbial community distributions across the different growth chambers and a higher degree of reproducibility within EcoFAB grown plants. We were also able to enrich for less characterized lineages such as *Verrucomicrobia* and *Acidobacteria* in addition to *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. Genome-resolved metagenomics indicate higher abundances of biofilm formation, flagellar assembly, and simple sugar transporter genes in microbial communities from the base and root tips in comparison to bulk soil. Furthermore, bulk soil and plant-base microbiomes were highly abundant in genes related to prokaryotic defense and transcription.

Next, we used rhizosphere samples from each growth chamber to enrich for rhizobacteria on single and mixed carbon media to examine the influence of carbon source on microbial community composition. We used media that reflected known *B. distachyon* exudates (i.e., glutamine, succinate, and asparagine) and sequentially transferred cultures every three days into fresh media to form reduced complexity communities amenable to targeted rhizosphere assembly studies. Our results indicate that different carbon sources resulted in the enrichment of differing abundances of genera. For example, we found that cultures from glutamine and succinate contained higher abundances of *Burkholderia* and *Enterobacter* but that mixed carbon sources such as R2A yielded smaller abundances of each but with a higher diversity of genera present. Furthermore, we measured key plant-growth promoting traits in axenic isolates obtained from these enrichments. Initial assessment of indole-3-acetic acid (IAA) phytohormone production indicated that all of our recovered isolates are capable of producing auxin to some degree (0.3 µg IAA $ml^{-1}$ to 10.54 µg IAA $ml^{-1}$). Further assays will include measuring other plant-growth promoting traits related to nutrient acquisition such as phosphorus solubilization and siderophore production.

Our research reveals patterns in the distribution of bacteria and gene abundances in different areas of the root. Furthermore, we showed the comparable performance of EcoFAB devices with regards to traditional plant growth chambers and also their higher reproducibility in plant-microbial studies. Lastly, reduced complexity communities formed using sequential transfers in media containing single and mixed known *B. distachyon* exudates indicate a common ability to produce indole-3-acetic acid, a phytohormone known to induce growth in plant roots.

*Funding statement.*
*This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) an SFA led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Examining Molecular Mechanisms Selecting for Rhizosphere Bacteria**

Dawn Chiniquy,[1] Hsiao-Han Lin[1], Spencer Diamond,[2] Marta Torres Béjar,[1] Catharine A. Adams,[2] Trenton K. Owen[1], Hans Carlson,[1] N. Louise Glass[2], Jenny C. Mortimer,[1,3] Jillian Banfield,[2] **Adam M. Deutschbauer[1],** and **Trent R. Northen[1*]**(TRNorthen@LBL.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley; [2] University of California, Berkeley; [3]University of Adelaide, Australia; [7]Lawrence Livermore National Laboratory, Livermore

http://mCAFEs.lbl.gov

**Project Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design.** To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

**Abstract**

The plant rhizosphere microbiome is both dynamic and complex, with many host and environmental factors governing its assembly and function. Here, we explored the assembly of the model plant *Brachypodium distachyon* root microbiome over time using a synthetic community (SynCom) of 17 microbes. Using 16S ribosomal sequencing, we demonstrate that both time of inoculation (0, 3, 8 days after germination) and duration of the experiment (14, 24 and 29 days) affected microbiome composition in fabricated ecosystems (EcoFABs). Next, we tested the role of root exudates and root glycans in this process. This was initially done by mapping the colonization pattern of a fluorescently tagged member of the SynCom, *Rhizobium* OAE497, in the *B. distachyon* plant mutant PMT that over expresses *p*-Coumaroyl-CoA:monolignol transferase (Petrik et al. 2014) and has an altered root cell wall composition. Further, we discovered a potential role for several plant cell wall-derived glycans in enriching for novel microbes that have proved challenging to culture, including members of the *Acidobacteria,* to develop reduced complexity microbial communities. Finally, we identified specific microbial genes and pathways responsible for colonization under multiple growth

conditions through use of an RB-Tnseq mutant library developed in another SynCom member, *Burkholderia OAS925*. Collectively, these orthogonal approaches deepen our knowledge of the genetic and metabolic basis of rhizosphere assembly.

## References

Petrik DL, Karlen SD, Cass CL, et al. p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in Brachypodium distachyon. *Plant J*. 2014;77(5):713-726. doi:10.1111/tpj.12420

*Funding statement.*
*This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) an SFA led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Spatio-temporal genome-scale metabolic modeling of the rhizosphere microbiome**

Ilija Dukovski[1*] (dukovski@bu.edu), Hui Shi[1], Melisa L. Osborne[1], Jing Zhang[1], Michael Silverstein[1], Peter Kim[2], Anup Singh[3], Daniel Segrè[1], Adam M. Deutschbauer[4], and **Trent R. Northen[4]**

[1]Boston University, Boston; [2]Sandia National Laboratory, Livermore; [3]Livermore National Laboratory, Livermore; [4]Lawrence Berkeley National Laboratory, Berkeley.

http://mCAFEs.lbl.gov ; https://runcomets.org

**Project Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design.** To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

One of the key methods we employ to understand the complexity of the rhizosphere microbial community is computational modeling. With the increasing availability of sequenced genomes and methods for building genome-scale metabolic models of organisms, such as the ones provided in the KBase platform[1], we are in position to harness the power of genome-scale stoichiometric reconstructions in order to reach predictive modeling capabilities[2]. Our software platform COMETS (Computation Of Microbial Ecosystems in Time and Space)[3] was developed with the purpose of predictive modeling of the complex dynamics of interactions in microbial ecosystems where the spatial structure may play a crucial role. COMETS combines genome-scale metabolic modeling using dynamic flux balance analysis (dFBA) with spatio-temporal simulations of the growth and propagation of the biomass and extracellular metabolites in spatially structured environments. With COMETS we are able to simulate a variety of spatially structured microbial communities, from simple assays in a Petri dish to complex natural environments such as a plant root. We have recently released a significantly improved version of COMETS[3] (2.0, see http://runcomets.org), which includes enhanced biophysical models for biomass propagation, evolutionary dynamics, extracellular enzyme activity and night/day light cycles. COMETS 2.0 also includes user-friendly Python and Matlab interfaces, as well as the documentation and the source code in a publicly accessible github repository (https://github.com/segrelab/comets).

Our current modeling effort is focused on two main goals. The first goal is to make the best use of our computational platform and the experimental capabilities of the m-CAFEs team to construct, curate and test metabolic models starting from sequenced genomes and phenotypic data. Notably, the dFBA core algorithm of COMETS allows us to generate predictions of whole growth curves, which require knowledge of kinetic uptake parameters. We have been focusing on the important rhizosphere bacterium

*Pseudomonas simiae* as a test case. After building a draft model using automated pipelines, we compared COMETS-predicted growth curves under different nutrient limitations to corresponding experimental measurements, and generated new hypotheses on the resources that limit growth under different conditions. Further comparison with experimental data, including growth curves and metabolite uptake/secretion rates under different media, as well as TnSeq data, will allow us to implement rounds of gap filling and refinements of the genome-scale reconstruction. This approach is in parallel being extended to systematically study multiple rhizosphere microbes and their interactions.

Second, we are enhancing COMETS in order to improve its capacity to capture biological and biophysical processes that may be crucial for the spatio-tempotral dynamics of communities around the plant roots. In order to start testing our capacity to accurately predict the dynamics of communities in gradients of nutrients and signals, we have been developing in parallel a microfluidic experimental device, the µEcoFab (MicroEcoFab), and a corresponding *in silico* COMETS version. In order to make COMETS simulations more realistic, we have introduced a new module that can simulate bacterial chemotaxis. This module is based on the Keller-Segel[4] model of chemotaxis, which adds to the biomass density equation a convective term proportional to the gradient of the chemoattractant (or chemorepellent). Preliminary *in silico* experiments demonstrate that our module can predict the dynamics of colony shapes as they propagate towards the chemotactic signal. Ongoing experiments with the µEcoFab are being compared to COMETS simulations, based on a newly developed layout that captures the properties of this device. We expect that this concerted effort, combining experimental measurements and computational simulations, will allow us to better understand how molecular complexity and spatial structure of the environment shape rhizosphere microbiome properties, further facilitating our ongoing efforts towards designing and controlling communities for specific goals.

**References**


1. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol. 2018;36: 566–569.

2. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, et al. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. Cell Rep. 2014;7: 1104–1115.

3. Dukovski I, Bajić D, Chacón JM, Quintin M, Vila JCC, Sulheim S, et al. A metabolic modeling platform for the computation of microbial ecosystems in time and space (COMETS). Nat Protoc. 2021;16: 5030–5082.

4. Keller EF, Segel LA. Initiation of slime mold aggregation viewed as an instability. Journal of Theoretical Biology. 1970. pp. 399–415. doi:10.1016/0022-5193(70)90092-5



*Funding statement.*
*This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) an SFA led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Evaluating Plant-Microbe Interactions, Persistence and Movement of Microbial Communities Across Scales**

Albina Khasanova[1*](akhasanova@lbl.gov), Daniel Naylor[1], Jonathan Diab[1], Shweta Priya[1], Jenny C. Mortimer[1,2], Aindrila Mukhopadhyay[1], John Vogel[1], Kateryna Zhalnina[1], Adam M. Deutschbauer[1], and **Trent R. Northen**[1*](TRNorthen@LBL.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]University of Adelaide, Australia

http://mCAFEs.lbl.gov

**Project Goals: Understanding the interactions, localization and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design.** To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

Plant root exudates recruit microbes and regulate interactions which vary over the plant life cycle. For example, root secreted metabolites can regulate plant-soil-microbiome interactions by affecting the colonization of symbiotic microbes in the root rhizosphere. These relationships in turn play an important role in plant health by altering nutrient availability and both biotic and abiotic stress tolerances over the course of plant development. Here we examine how these interactions influence plant productivity and morphology across scales using fabricated ecosystems (centimeter-scale EcoFABs and meter-scale EcoPODs). EcoFABs and EcoPODs facilitate the examination of complex interactions under highly controlled and replicated conditions to explore plant-microbial-soil interactions. Both systems are specifically designed to be compatible with biosafety workflows. EcoFABs allow for the extremely controlled manipulation of abiotic and biotic factors in small scale, high-throughput experiments with plant seedlings, while the EcoPOD system allows for precise control and monitoring of conditions across several growth parameters, including light intensity, temperature, humidity, water availability, and other important climatic parameters both above and below ground over the main developmental stages of the plant. The EcoPOD's large size and soil depth allows for the experiments where plants experience conditions similar to the field, yet provides for a high degree of control in the manipulation of engineered microbial communities. Together, these two systems allow for complete system development to study the persistence and fate of engineered genes, microbes and SynComs over the entirety of plant life cycles as well as the role of these interactions in tolerances to biotic and abiotic stressors. Currently we investigate plant-microbe interactions across scales (**i**) to uncover the role of aromatic acid in rhizosphere microbiome

assembly and stability; (**ii**) to engineer and test microbial enrichments under limited N conditions; (**iii**) to study the effect of drought stress on plant-microbial relationships.

Aromatic acids are one of the important classes of metabolites involved in rhizosphere-microbiome interactions and are widely exuded by plants. However, the role and mechanisms through which these molecules influence rhizosphere microbiome establishment are unknown. To dissect the role of aromatic acids as metabolic handoffs that strengthen plants-microbe interactions we identified *Brachypodium distachyon* accession lines with altered aromatic acid exudate profiles using EcoFABs. We pair these lines with SynComs utilizing these aromatic acids and evaluate dynamics of SynComs in response to different levels of exuded aromatic acids. In parallel, we utilized a RB-TnSeq mutant library of *Bulkholderia* and identified the microbial genes protocatechuate 3,4-dioxygenase alpha chain and 3-oxoadipate CoA-transferase subunit B, which are related to the microbial degradation of aromatic acids, as required for optimal growth on shikimic acid. We are currently constructing targeted mutants of these genes to demonstrate how the structure of SynCom changes when microbes are unable to utilize aromatic acids. Lastly, we used the EcoPOD to conduct an experiment that paired SynComs in tandem with *Brachypodium distachyon* accession lines that expressed altered production of aromatic acids at a field-relevant scale.

Limited N availability is a significant factor affecting plant growth. In this project, we aim to use a host-mediated microbiome engineering approach to develop a beneficial microbiome for low-N conditions. In this approach, *Brachypodium distachyon* plants are inoculated with a native soil microbiome, exposed to low-N inside EcoFABs over several rounds of selection, and the best-performing plants will have their microbiomes harvested and perpetuated onto the next round of selection, in this way 'evolving' a beneficial microbiome with a tight host association. We have successfully established an assay for inducing negative growth phenotypes in *Brachypodium* in response to decreasing N levels and performed pilot inoculation with soil enrichments of microbes. Future experiments will involve testing alternate soil sources to find one with more beneficial microbes. Once the effect of the enrichment is established in the EcoFAB, it will be tested at a more field-relevant scale using the EcoPOD.

Plant response to drought stress may be in part affected by root microbial relationships. We have used the EcoFAB system to develop a baseline dataset of *Brachypodium distachyon* response to drought in the presence of PEG6000, and identified an osmotic stress level that reduced growth without inducing mortality. We then used a soil bacterium (*Pseudomonas putida* KT2440) expressing mCherry as a proof of concept to test the effects of drought on root colonization. The EcoPOD system carefully controls water availability and enables mimicking field drought in the lab. We are currently testing the *Brachypodium-Pseudomonas*-drought system in the EcoPOD-mini, and comparing the findings with our small scale EcoFAB results.

*Funding statement.*
*This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) an SFA led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

# A rapid *Brachypodium distachyon* transformation method using leaf whorl explants

Hsiao-Han Lin[1*](hsiaohanlin@lbl.gov), Tallyta Silva[1], Nick Downs[1], Jenny C. Mortimer[1,2*](jcmortimer@lbl.gov), Adam M. Deutschbauer[1], and **Trent R. Northen**[1*](TRNorthen@LBL.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley; [2] University of Adelaide, Australia

http://mCAFEs.lbl.gov

**Project Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design.** To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

## Abstract

To fully understand grass rhizosphere communities, it is important to consider the role of the plant root in shaping the rhizosphere microbiome. However, this is poorly understood, in part due to the difficulties in generating plant mutants, since plant transformation, particularly of grasses, is time-consuming and laborious. The m-CAFEs project uses *Brachypodium distachyon*, as it is an excellent model for bioenergy feedstocks. The current transformation method for *B. distachyon* ecotype Bd21-3 uses immature embryos as the explant material, and has a cycle of 22-31 weeks from non-transgenic seed to the first generation of transgenic seed. Immature embryos are both technically challenging to isolate and only available from plants during a short developmental window. Recently, a novel method using leaf whorls as the explant material has been developed in sorghum (Silva et al., 2020). Here, we demonstrate that the leaf whorl-based transformation can be adapted to *B. distachyon*, thereby shortening the cycle to 14-20 weeks. Advantages include that since the explant material is derived from 3-week-old seedlings, rather than mature plants, total experiment length is reduced. Second, callus can be generated from the plants continuously until the plant starts to set seed (~8-12 weeks depending on ecotype). Third, isolating leaf whorls is much less technically challenging and laborious compared to immature

embryos. Finally, we have been successful in generating calli from late-flowering ecotypes, such as Bd1-1, which can be difficult to isolate immature embryos from under lab conditions (Schwartz et al., 2010). This method can be integrated with other advances in plant transformation, such as the expression of morphogenic regulations (e.g. BBM/WUS or GRF-GIF4) to further accelerate plant transformation.

**References**

**Schwartz, C.J., Doyle, M.R., Manzaneda, A.J., Rey, P.J., Mitchell-Olds, T., and Amasino, R.M.** (2010). Natural variation of flowering time and vernalization responsiveness in *Brachypodium distachyon*. BioEnergy Research **3**: 38–46.

**Silva, T.N., Kelly, M.E., and Vermerris, W.** (2020). Use of *Sorghum bicolor* leaf whorl explants to expedite regeneration and increase transformation throughput. Plant Cell, Tissue and Organ Culture (PCTOC) **141**: 243–255.

*Funding statement.*

*This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) an SFA led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**m-CAFEs Phage Engineering for Targeted Editing of Microbial Communities**

Avery Roberts[1]*(larober4@ncsu.edu), Matthew A. Nethery[1], Ben Adler[2], Brady Cress[2], Ben Rubin[2], Spencer Diamond[2], Jill Banfield[2], Jennifer A. Doudna[2], N. Louise Glass[3], Rodolphe Barrangou[1], Adam M. Deutschbauer[3], and **Trent R. Northen[3]**

[1]North Carolina State University, Raleigh; [2]University of California, Berkeley; [3]Lawrence Berkeley National Laboratory, Berkeley.
https://mcafes.lbl.gov/

**Project Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design.** To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

**Abstract:** Investigation of microbial gene function is essential to elucidation of the ecological roles and complex genetic interactions that take place in microbial communities. While microbiome investigations have increased in prevalence, the lack of viable *in situ* editing strategies impedes experimental design and progress, hindering genetic manipulation of microbial communities and genetic discovery in community contexts. Here, we demonstrate the utility of phage-delivered CRISPR-Cas payloads to perform targeted genetic manipulation within a community context deploying a fabricated ecosystem (EcoFAB) as an analog for the soil microbiome. We provide a roadmap for engineering phages T7 and λ for community editing using Cas9-mediated recombination in non-essential genes. We further engineer λ to deliver

antibiotic resistance and fluorescent genes to an *Escherichia coli* host. Expanding on this platform, we engineer λ to express an APOBEC-1-based cytidine base editor (CBE), which we leverage to perform C to T point mutations guided by a nuclease-deficient Cas9 (dCas9). We strategically introduced these base substitutions to create premature stop codons in-frame, inactivating both chromosomal (*lacZ*) and plasmid-encoded genes (mCherry and ampicillin resistance) without perturbing the surrounding genomic regions and maintaining host viability. Further, using a multi-genera synthetic soil community, we employed phage-assisted base editing to induce host-specific phenotypic alterations in a community context both *in vitro* and within the EcoFAB, observing editing efficiencies from 10% to 28% across the entire bacterial population. Future work aims to harness the enzymatic diversity of CRISPR-Cas systems to expedite and innovate on phage engineering workflows. The concurrent use of Cas nucleases, a synthetic microbial community, soil matrix, and EcoFAB device provides a controlled and reproducible model to more closely approximate *in situ* editing of the soil microbiome.

*This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) a Project led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231.*

**Targeted DNA Editing Within Microbial Communities**

Benjamin E. Rubin[1](brubin@berkeley.edu), Spencer Diamond[1], Brady Cress[1], Trenton K. Owens[3], Christine He[1], Alex Crits-Christoph[1], Zeyi Zhou[1], Michael Xu[1], Kimberly Tang[1], Dylan C. Smock[1], N. Louise Glass[3], Rodolphe Barrangou[2], **Jillian Banfield[1], Jennifer A. Doudna[1],** Adam M. Deutschbauer[3], and Trent R. Northen[3]

[1]University of California, Berkeley; [2]North Carolina State University, Raleigh; [3]Lawrence Berkeley National Laboratory, Berkeley.

http://mCAFEs.lbl.gov

**Project Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design.** To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

Our functional understanding of microbial DNA is predominantly founded on the principles of isolate genetics, where the effects of genetic manipulations on cultivable organisms are observed in isolation. Unfortunately, this provides limited insight into the workings of genes in the complex and societally relevant microbial communities that exist in nature. In order to move beyond the paradigm of manipulating microbes in confinement, we have created a generalizable toolset for targeted genome editing of individual organisms within complex microbial communities. First, we have developed environmental transformation sequencing (ET-Seq) to determine *in situ* which microbes within a community can be edited by untargeted transposases, and with what efficiency. Second, we have repurposed RNA-guided CRISPR-Cas transposases to paste customized DNA into unique target sites within the genomes of specific microbes in a community. Third, we have applied these technologies to track the fitness effects of genetic mutants in communities, as well as for enrichment and isolation of edited organisms. The ability to make organism- and locus-specific changes within microbiomes will lead to improved understanding of microbial communities and enable us to effect meaningful changes within them.

*Funding statement.*
*This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) an SFA led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

# The Filamentous Fungus *Trichoderma atroviride* as a Model System for Understanding Fungal Genetics, the Plant-Fungal Symbiosis, and Interactions with Diverse Bacteria

José Manuel Villalobos Escobedo[2] (jose.villalobos@berkeley.edu), Catharine A. Adams[2], Ran Shi[2], Hsiao-Han Lin[2], Trenton K. Owens[1], Jens Heller[1], Dawn M. Chiniquy[1], Jenny C. Mortimer[1,6], N. Louise Glass[2] **(lglass@berkeley.edu)**, Adam M. Deutschbauer[1], and **Trent R. Northen[1*]**(TRNorthen@LBL.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley; [2] University of California, Berkeley; [3]Sandia National Laboratory, Livermore; [4]North Carolina State University, Raleigh; [5]Boston University, Boston; [6]University of Adelaide, Australia; [7]Lawrence Livermore National Laboratory, Livermore

http://mCAFEs.lbl.gov

**Project Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design.** To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

The plant rhizosphere is ecologically important and houses diverse microbes including Archaea, bacteria, and fungi. Filamentous fungi in the genus *Trichoderma* are ubiquitous in soil, and have well characterized mycoparasitic, biocontrol, and plant growth promoting effects (Harman 2007). In this work we carry out the mutagenesis of all the genes of the *Trichoderma atroviride* genome to characterize their function under different growth conditions and in interactions with bacteria and the rhizosphere. To do this, we used t-DNA and *Agrobacterium tumefaciens* to insert barcodes throughout the fungal genome (Coradetti et al., 2018).

The preliminary t-DNA-seq in *T. atroviride* showed that more than 75% of the reads obtained from the genome have inserted barcodes, and we did not observe bias due to %GC content or an enrichment of insertions in any contigs. The barcode analyses showed that 7,115 of the 11,816 predicted genes in the *T. atroviride* genome have at least one insertion. We then performed a Bar-seq experiment under nutrient deficiency conditions to assess fitness of our insertional library of mutants. We observed that under glucose deficiency, strains with mutations in 67 genes showed lower fitness, while mutants in 36 genes have a positive fitness under these conditions.

Under nitrogen starvation, mutants in 53 genes had a negative effect while 36 mutations had a positive effect on fitness. Under phosphate deficiency, only mutants in 13 genes had a significant change in fitness. These results pave the way to deciphering genes and pathways important for nutritional signaling and utilization in *T. atroviride*, as well as tools to define processes important for interactions with other organisms in rhizosphere communities. .

**References**

1. Harnam, G. E. "Overview of Mechanisms and Uses of *Trichoderma* spp." *Phytopathology.* 2007.

2. Coradetti, Samuel T., et al. "Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides." *Elife* 7 (2018): e32110.

*Funding statement.*
*This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) an SFA led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231*

**Title: The Development of Plant and Soil Fabricated Ecosystems (EcoFABs) for Standardized Microbiome Experiments**

**Authors:** Peter Andeer,[1*] (pfandeer@lbl.gov), Kate Zhalnina[1], Kirsten Hofmockel,[2] Arunima Bhattacharjee,[2] Natalie Sadler,[2] Eoghan King,[3] John Vogel,[3] Joanna Coker,[4] Deepan Thiruppathy,[4] Karsten Zengler,[4] and **Trent Northen[1,3]**

**Institutions:** [1]Lawrence Berkeley National Laboratory, [2]Joint Genome Institute, [3]Pacific Northwest National Laboratory, [4]University of California, San Diego,

**Website:** https://ecofab-teams.lbl.gov/

**Project Goals: Short statement of goals.** The TEAMS project (Trial Ecosystems for the Advancement of Microbiome Science) is focused on the development of technologies and protocols to facilitate plant and soil microbiome experiments using fabricated ecosystems (EcoFABs). EcoFABs are intended to provide standardized methods for understanding microbial interactions in complex environments through the creation of environments that capture relevant aspects of their natural environment like spatial heterogeneity, but do so in a reproducible manner better suited for inter-laboratory investigations and designed to allow for multiple avenues of investigation (e.g., microscopy, metabolomics). Through the TEAMS project two types of EcoFABs have been developed, the EcoFAB 2.0, which can be mass produced and is compatible with robotics for automated plant-microbiome experiments, and the soil EcoFAB, which is designed to study microbial interactions within a defined spatial structure. Finally, the EcoBOT is being developed to facilitate automated EcoFAB experiments in a single system containing an integrated growth chamber, liquid handling robot, inverted microscope and hyperspectral camera. An automated arm allows the remote or programmed transfer of EcoFABs between stations to monitor plant development and collect metabolite and microbial samples throughout its growth to facilitate automated EcoFAB experimentation.

**Abstract text:**

The TEAMS project has organized inter-laboratory experiments to evaluate the reproducibility of microbial community analyses in plant microbiome experiments within the EcoFAB 2.0. These experiments follow up on a previously published inter-laboratory study that demonstrated that the impacts of growth media on the exudation, and root and shoot phenotypes of *Brachypodium distachyon* 21-3 were reproducible between laboratories when cultivated in EcoFABs[1]. This iteration is focused on the interactions between *B. distachyon* 21-3 and microbial communities and includes DNA sequencing to evaluate microbial colonization as well as metabolomics and phenotyping analyses. This study utilizes a standardized consortium of isolates made from a field grown switchgrass plant and selected to represent a phylogenetically diverse set of bacteria. All 17 of these isolates have been sequenced and evaluated for their reproducible growth *in vitro* under different nutrient conditions. Pilot EcoFAB experiments have demonstrated that inclusion of the entire community vs 16 members that exclude a

*Paraburkholderia* sp. have a significant impact on shoot fresh weight and root/shoot mass ratios. The details of this process will be investigated as part of a five institution intercomparison study. The results of this study will lay the foundation for an comparison study by scientists at all four of the DOE Bioenergy Research Centers. Together these studies will establish a common plant-microbiome experimental system to enable collaborative research studies. The results of these intercomparison studies will also be used to benchmark the automated EcoBOT system vs. conventional experimentation.

Researchers at PNNL have developed soil EcoFAB devices designed to complement the plant-microbe focused devices described above. These EcoFABs take advantage of custom polymers compatible with advanced mass spectrometry approaches to facilitate spatial mass spectroscopy imaging. They also provide solid structures of defined sizes and spacing to support microbial growth. Initial experiments were performed by inoculating each quadrant of the EcoFABs with 1 of 4 phylogenetically diverse bacteria expressing a fluorescent protein. After a week, confocal microscopy revealed that the bacteria each colonize the device in distinct, but reproducible manners. The reproducibility of the developed soil EcoFAB experiments will also be assessed using inter-laboratory evaluations involving laboratories at 4 separate institutions.

Together these capabilities provide a powerful new platform for scientists to build on each other's research and collaborate to advance microbiome science.

**References/Publications**

1. Sasse, J., Kant, J., Cole, B. J., Klein, A. P., Arsova, B., Schlaepfer, P., … Northen, T. R. (2019). Multilab EcoFAB study shows highly reproducible physiology and depletion of soil metabolites by a model grass. New Phytologist, 222(2), 1149–1160. https://doi.org/10.1111/nph.15662

**Funding statement:** This material by TEAMS – Trial Ecosystems for the Advancement of Microbiome Science is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory and FWP 77195 at Pacific Northwest National Laboratory. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

**Crosstalk: Interkingdom interactions in the mycorrhizal hyphosphere and ramifications for soil C cycling**

Edith Lai[1], Anne Kakouridis[2], Peter Kim[2], Trent Northen[2], Mary Firestone[2,3], **Erin E Nuccio**[1*] (nuccio1@llnl.gov)

[1]Lawrence Livermore National Laboratory, Livermore, CA, [2]Lawrence Berkeley National Laboratory, Berkeley, CA, [3]University of California, Berkeley, Berkeley, CA

**Project Goals: Arbuscular mycorrhizal fungi (AMF) are ancient symbionts that form root associations with most plants. AMF play an important role in global nutrient and carbon cycles, and understanding their biology is crucial to predict how carbon is stored and released from soil. This Early Career research investigates the basic mechanisms that underpin synergistic interactions between AMF and microbes that drive nitrogen and carbon cycling, addressing DOE's mission to understand and predict the roles of microbes in Earth's nutrient cycles. By coupling isotope-enabled technologies with next generation DNA sequencing techniques, our project investigates soil microbial interactions *in situ* using natural levels of soil complexity. This work will provide a greater mechanistic understanding needed to determine how mycorrhizal fungi influence organic matter decomposition and will shed light on large-scale nutrient cycling processes in terrestrial ecosystems.**

The arbuscular mycorrhizal association between the Glomeromycota fungi and land plants is ancient and widespread; approximately 80% of all land plants form symbiotic associations with AMF. Recent work has shown that mycorrhizal fungal type is one of the key predictors of soil C storage in terrestrial ecosystems. However, studies have come to opposing conclusions about the direction and magnitude of soil C accumulation in relation to mycorrhizal colonization. A greater mechanistic understanding is needed to determine how mycorrhizal fungi alter decomposition to help predict large scale C cycling processes in terrestrial ecosystems.

While AMF are obligate symbionts that depend on their host plant for C and cannot decompose soil organic matter (SOM), AMF can stimulate the decomposition of SOM and dead plant material. Our prior research strongly suggests that AMF partner with their microbiome in the zone surrounding hyphae, or hyphosphere, to encourage decomposition. The molecular mechanisms underpinning synergistic interactions between AMF and the microbial community during N uptake from SOM is a key knowledge gap limiting our ability to model these interactions. To determine how AMF harness hyphosphere microbial communities to stimulate decomposition, we are first using $^{13}$C-Stabile Isotope Probing (SIP) to identify the genomes of organisms in the fungal hyphosphere. We have sequenced $^{13}$C-SIP hyphosphere metagenomes collected from a $^{13}CO_2$ labeling experiment using living soil. Additionally, we are developing sterile plant-mycorrhizal microcosms ("MycoChips," based off the EcoFAB platform) that we can use to interrogate hyphal-microbial interactions *in-situ*. For notoriously heterogeneous environments such as soil, it is critical to develop and apply systems biology tools with the ability to interrogate soil microbial communities at their natural levels of complexity. Using this framework, our work aims to deconstruct complex interkingdom interactions in living soil.

*This research is supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Early Career award SCW1711, award numbers DE-SC0016247 and DE-SC0020163 to UC Berkeley, work at Lawrence Berkeley National Laboratory is supported by the m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils (m-CAFEs@lbl.gov) Scientific Focus Area. Work conducted at Lawrence Livermore National Laboratory was supported under the auspices of the U.S. DOE under Contract DE-AC52-07NA27344, and work at Lawrence Berkeley National Laboratory was performed under the auspices of the U.S. Department of Energy Contract No. DE- AC02-05CH11231.*

**Title:** Comparison of Michaelis-Menten and Elementary Decomposition Rate Kinetics for Predicting Enzyme Progress Curves

**Authors:** Wheaton Schroder [1,2]* (wheaton.schroeder@gmail.com), Costas D. Maranas,[1,2] and **Daniel G. Olson [1,3]**

**Institutions:** [1]The Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; [2]The Pennsylvania State University, University Park, PA; [3]Dartmouth College, Hanover, NH

**Project Goals:** *Clostridium thermocellum* is a promising organism for the production of fuels and chemicals from cellulose, but it has an atypical glycolysis that is poorly understood, which limits commercial applications. The overall goal of this project is to develop an improved systems-level understanding of *C. thermocellum* metabolism. We will do this by performing kinetic measurements of purified enzymes and kinetic analysis of enzyme modules in cell lysates. Results from these analyses will be synthesized with a novel multi-scale kinetic modeling approach. This improved understanding will be demonstrated by producing 2,3-butanediol from cellulose *in vivo*.

**Abstract Text:** Elementary decomposition (ED) provides a way to generate mechanistic kinetic models from reaction stoichiometry, and also a mathematical framework for determining kinetic parameters through non-linear optimization (Gopalakrishnan, Dash, and Maranas 2020) or screening sets of stochastically chosen kinetic parameters (Tran, Rizk, and Liao 2008), both of which attempt to determine enzyme kinetics through global fitting of kinetic parameters (i.e. a top-down approach). These approaches are somewhat limited in that the kinetic parameters estimated in this manner are based on relative enzyme and metabolite concentrations and often make use of lumped reactions and kinetic parameters. For critical metabolic enzymes where more accurate kinetic knowledge is desired, we are interested in using the ED approach for estimating kinetic parameters from progress-curve data for individual enzymes for parameterizing kinetic models of *C. thermocellum* (i.e. a bottom-up approach). To do this, we have created a Nonlinear Programming (NLP) approach. Using spectroscopy data, our approach minimizes error between the measured and calculated concentrations of one reaction participant by solving a system of equations describing the ED kinetics with variable kinetic parameters and participating species concentrations. The process as a whole then performs a bottom-up and absolute estimation of kinetic parameters. To validate our system, we are testing it on formate dehydrogenase (EC 1.2.1.2), a commercially available enzyme that is commonly used for recycling of $NAD^+$ to NADH in redox reactions. Furthermore, this is one of the few enzymes where a kinetic model has been validated on a complete progress curve, rather than just initial rates (Schmidt et al. 2010). We are interested in addressing the following questions:

1. How does prediction accuracy compare between the two models?
2. For each model, which parameters are well-resolved by time-course kinetic data?
3. How much inaccuracy is introduced by assuming the reaction is irreversible?

4. To what extent can the kinetic mechanism be determined from progress curve data?

**References/Publications**

Gopalakrishnan, Saratram, Satyakam Dash, and Costas Maranas. 2020. “K-FIT: An Accelerated Kinetic Parameterization Algorithm Using Steady-State Fluxomic Data.” *Metabolic Engineering* 61 (September): 197–205. https://doi.org/10.1016/j.ymben.2020.03.001.

Schmidt, T, C Michalik, M Zavrel, A Spiess, W Marquardt, and M B Ansorge-Schumacher. 2010. “Mechanistic Model for Prediction of Formate Dehydrogenase Kinetics under Industrially Relevant Conditions.” *Biotechnology Progress* 26 (1): 73–78. https://doi.org/10.1002/btpr.282.

Tran, Linh M, Matthew L Rizk, and James C Liao. 2008. “Ensemble Modeling of Metabolic Networks.” *Biophysical Journal* 95 (12): 5606–17. https://doi.org/10.1529/biophysj.108.135442.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012345.

**Title:** Engineering Synthetic Anaerobic Consortia Inspired by the Rumen for Biomass Breakdown and Conversion

**Authors:** Elaina Blair[1]* (elainablair@umail.ucsb.edu), Patrick Leggieri,[1] Joel Howard,[2] Scott Baker,[3] Chris Lawson[2], and **Michelle A. O'Malley[1]**

**Institutions:** [1]University of California, Santa Barbara, Santa Barbara, CA; [2]University of Toronto, Toronto, Canada; [3]Pacific Northwest National Laboratory, Richland, WA

**Project Goals:** This project will leverage a "synthetic rumen" consortium composed of anaerobic fungi and chain-elongating bacteria to study which metabolites are shared and exchanged between microbes and identify strategies to bolster lignocellulose conversion to value-added products. Our approach will develop high-throughput systems and synthetic biology approaches to realize stable synthetic consortia that route lignocellulosic carbon into short and medium chain fatty acids (SCFAs/MCFAs) rather than methane. Key research objectives are to (1) design and predict anaerobic fungal and bacterial consortia that efficiently convert lignocellulosic biomass into medium-chain fatty acids (MCFAs), (2) understand how fermentation parameters and microbe-microbe interactions regulate and drive microbiome metabolic fluxes, and (3) use genomic editing to alter the fermentation byproducts of anaerobic fungi and bolster MCFA titers and yields.

**Abstract Text:** Lignocellulose deconstruction and conversion in nature is driven by mixed microbial partnerships rather than the action of a single microbe. For example, microbes are particularly well optimized to recycle organic matter in anaerobic habitats, ranging from landfills to intestinal tracts, via interspecies $H_2$ transfer and methane release. Compared to aerobic processes, anaerobic digestion can be far more efficient in converting substrate to chemical products, largely because far less carbon is funneled to cell growth resulting in higher yields, and far less energy inputs are required because pre-treatment, aeration, mixing, and heat removal are greatly reduced. Compartmentalizing difficult biomass deconstruction and production steps among "specialist" anaerobes is an exciting new route to convert biomass into value-added products, especially if consortia can be built predictively, and engineered for stability.

Recently, we established a model bacterial consortium enriched from the rumen that converts lignocellulosic biomass into high titers of C4 VFAs (butyrate), based on a chain elongation process that inhibits archaeal methanogenesis. Several consortia were selectively enriched from the feces of cows and goats using sorghum as a substrate while suppressing methanogenesis with 2-bromoethanesulfonate (BES) addition. Metagenomic and metatranscriptomic analysis identified the key chain elongating bacteria in these consortia to be unique species of the *Lachnospiraceae* family that maintain high expression of the reverse β-oxidation pathway responsible for C4-C8 VFA production. This analysis also revealed several other bacterial species in the consortia that compete with *Lachnospiraceae* and reduce overall C4-C8 VFA yields by diverting carbon to unwanted products. Therefore, building synthetic consortia that

eliminate these competing bacteria would bolster product yields and enable great control over VFA chain length. In parallel, we also demonstrated that anaerobic rumen fungi within the *Neocallimastix* genus are superior biomass degraders compared to anaerobic bacteria from these enrichments. Moreover, the biomass degradation products lactate, acetate, and ethanol from *Neocallimastix* are optimal substrates for *Lachnospiraceae* and other chain elongators. Accordingly, partnering anaerobic fungi and chain elongating bacteria in synthetic consortia represents a novel strategy for maximizing lignocellulose conversion to C4-C8 VFAs. Here, we further describe our efforts to systematically characterize and model the production of VFAs from synthetic bacterial/fungal communities that have been grown on representative lignocellulosic grass substrates.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022142.

# Understanding Syntrophies within Methane Oxidizing Microbial Consortia: integrating Genome Scale Metabolic Models and Reactive Transport

**Nidhi Gupta**[1,*] (ngupta@anl.gov), Filipe Liu[1], José P Faria[1], Janaka Edirisinghe[1],Ranjani Murali[2], Sergio Parra[2], Heewon Jung[5], Shahram Asgari[3], Christof Meile[3], Mark Ellisman[4], Victoria Orphan[2] and Christopher Henry[1]

[1]Argonne National Laboratory, Lemont, IL, [2]California Institute of Technology, Pasadena, CA, [3]University of Georgia, Athens, GA, [4]University of California, San Diego, CA, [5]Chungnam National University, Daejeon, Republic of Korea

## Project Goals:

The overarching goal is to expand the understanding of interactions and fundamental activities involved in cycling of carbon and nutrients by syntrophic methanotrophic archaeal-bacterial consortia and associated viruses in anoxic sedimentary environments. Specific objectives are to (1) quantify energy and nutrient exchange [e.g., nitrogen (N), phosphorus (P), iron (Fe) and vitamins] within AOM consortia and between ANME-bacterial partners; (2) identify virus-host interactions associated with AOM and assess C and N transfer through viruses in methane - impacted sediment ecosystems; (3) model energy and nutrient exchange in AOM consortia and viral-host interactions (i.e., viral activity), and their environmental distribution patterns.

## Abstract:

Understanding the cycling of carbon and nutrients in the sediment environments requires an accounting of the feedback between microbial activities and environmental conditions. We approach this through the development of (1) flux balance community metabolic models consisting of methanotrophic archaea (ANME) and sulfate reducing bacteria (SRB) and (2) a reactive transport model that establishes a connection between processes at the pore scale and the macroscopic environment.

To build a community flux model, we started with the construction of flux balance models of ANME and SRB individually using ModelSEED in the DOE Systems Biology Knowledgebase (KBase) platform (www.kbase.us). ANME microbes are related to methanogenic bacteria but these organisms support “reverse methanogenesis” as the biochemical model for methane oxidation pathway and eventually for defining syntrophy. Towards this end, we curated pathways and model templates for archaeal microbes in our latest build of the ModelSEED genome-scale model reconstruction tool. The ModelSEED now has an archaea template with an archaea-specific biomass reaction and a more complete representation of archaeal metabolic pathways. On average, models of archaeal species constructed by the ModelSEED have 150 additional genes compared to before our curation efforts. To correctly mimic methane utilizing consortia behavior, flux constraints were also adjusted to ensure proper implementation of ANME behavior with respect to the flow of electrons along the methanogenesis pathway. In addition to archaeal template, significant progress was also made to improve the representation of the bacterial sulfur reduction pathway in the ModelSEED,curating the reactions and annotations based both on literature and

experimental data. This resulted in the addition of at least ten new reactions and unique annotations in a single sulfate reduction pathway. We will apply this improved ModelSEED tool to construct models for exemplar ANME and SRB strains assembled and binned from metagenomic data collected for this project. Presently, we are combining these models together in community flux simulations, while integrating transcriptomic data, to predict potential trophic interactions between various ANME and SRB strains under various conditions. Ultimately, these predictions will be tested in the reactive transport simulations described below, as well as new experimental studies.

To establish a connection between processes at the pore scale and the macroscopic environment, ongoing work aims at integrating microCT and multi-model imaging analysis into models that resolve the pore structure of carbonate rocks formed due to the alkalinity produced in the anaerobic oxidation of methane. CT scans of the rocks are collected through an x-ray microscope at an 8-micron isotropic resolution for the modeling component. Select, higher resolution scans are also collected at 0.8 microns for resolution of smaller pore network structures and the biomass of interest. This structural information is then being used to estimate the residence time distribution and local biogeochemical conditions that result from and shape niches for microbial metabolism.

**Funding statement** - This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0020373.

# Characterization of viral assemblages in methane-saturated sediments and their Spatio-temporal Dynamics

Alon Philosof[1*], Aditi Narayanan[2], Maria Alvarez Sanchez[3], Aitana Llorenç Vicedo[3], Francisco Martinez-Hernandez[3], Manuel Martinez Garcia[3], Mason Mackey[4], Eric Bushong[4], Thomas Deerinck[4], Mark Ellisman[4], **Victoria J. Orphan**[1,2]

*[1]Division of Geological and Planetary Sciences, Caltech, Pasadena, California, USA; [2]Division of Biology and Biological Engineering, Caltech, Pasadena, California, USA; [3]Departamento Fisiologia, Genetica y Microbiologia, Universidad de Alicante, Spain; [4]National Center for Microscopy and Imaging Research, University of California San Diego, USA*

**philosof@caltech.edu*

## Project Goals:

The overarching goal is to characterize the role of viruses in the ecology and biogeochemical cycling in methane saturated sediments , especially their impact on the syntrophic archaeal-bacterial consortia that perform the anaerobic oxidation of methane (AOM). Specific objectives are to (1) Characterize the viral community in methane-saturated sediments and their environmental distribution patterns using viral metagenomics and transmission electron microscopy; (2) Identify virus-host interactions associated with AOM and potential viral auxiliary metabolic genes (AMG) involved in key sedimentary biogeochemical cycles. (3) Develop activity-based methods for fluorescently-labeling viruses for microscopy and flow sorting to achieve single virus level resolution of newly produced viruses and their genomic diversity; (4) Quantify viral activity and constrain how the production of new viruses relates to host physiology and AOM.

Viral communities in sediment and soil environments are largely unexplored, though initial evidence exists for their large contribution to the dissolved organic carbon pool. In methane-saturated anoxic sediments, the process of anaerobic oxidation of methane by syntrophic consortia of methanotrophic archaea (ANME) and sulfate-reducing bacteria (SRB), is the dominant pathway by which methane is oxidized prior to release to the atmosphere.

To understand the role of viruses in this process, we have established a series of microcosm anoxic incubations of multiple sediment depths from four discrete cores collected from sites of active AOM. From these incubations, we have assembled viral metagenomes that yielded more than 3,400 complete and nearly complete viral metagenome-assembled genomes (vMAGs), along with thousands more contigs longer than 15kb. The viral community from this environment is very diverse and the majority of the vMAGs are unclassified followed by Siphoviruses and Halovirus-like viruses. Our data are dominated by site-specific abundance patterns rather than clustering by sediment depths. Auxiliary metabolic genes (AMG) involved in heme metabolism and assimilatory sulfate reduction are highly abundant throughout our dataset, pointing at the potential role of viruses in sulfur metabolism and possibly extracellular electron transfer.

To further characterize the impact of viruses on AOM and sulfate reduction a set of incubations were amended with $^{15}NH_4^+$, $^{13}CO_3$ and L-Homopropargylglycine (HPG). An additional set of incubations were also amended with anthraquinone-2,6-disolfonate (AQDS), an external electron acceptor that decouples the syntrophic relationship between the ANME and SRB. TEM analysis of a previous incubation series showed preliminary differences in viral capsid morphologies between decoupled and non-decoupled samples. The current incubation series seeks to track such changes in the viral community over time, and correlate it with metabolic activity of the hosts, viral production rates and the rate of elemental transfer between hosts and virus-like particles (VLP). Biorthogonal Non-Canonical Amino Acid Tagging (BONCAT)[2] with click-chemistry was used to fluorescently label newly-synthesized viruses and quantify viral production. BONCAT methodology was optimized for maximizing the ratio of viral signal to noise, which is crucial for flow cytometry detection and sorting of positive BONCAT tagged viral particles. Finally, we present preliminary results of a method that combines optimized viral-BONCAT with fluorescence-activated sorting (FACS) on an Influx cell sorter with Single Virus Genomics[3] (SVG) that can be used to identify and sequence the genomes of newly produced viruses in sediments from active host cells, correlated to geochemical data from the incubation.

**References**


1. Scheller, Silvan, Hang Yu, Grayson L. Chadwick, Shawn E. McGlynn, and Victoria J. Orphan. 2016. "Artificial Electron Acceptors Decouple Archaeal Methane Oxidation from Sulfate Reduction." Science (New York, N.Y.) 351 (6274): 703–7. https://doi.org/10.1126/science.aad7154.
2. Pasulka, Alexis L., Kimberlee Thamatrakoln, Sebastian H. Kopf, Yunbin Guan, Bonnie Poulos, Annie Moradian, Michael J. Sweredoski, et al. 2018. "Interrogating Marine Virus-Host Interactions and Elemental Transfer with BONCAT and nanoSIMS-Based Methods." Environmental Microbiology 20 (2): 671–92. https://doi.org/10.1111/1462-2920.13996.
3. Martinez-Hernandez, Francisco, Oscar Fornas, Monica Lluesma Gomez, Benjamin Bolduc, Maria Jose de la Cruz Peña, Joaquín Martínez Martínez, Josefa Anton, et al. 2017. "Single-Virus Genomics Reveals Hidden Cosmopolitan and Abundant Viruses." Nature Communications 8 (June): 15892. https://doi.org/10.1038/ncomms15892



*Funding statement.*

*This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0020373.*

# Development of *Anabaena* 33047, a fast-growing $N_2$-fixing cyanobacterium, as a carbon neutral bioproduction platform

Anindita Banerjee[1]* (anindita@wustl.edu), Deng Liu[1], John I. Hendry[2], Costas D. Maranas[2], Maciek R. Antoniewicz[3] and **Himadri B. Pakrasi**[1]

[1]Washington University in St. Louis, MO; [2]Pennsylvania State University, University Park, PA, [3]University of Michigan, Ann Arbor, MI

## Abstract

Cyanobacteria, oxygenic photosynthetic prokaryotes, are emerging as promising platforms for cost effective conversion of sunlight and $CO_2$ into high value end products. Recent discovery of fast-growing cyanobacterial strains, with growth rates comparable to their heterotrophic counterparts, have alleviated one of the long-standing bottlenecks in the commercialization of these organisms for bio-production. *Anabaena* sp. ATCC 33047, a heterocystous cyanobacterium, stands out in this category for its ability to utilize atmospheric $N_2$ in addition to $CO_2$ and generate biomass at unprecedented rates. Developing such a strain as a chassis can eliminate the need for fixed nitrogen in production systems, a significant step towards cost reduction. Our work aimed to develop *Anabaena* 33047 into such a bioproduction platform.

*Anabaena* 33047 exhibits rapid growth (~3.8h doubling time) utilizing high light, $CO_2$ and $N_2$ [1]. Under controlled culture conditions, this strain can fix up to 3.0 g $CO_2$ $L^{-1}$ $day^{-1}$, the highest conversion rate of atmospheric $CO_2$ known for cyanobacteria. Much of this fixed carbon (~47%) is secreted into the medium in the form of exopolysaccharides (EPS) with production rates as high as 1.4 g $L^{-1}$ [2]. Cyanobacterial EPS exhibits unique physical and chemical properties and harbors bioactive potential which can be explored for a variety of applications [3]. However, this strain was previously known to be recalcitrant to genetic manipulation and hence, despite its many appealing traits, remained largely unexplored. During this project, we have successfully developed a genetic manipulation system that has enabled targeted genome modifications.

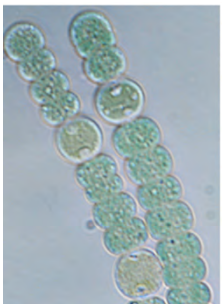


**Figure 1**. ***Anabaena* 33047** image showing high frequency of heterocysts.

Genome-scale metabolic models (GSM) facilitate comprehensive understanding of the metabolism of non-model organisms and metabolic engineering for over production of target chemicals [4]. In the course of this project, we developed a genome scale metabolic model, *i*AnC892 [5] for *Anabaena* 33047. This model will help expand our understanding of its metabolism and identify genetic interventions to divert carbon and nitrogen flux towards products of interest. The model was constructed by retrieving annotations from multiple databases: KEGG [6], MetaCyc [7] and ModelSEED [8] and a recently published model for the closely related strain *Anabaena* 7120 [9]. *i*AnC892 contains 953 unique reactions representing the annotation of 892 genes. The diazotrophic life cycle of *Anabaena* 33047 is captured by accounting for both vegetative and heterocyst cell types. This is achieved by creating super-compartments that reflect the metabolic differences and interactions between these two cell types (Figure 2). The model provided insight into importance of light dependent electron transport in the heterocyst and pointed pathway combinations that can supply reducing equivalents and ATP in the appropriate ratio for optimal $N_2$ fixation [5]. The model was used alongside the strain design algorithm, OptForce [10], to identify genetic interventions that would lead to overproduction of nitrogen-rich compounds [5].

To validate the existing metabolic models and to have a better understanding of the metabolism of *Anabaena* sp. ATCC 33047 under mixotrophic, heterotrophic, autotrophic and diazotrophic growth conditions, we performed a series of stable-isotope tracing experiments and quantified metabolic fluxes using state-of-the-art tools for $^{13}$C-metabolic flux analysis ($^{13}$C-MFA). Our results validated the core

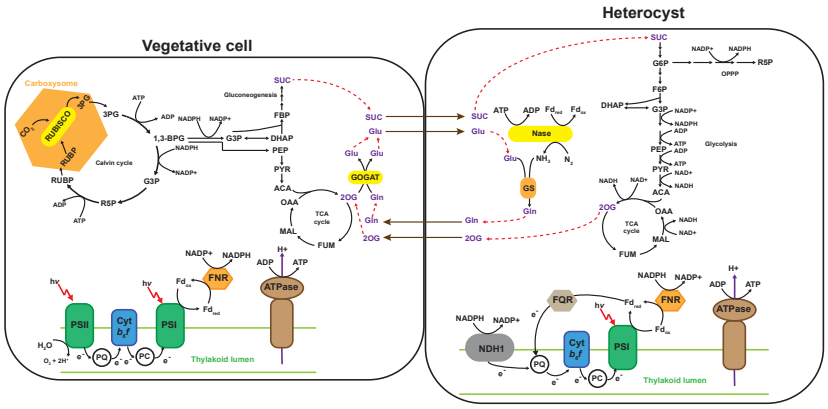



**Figure 2: Two cell model of *Anabaena* 33047.** The genome scale metabolic model, *i*AnC915, has two super-compartments, the vegetative cell and the heterocyst, in-order to capture the diazotrophic and heterocyst forming lifestyle of this cyanobacterium.

metabolic network model assumptions for this organism. To assess metabolite exchange between vegetative cells and heterocysts during diazotrophic growth, we attempted to obtain independent metabolomics and stable-isotope labeling data from vegetative cells and heterocysts. While we were able to successfully separate the two cell types, we observed significant leakage of metabolites from the cells, suggesting that further refinement of our protocol is needed to obtain more reliable metabolomics data.

Using our newly developed engineering strategy and based on our model predictions, we generated a Δ*nblA* strain of *Anabaena* 33047. NblA is involved in the degradation of phycobilisomes (PBS), light harvesting antenna proteins. The Δ*nblA* mutant exhibited resistance to PBS degradation and retained high amounts of these antenna proteins in its heterocysts. Quantitative analysis of PBS in individual heterocysts of the mutant and the WT (using a Fluorescence Kinetic Microscope - FKM) revealed ~ 8-fold higher amounts of PBS in the heterocysts of the mutant (Figure 3) [1]. Intriguingly, the Δ*nblA* mutant displayed ~ 2.5 folds higher rates of nitrogen fixation compared to the WT. Spectroscopic analysis revealed altered PSI kinetics in the mutant, with increased cyclic electron flow around PSI, a route that contributes to ATP generation and nitrogenase activity in heterocysts. Thus the Δ*nblA* mutant of *Anabaena* 33047 offers an improved platform for bioproduction. Overall, this project has laid the foundation for developing the non-model cyanobacterium *Anabaena* 33047 into a highly efficient chassis for bioconversion of solar energy and atmospheric $CO_2$ and $N_2$ into high value bio products.

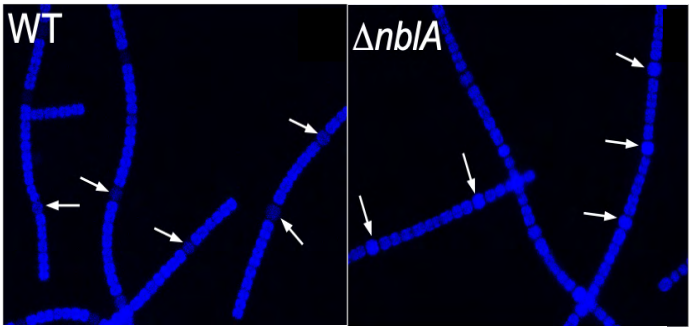



**Figure 3. FKM analysis of WT and Δ*nblA* heterocysts.** Bright signal from PBS in heterocysts of the Δ*nblA* mutant compared to the WT (arrows).

**References**

1. Bandyopadhyay A, Ye Z, Benedicty Z, Trtilek M, Pakrasi HB: Antenna modification leads to enhanced nitrogenase activity in a high light tolerant cyanobacterium. *mBio* 2021, **12**:e034082.
2. González López CV, Acién Fernández FG, Fernández Sevilla JM, Sánchez Fernández JF, Cerón García MC, Molina Grima E: Utilization of the cyanobacteria *Anabaena* sp. ATCC 33047 in $CO_2$ removal processes. *Bioresour Technol*. 2009, **100**:5904-10.
3. Pereira SB, Sousa A, Santos M, Araújo M, Serôdio F, Granja P, Tamagnini P: Strategies to Obtain Designer Polymers Based on Cyanobacterial Extracellular Polymeric Substances (EPS): *Int J Mol Sci*. 2019, **20**:5693.

4. Simeonidis E, Price ND: Genome-scale modeling for metabolic engineering. *J Ind Microbiol Biotechnol* 2015, **42**:327-338.
5. Hendry JI, Dinh HV, Sarkar D, Wang L, Bandyopadhyay A, Pakrasi HB, Maranas CD: A Genome-Scale Metabolic Model of Anabaena 33047 to Guide Genetic Modifications to Overproduce Nylon Monomers. *Metabolites* 2021, **11**.
6. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: The KEGG resource for deciphering the genome. *Nucleic Acids Res* 2004, **32**:D277-280.
7. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, et al.: The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res* 2018, **46**:D633-D639.
8. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 2010, **28**:977-982.
9. Malatinszky D, Steuer R, Jones PR: A comprehensively curated genome-scale two-cell model for the heterocystous cyanobacterium *Anabaena* sp. PCC 7120. *Plant Physiol* 2017, **173**:509-523.
10. Ranganathan S, Suthers PF, Maranas CD: OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLOS Computational Biology* 2010, **6**:e1000744.

Funding statement: This study is being supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019386

**Enhanced Resistance Pines for Improved Renewable Biofuel and Chemical Production**

**Gary F. Peter**[1], gfpeter@ufl.edu

[1]University of Florida, Gainesville, FL

**Project Goals:** Our goal is to genetically increase constitutive terpene defenses of loblolly and slash pine to enhance protection against pests and pathogens and at the same time expand terpene supplies for renewable biofuels and chemicals.

**Abstract**: Today, the southeastern U.S. hosts the world's largest biomass supply chain, annually delivering 17% of global wood products, more than any other country. This well-developed regional supply chain supports southern pine genetic improvement, seedling production and planting, silviculture, harvesting, and transportation annually delivering ~250 million tons of pine wood to integrated manufacturing facilities. In the SE U.S., 39 million acres of land not suited for food production are planted with genetically improved loblolly and slash pine seedlings selected and managed for fast growth and high wood yields. The SE also houses the U.S. pine chemicals industry the oldest and one of the largest renewable hydrocarbon chemical industries with favorable cost-competitiveness with petroleum derived feedstocks. Our focus is on increasing constitutive terpene production to enhance loblolly and slash pine resistance to pests and pathogens. Enhanced resistance in these commercial species is critical to protect against widespread losses as biotic pressures increase due to global warming, landuse change and introduced exotic organisms.

Increasing pine terpenes also is aligned well with the needs of the developing bioeconomy. Today, commercial scale collection of pine terpenes occurs from live trees by tapping, stumps by solvent based steam extraction, and from pulp mills as co-products. US pulp mills recover ~900,000 tonnes $y^{-1}$ of terpenes and fatty acids supporting specialty chemical biorefineries that compete in markets with petroleum derived feedstocks supporting our concept that biofuels from pine terpenes could be profitable without subsidy if supply was increased. Proven technologies exist to efficiently convert pine terpenes to biofuels: a 30 million gallon $y^{-1}$ bio-refinery produces renewable diesel from pine terpenes and fatty acids and pine monoterpenes can be efficiently dimerized to produce a replacement for JP10, the highest density jet fuel. Higher wood terpene content will increase the yield of bioenergy per unit mass in the rapidly developing wood pellet industry. Pine terpene supply is limited by the relatively low average wood terpene content.

Pine terpenes evolved as a primary chemical and physical defense system and are a main component of a durable, quantitative defense mechanism against pests and pathogens. The terpene defense traits are under genetic control and can be improved by breeding and genetic engineering. Our goal is to genetically increase constitutive terpene defenses of loblolly and slash pine to enhance protection against pests and pathogens and at the same time expand terpene supplies for renewable biofuels and chemicals. Objective one will integrate existing and new genome wide association genetic results with RNA expression, QTL mapping, and allele

frequency information in known high oleoresin flow selections and our breeding populations to discover and validate loblolly and slash pine alleles/genes that are important for resistance. Objective two will use information from objective one to accelerate breeding for increased resistance in loblolly and slash pine through marker assisted introgression, and will develop and test genomic selection models to accelerate breeding of resistant slash pine.

**Funding**: U.S. Department of Energy – USDA Biomass Feedstocks DE-SC0 019099

**SIP-OMICS: A Semi-Automated Pipeline for Isotopically-Targeted Community Analysis and Metagenomics**

Erin Nuccio[1], **Steve Blazewicz[1]*** (blazewicz 1@llnl.gov), Ashley Campbell[1], Marissa Lafler[1], Rex Malmstrom[2], **Bruce Hungate[3], Egbert Schwartz[3], Jennifer Pett-Ridge[1]**

[1]Lawrence Livermore National Laboratory, Livermore, CA; [2]Joint Genome Institute, Berkeley, CA; [3]Northern Arizona University, Flagstaff, AZ

**Website:** https://sc-programs.llnl.gov/soil-microbiome

**Project Goals: Microorganisms play key roles in soil carbon turnover and stabilization of persistent organic matter via their metabolic activities, cellular biochemistry, and extracellular products. Microbial residues are the primary ingredients in soil organic matter (SOM), a pool critical to Earth's soil health and climate. We hypothesize that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and we are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. We focus on soil moisture as a 'master controller' of microbial activity and mortality, since altered precipitation regimes are predicted across the temperate U.S. *Our SFA's ultimate goal is to determine how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues under changing moisture regimes.***

**Abstract:** Linking the identification of uncultivated microbes with their environmental function is a key ambition for microbial ecologists. While many techniques attempt to meet this goal, stable isotope probing—SIP—remains the most comprehensive for studying whole microbial communities *in situ*. In DNA-SIP, microbes who take up an isotopically heavy substrate end up with heavier DNA, which can be divided into multiple fractions containing DNA of different densities. Compounds labeled with $^{13}C$ or $^{15}N$ are frequently used to study the ecophysiology of organisms that consume a substrate of interest, while $^{18}O$ water is used as a universal tracer to measure the taxon-specific growth of all active taxa. However, SIP is not as broadly used as it could be because it requires specialized equipment, requires expensive reagents, is relatively low throughput and very time-consuming.

We designed a high throughput semi-automated DNA-SIP pipeline that can be combined with either amplicon or metagenomic sequencing. Our pipeline decreases operator time, reduces operator error, and improves reproducibility by targeting the most labor-intensive steps of traditional SIP—fraction collection, cleanup, and DNA processing. Fractionation is accomplished by connecting a SIP tube to an HPLC fraction collector (Agilent Infinity Fraction Collector), which aliquots four SIP tubes into a single plate 96-well plate. DNA precipitation and Picogreen quantification are then automated on a liquid handling robot (Hamilton STAR), which allows us to process density fractions from 16 SIP samples simultaneously. In addition, we have developed a method for pre-screening nucleic acids for isotopic enrichment, to ensure samples are adequately enriched prior to density gradient separation. Since establishing our pipeline, we have run over 1000 SIP samples, including well-replicated studies of annual grassland soil taxa active during key points in the water-year (fall wet up, spring growing season), plus analyses of

soil water limitation, redox, habitat and mycorrhizal effects. Overall, the pipeline reduces the per sample processing time from 9 hours to 1.7 hours, with ca. 30 minutes of manual work per sample. Using *in silico* analysis, we determined that 9 fractions are an ideal number to identify enriched organisms in samples with greater than 5 atom percent enrichment—this level of resolution balances of financial costs of extra fractions versus the benefits of error reduction [1].

We further improved our SIP practices by including internal standards during ultracentrifugation, which can calibrate per sample conversions of GC content to mean weighted density and determine the $^{18}O$ atom percent enrichment of a taxon's genome. The internal standards are two 9Kb PCR products of known isotopic enrichment. The more isotopically enriched standard typically appears in ultracentrifuge tubes approximately 0.05 g/ml heavier than the non-enriched standard, indicating about 3.3 extra neutrons per base pair are required for a DNA molecule to be isolated in a subsequent fraction. Using this methodology, the atom percent enrichment of a taxon's genome can be quantified in comparison to the internal calibration standard, which greatly improves the reproducibility of SIP runs because the calculation of sequence enrichment does not rely solely on refractometry. Refractometry measurements may vary between different operators and laboratories leading to incorrect measurements of isotopic enrichment of DNA contained in that fraction. Our semi-automated SIP approach and calibration of internal standards should make isotope-enabled techniques more high-throughput, more reproducible, accessible to the greater scientific community and allow better comparison of results among different experiments.

## References

1. Sieradzki, E.T., et al., *Measurement Error and Resolution in Quantitative Stable Isotope Probing: Implications for Experimental Design.* mSystems, 2020. **5**(4): p. e00151-20.

*This research is based upon work of the LLNL 'Microbes Persist' Soil Microbiome SFA, supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number SCW1632 to the Lawrence Livermore National Laboratory, and a subcontract to Northern Arizona University. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.*

# The biogeographic distribution of genomic traits between soil microbial communities

Peter F. Chuckran[1]* (pfc25@nau.edu), Cody Flagg[2], Jeffrey Propster[1], William A. Rutherford[3], Ella Sieradzki[4], **Steven J. Blazewicz[5], Bruce Hungate[1], Jennifer Pett-Ridge[5], Egbert Schwartz[1], Paul Dijkstra[1]**

[1]Center for Ecosystem Science and Society (ECOSS) and Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona; [2]National Ecological Observation Network (NEON), Boulder, Colorado; [3]School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona; [4]Department of Environmental Science, Policy, and Management, University of California, Berkeley, California; [5]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California

**Website:** https://sc-programs.llnl.gov/soil-microbiome

**Project Goals: Microorganisms play key roles in soil carbon turnover and stabilization of persistent organic matter via their metabolic activities, cellular biochemistry, and extracellular products. Microbial residues are the primary ingredients in soil organic matter (SOM), a pool critical to Earth's soil health and climate. We hypothesize that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. We focus on soil moisture as a 'master controller' of microbial activity and mortality, since altered precipitation regimes are predicted across the temperate U.S. *Our SFA's ultimate goal is to determine how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues under changing moisture regimes.***

**Abstract:**
Linking microbial communities to broad ecosystem scale processes is a central focus in microbial ecology. Here, we analyzed over 500 metagenomes to examine how fundamental microbial characteristics relate to environmental parameters, primarily in soil microbial communities. Our study combines traditional statistical analyses and a machine-learning approach to determine the environmental factors which most strongly control the genomic traits in soil microbial communities. This work provides valuable insight into the forces dictating microbial life-strategies in soil, which could assist in determining the environmental services soil microbes may provide.

Environmental conditions exert selective pressure on bacteria in microbial communities, and these pressures are often exhibited in the genomic traits of these bacteria. Traits such as genome size, GC content, and amino acid and codon frequency can lend valuable insight into the relationship between environment pressures and bacterial life-strategy [1]. For example, bacteria in nutrient limited environments often exhibit smaller genomes in order to curb the cost of reproduction, and lower GC content in order to conserve nitrogen (as the GC base pair has a higher nitrogen content than the AT base pair) [2]. While genomic traits could be used in the assessment of microbial communities across broad scales, little work has been done to examine how genomic traits are distributed among different environments. Further, the factors which

shape traits in soil microbial communities remains especially understudied. Here we present findings from two studies which look at the distribution of genomic traits between microbial communities. In the first, we compared 100+ publicly available metagenomes from the Joint Genome Institute [3] which were collected from soils, oceans, animal microbiomes, and hot springs, to evaluate how genomic traits were distributed across fundamentally different systems. We found that the relationships between genomic traits were unique among these different ecosystems and notably, that the relationship between the GC content and average genome size of communities was different between soil and marine environments. The GC content and average genome size of ocean microbial communities were positively correlated—consistent with results from previous studies examining individual bacteria taxa. However, in soils we found that larger genomes demonstrated comparatively lower GC content, indicating that a different set of environmental constraints are dictating the traits of soil bacteria. In our second study we examined this relationship more closely and compared the genomic traits of soil microbial communities across 400 metagenomes and assessed how these traits were related to numerous environmental parameters. These data, accessed from the National Ecological Observation Networks [4], showed that the genomic traits of soil bacteria were most closely related to pH. In low pH soils, soil microbial communities tended to have larger genomes with a lower GC content, and communities in high pH soils often had smaller genomes with higher GC content. We suggest that this reflects the several environmental parameters which are reflected in soil pH, specifically: soil carbon to nitrogen ratios, moisture availability, and physiological stress.


1. Westoby M, Gillings MR, Madin JS, Nielsen DA, Paulsen IT, Tetu SG. Trait dimensions in bacteria and archaea compared to vascular plants. *Ecol Lett* 2021; **24**: 1487–1504.
2. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. *ISME J* 2014; **8**: 1553–1565.
3. Chen IMA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* 2019; **47**: D666–D677.
4. National Ecological Observatory Network (NEON). Soil microbe metagenome sequences (DP1.10107.001). 2021.



*This research is based upon work of the LLNL 'Microbes Persist' Soil Microbiome SFA, supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number SCW1632 to the Lawrence Livermore National Laboratory, and a subcontract to Northern Arizona University. Support for PC was also provided from the USDA National Institute of Food and Agriculture Foundational Program (award #2017-67019-26396). Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.*

**Microbes Persist: Towards Quantitative Theory-Based Predictions of Soil Microbial Fitness, Interaction and Function in KBase**

Ulas Karaoz[3]* (ukaraoz@lbl.gov), Gianna Marschmann[3], Jeff Kimbrel[1], Ben Koch[2], **Steve Blazewicz**[1], **Bruce Hungate**[2], **Eoin Brodie**[3], **Jennifer Pett-Ridge**[1]

[1]Lawrence Livermore National Laboratory, Livermore, CA, [2]Northern Arizona University, Flagstaff, AZ; [3]Lawrence Berkeley National Laboratory, Berkeley, CA

**Website:** https://sc-programs.llnl.gov/soil-microbiome

**Project Goals: Microorganisms play key roles in soil carbon turnover and stabilization of persistent organic matter via their metabolic activities, cellular biochemistry, and extracellular products. Microbial residues are the primary ingredients in soil organic matter (SOM), a pool critical to Earth's soil health and climate. We hypothesize that microbial cellular chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and we are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. We focus on soil moisture as a 'master controller' of microbial activity and mortality, since altered precipitation regimes are predicted across the temperate U.S. *Our SFA's ultimate goal is to determine how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues under changing moisture regimes.***

**Abstract:** Our SFA has pioneered new methods to quantify element fluxes with genome-resolved taxonomic resolution —especially quantitative stable isotope probing (**qSIP)**, which allows us to evaluate *in situ* activity of individual taxa in complex communities by adding isotope tracers such as $^{18}$O-enriched 'heavy water' or $^{13}$C-enriched compounds. We have recently developed a computational workflow that accepts both amplicon or metagenomic sequence SIP input and calculates atom fraction excess (enrichment) as well as ***growth*** and ***mortality*** rates for individual amplicon sequence variants (ASVs) or metagenome assembled genomes (MAGs)**.** Experiments using $^{18}$O-$H_2O$ labeling and qSIP provide critical information on organism growth rates and mortality *in situ*. The analytical pipelines we are developing within KBase establish a standard qSIP analytical workflow, and a qSIP database suitable for robust cross-site comparisons and for model benchmarking. The workflow will enable uniform bioinformatics and calculations of qSIP data (e.g., a uniform approach to density shift calculations), and the database will facilitate robust comparisons across experiments. Integration within KBase will support analyses that compare traits of organisms with their performance in nature, across environments.

The qSIP pipeline is fully integrated with a genomes-to-trait workflow (**microTrait**) and compatible with a dynamic energy budget-based trait-based model (**DEBmicroTrait**). With microTrait and DEBmicroTrait, we have developed and tested a computational workflow to (1) infer ecologically relevant traits from microbial genomes, (2) systematically reduce the high-dimensionality of genome-level microbial trait data by inferring functional guilds (sets of organisms performing the same ecological function irrespective of their phylogenetic origin), (3) quantify within-guild trait variance and capture trait linkages in trait-based models, (4) explore trait-based simulations under different scenarios with varying levels of microbial community and

environmental complexity, and (5) benchmark emergent model substrate utilization (digested as chemical abundance data) and qSIP-derived growth and mortality rates (from qSIP database).

Combining both the qSIP and DEB-microTrait tools within KBase will provide a strong foundation for researchers who wish to use quantitative *in situ* measurements of microbial ecophysiology and population dynamics to benchmark models and build a predictive understanding of biological processes controlling material fluxes in complex environments.

*This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program as part of a KBase supplementary project to the LLNL 'Microbes Persist' Soil Microbiome SFA, under Award Number SCW1632 and SCW1746 to the Lawrence Livermore National Laboratory, Lawrence Berkeley National Laboratory and Northern Arizona University. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344. Work at Lawrence Berkeley National Laboratory was performed under the auspices of the U.S. Department of Energy Contract No. DE- AC02-05CH11231.*

**Trait-based modeling approach to carbon use efficiency**

Gianna Marschmann[1*] (glmarschmann@lbl.gov), **Jinyun Tang[1]**, Ulas Karaoz[1], Kate Zhalnina[1], Heejung Cho[1], **Jennifer Pett-Ridge**[2], **Eoin Brodie[1]**

[1]Earth and Environmental Sciences Area, Lawrence Livermore National Laboratory, Berkeley, CA; [2]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA

**Website:** https://sc-programs.llnl.gov/soil-microbiome

**Project Goals: Microorganisms play key roles in soil carbon turnover and organic matter stabilization via their metabolic activities, cellular biochemistry, and extracellular products. Microbial residues are the primary ingredients in soil organic matter (SOM), a pool critical to Earth's soil health and climate. We hypothesize that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and we are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. We focus on soil moisture as a 'master controller' of microbial activity and mortality, since microbial substrate uptake depends strongly on moisture availability and altered precipitation regimes are predicted across the temperate U.S.** ***Our SFA's ultimate goal is to determine how soil microbial ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues under changing moisture regimes.***

**Abstract:** To understand the ecology and biogeochemistry of plant-soil-microbe interactions, it is necessary to quantify and model how substrates control growth and interactions among biological organisms (and abiotic factors, e.g., adsorptive mineral soil surfaces). To address these substrate-consumer relationships, many substrate- kinetics and growth rules have been developed, including the famous Monod kinetics for single-substrate-based growth and Liebig's law of the minimum for multiple-nutrient-colimited growth. However, the mechanistic basis that leads to these various concepts and mathematical formulations and the implications of their parameters are often quite uncertain. Here, we show that an analogy based on Ohm's law in electric circuit theory can unify many of these different concepts and mathematical formulations [1]. We outline how this Ohm's law analogy can be integrated with Dynamic Energy Budget (DEB) models of microbial metabolism that partition biomass into reserve and structural components, and compute structural growth based on reserve dynamics. This biomass partitioning separates extracellular substrate competition from internal microbial physiological tradeoffs (e.g., maintenance vs. growth vs. enzyme production), enabling the tracking of dissipative loss associated with microbial substrate assimilation, somatic maintenance, structural biomass growth, and extracellular enzyme production, and allowing power-yield tradeoffs in metabolism to emerge dynamically.

Evolutionary trade-offs between power and yield can be observed in genomes where higher rRNA operon copy numbers scale positively with power (growth rate) and negatively with yield (carbon-use efficiency - CUE). Recent modeling work [2] predicted that intrinsic physiological differences between microorganisms can account for substantial variance in predicted yield, with

genome size negatively correlating with yield. In a model system representing rhizosphere and bulk soil – derived from an annual grassland soil in California, where we have 39 sequenced genomes representing *Proteobacteria, Actinobacteria*, *Bacteroidetes, Firmicutes and Acidobacteria*, we find that the power-yield phase space is partitioned into distinct trade-off regions, either dominated by maintenance requirements (positive relationship between power and yield) or relating to carbon and energy allocation (negative relationship). For growth on multiple distinct carbon sources, power-yield strategies of isolates fall on a Pareto optimal curve. Bacteria with larger genomes have less flexible power-yield trade off strategies, suggesting a trade-off between catabolic diversity and metabolic flexibility in isolates occupying distinct ecological niches in rhizosphere vs. bulk soil. Analysis of variance of genomic features and substrate properties furthermore revealed that yield might be a species-specific metabolic property: across substrates, isolate identity explained 38% of the variation in CUE and 88% of the variation within a substrate type. Class (phylum) taxonomic order explained 20% (13%) of the variation in CUE across all substrates and 69% (63%) of the variation within each substrate type. Changes in substrate supply will alter the performance of individual isolates; e.g., in the rhizosphere, we find that low Gibbs energy substrates that are preferentially assimilated select for organisms with metabolic traits that result in higher CUE, but we predict that these changes will not be as strong as changes in CUE that arise owing to differences in community composition. These changes and the potential for manipulating CUE through substrate addition in the rhizosphere can be further explored within our genome-informed, network-based modeling framework for microbial metabolism and plant biogeochemistry.

**References**

1. Tang, J., Riley, W. J., Marschmann, G. L., & Brodie, E. L. (2021). Conceptualizing biogeochemical reactions with an Ohm's law analogy. *Journal of Advances in Modeling Earth Systems*, *13*(10), e2021MS002469.
2. Saifuddin, M., Bhatnagar, J. M., Segrè, D., & Finzi, A. C. (2019). Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nature Communications*, *10*(1), 1-10.

*This research is based upon work of the LLNL 'Microbes Persist' Soil Microbiome SFA, supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number SCW1632 to the Lawrence Livermore National Laboratory, and a subcontract to Lawrence Berkeley National Laboratory. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344. Work at Lawrence Berkeley National Laboratory was performed under the auspices of the U.S. Department of Energy Contract No. DE- AC02-05CH11231*

**Functional genomics of replicating microbes and viruses following rewetting of a Mediterranean grassland soil**

Ella T. Sieradzki[1]* (ella.shir@gmail.com), Alexa Nicolas[1], Katerina Estera-Molina[3], **Jennifer Pett-Ridge[3], Mary K. Firestone[1,2], Jillian Banfield[1,2], Steven J. Blazewicz[3]**

[1]University of California Berkeley, Berkeley, CA; [2]Lawerence Berkeley National Laboratory, Berkeley, CA; [3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA

**Website**: https://sc-programs.llnl.gov/soil-microbiome

**Project Goals: Microorganisms play key roles in soil carbon turnover and stabilization of persistent organic matter via their metabolic activities, cellular biochemistry, and extracellular products. Microbial residues are the primary ingredients in soil organic matter (SOM), a pool critical to Earth's soil health and climate. We hypothesize that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. We focus on soil moisture as a 'master controller' of microbial activity and mortality, since altered precipitation regimes are predicted across the temperate U.S. *Our SFA's ultimate goal is to determine how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues under changing moisture regimes.***

**Abstract**: Rewetting of soil stimulates a succession of microbial growth and mortality, a process that could potentially become more frequent as climate change in semi-arid zones is predicted to lead to less rain events, potentially allowing for soil dry-down between events. Hypothetically, certain microbial traits, such as degradation of carbohydrates and acquisition of nitrogen, underlie this succession and confer advantages for growth as both the soil microbial community and available resources change over time. We hypothesized that some of the mortality during this succession is due to viral predation of growing organisms (i.e., Lotka–Volterra "kill the winner" dynamics). We also hypothesized that the summer dry down of soil would drive phages to integrate into host chromosomes and that wet-up of dry soil serves an environmental inducer of temperate phages.

To determine the mechanisms driving microbial growth and mortality during wet-up, we performed a wet-up experiment using soils that had been previously $^{13}CO_2$ labeled and maintained under one of two precipitation regimes: the historical average precipitation (100%) and a 50% water reduction. Following the annual summer dry period, soils were collected and incubated with multiple isotopic treatments. 'Heavy water' ($^{18}O$-$H_2O$) additions were used to specifically target the active portion of the microbiome and virome. Samples were harvested at six times following rewetting (0, 3, 24, 48, 72, 168 hr) for DNA-quantitative stable isotope probing (qSIP), metagenomics, viromics, and $CO_2$ production.

While total soil respiration did not vary between soils exposed to 100% versus 50% precipitation, respiration of new (labeled) rhizodeposits was higher in the 100% soils, implying functional differences between precipitation groups. This result was supported by differential abundance of traits found in growing ($^{18}O$ labeled) microorganisms, which revealed differences between precipitation treatments, as well as successional patterns with time. Differential abundance of traits also revealed an extreme legacy effect of historic precipitation, which was 1-3 orders of magnitude higher than any temporal changes. The effect peaked at 0 h but waned quickly and disappeared by 168 h, implying that the community "restarts itself" annually. Abundance of pathways for carbohydrate degradation in growing organisms varied over time, e.g., an increase in abundance of cellulose degradation at 48 h and 72 h, implying changes in complex C availability. N acquisition, while varying little over time, appeared to depend mainly on ammonium transport and assimilation pathways, as well as extracellular proteases, but not other complex N degradation pathways or dissimilatory inorganic N processes.

In comparison to temporal abundance patterns in microorganisms, viruses displayed spatially heterogeneity in addition to temporal community changes. Actively replicating viruses were mostly phages and were detected in dry soil (0 h) as virus-like particles. The vast majority of viruses sampled (83%) did not encode an integrase gene, implying a non-lysogenic life cycle. The low prevalence of putatively lysogenic phage in this wet-up dataset is also underscored by the fact that non-integrase-containing viruses increased by 24 h after rewetting, while integrase-containing viruses did not increase in abundance with time, suggesting that wetting of dry soil does not induce integrated phages. We conclude that the role of lysogeny in soil viral infections is lesser than generally hypothesized.

In summary, we observed temporal changes in active microbial and viral communities following wet-up which were underpinned by organic carbon and organic nitrogen degradation capabilities, as well as by lytic infection by viruses.

*This research is based upon work of the LLNL 'Microbes Persist' Soil Microbiome SFA, supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number SCW1632 to the Lawrence Livermore National Laboratory, with a subcontract to the University of California, Berkeley. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.*

**Soil habitats and water limitation shape microbial traits correlated with formation of mineral-associated organic matter**

Noah W. Sokol (sokol1@llnl.gov)[1]*, Megan M. Foley[2], Katerina Estera-Molina[1,3], Alex Greenlon[3], Eric Slessarev[1], Jose Liquet[1], **Mary Firestone[3], Bruce Hungate[2], Steven Blazewicz[1], Jennifer Pett-Ridge[1]**

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, [2]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, [3]Department of Environmental Studies, Policy, and Management; University of California Berkeley, Berkeley, CA

**Website:** https://sc-programs.llnl.gov/soil-microbiome

**Project Goals: Microorganisms play key roles in soil carbon turnover and stabilization of persistent organic matter via their metabolic activities, cellular biochemistry, and extracellular products. Microbial residues are the primary ingredients in soil organic matter (SOM), a pool critical to Earth's soil health and climate. We hypothesize that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and we are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. We focus on soil moisture as a 'master controller' of microbial activity and mortality, since altered precipitation regimes are predicted across the temperate U.S. *Our SFA's ultimate goal is to determine how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues under changing moisture regimes.***

**Abstract:** Soil microorganisms are frontline managers of the terrestrial carbon cycle. To better understand and model their effects under a changing climate, it is critical to determine which microbial ecophysiological traits are associated with soil organic matter formation – particularly mineral-associated organic matter (MAOM). Yet major uncertainty surrounds the traits that regulate this process, and how environmental context (e.g. spatial habitat, moisture conditions) shapes the manifestation of these traits. Microbial carbon-use efficiency (CUE) is posited to be positively correlated with MAOM formation, yet direct evidence for this relationship is sparse, and few other microbial traits have been directly tested as predictors of MAOM formation.

To investigate the relationship between different microbial traits and MAOM, we conducted a 12-week $^{13}C$ tracer study to track the movement of rhizodeposits and root detritus into microbial communities and SOM pools under moisture replete (15 ± 4.2 %) or water-limited (8 ± 2%) conditions. Using a continuous $^{13}CO_2$-labeling growth chamber system, we grew the annual grass *Avena barbata* for 12 weeks and measured formation of $^{13}C$-MAOM from either $^{13}C$-enriched rhizodeposition or decomposing $^{13}C$-enriched root detritus. We also measured active microbial community composition (via $^{13}C$-quantiative stable isotope probing; qSIP) and a suite of microbial traits including carbon-use efficiency, growth rate, and turnover (via the $^{18}O$-$H_2O$ method), extracellular enzyme activity, bulk $^{13}C$-extracellular polymeric substances (EPS), and total microbial biomass carbon ($^{13}C$-MBC).

We found that microbial traits associated with MAOM formation were distinct between the rhizosphere versus the detritusphere, and their effect was influenced by soil moisture. In the rhizosphere, fast growth and turnover were positively associated with MAOM, as were total $^{13}$C-MBC and $^{13}$C-EPS production. In contrast, growth rate was negatively associated with MAOM formation in the detritusphere, as were CUE, $^{13}$C-MBC, and $^{13}$C-EPS. However, total extracellular enzyme activity was positively associated with MAOM in the detritusphere. These results, paired with data on the chemical composition of MAOM (via STXM-NEXAFS) suggest that traits associated with fast growth and turnover, as well as high necromass yield, generate microbial-derived MAOM in the rhizosphere, whereas traits associated with resource acquisition generate plant-derived MAOM in the detritusphere. In these soil habitats, $^{13}$C-qSIP indicated that fungal taxa were more active in the detritusphere, whereas certain bacterial phyla (e.g., *Firmicutes*) were more active in the rhizosphere. Together, our results show that the rhizosphere has distinct traits, communities, and pathways of MAOM formation relative to the detritusphere. Future research should consider a broad suite of microbial traits – including but not limited to CUE – to model the role of microbes in MAOM formation in distinct soil environmental conditions.

*This research is based upon work of the LLNL 'Microbes Persist' Soil Microbiome SFA, supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number SCW1632 to the Lawrence Livermore National Laboratory, and a subcontract to the Northern Arizona University. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.*

# Environmental conditions shape active viral community structure and virus-host dynamics in soil ecosystems

Gareth Trubl[1]* (trubl.1@llnl.gov), Jeffrey A. Kimbrel[1], Ashley Campbell[1], Dariia Vyshenska[2], Simon Roux[2], Rex Malmstrom[2], Emiley Eloe-Fadrosh[2], Mark P. Waldrop[3], **Steven J. Blazewicz[1]**, **Jennifer Pett-Ridge[1]**

[1]Lawrence Livermore National Laboratory, Livermore, CA; [2]Joint Genome Institute, Berkeley, CA; [3]U.S. Geological Survey, Menlo Park, CA

**Website:** https://sc-programs.llnl.gov/soil-microbiome

**Project Goals: Microorganisms play key roles in soil carbon turnover and stabilization of persistent organic matter via their metabolic activities, cellular biochemistry, and extracellular products. Microbial residues are the primary ingredients in soil organic matter (SOM), a pool critical to Earth's soil health and climate. We hypothesize that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. We focus on soil moisture as a 'master controller' of microbial activity and mortality, since altered precipitation regimes are predicted across the temperate U.S. *Our SFA's ultimate goal is to determine how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues under changing moisture regimes.***

**Abstract:** Soil viruses play a large but uncharacterized role as regulators of microbial growth dynamics and microbial biogeochemistry, largely by controlling microbial activity via cell lysis and redirection and modulation of host metabolism. Research on the diversity, functional capabilities, and host predictions of soil viruses is expanding near-exponentially, however, current methods are fragmented, limiting our ability to make connections important to soil ecology. We hypothesize that virus-microbe dynamics, and microbial cellular-chemistry and ecophysiology fundamentally shape soil carbon persistence. Our SFA team is characterizing these phenomena via stable isotope probing (SIP) combined with meta-omics approaches including metagenomics, viromics, metatranscriptomics, and eDNA. Here, we used stable isotope probing (SIP) targeted metagenomics to reveal the genomic potential of active microbial and viral communities using two different isotopically labelled substrates. We used heavy-water ($^{18}O$) SIP in Arctic peat soils to identify and track active microbes and viruses over a year under subzero and anoxic conditions. We found that active bacterial populations represented only a small portion of the detected microbial community, were capable of fermentation and organic matter degradation, and were responsible for significant $CO_2$ production throughout the entire incubation. In contrast, the active viral populations in our arctic soil represented a large portion of the detected viral community and one-third were linked to active bacterial populations. In a second study, we incubated $^{13}C$-plant biomass in a tropical forest soil under four redox treatments to track viruses infecting microbes that degrade organic matter. Viral diversity was highest in the oxic samples and decreased in soils with lower $O_2$ exposure. More than a quarter of the soil viruses infected key active microbial organic matter degraders, and many of these were present only in anoxic samples. These findings highlight the impact temperature and soil redox conditions have on microbial and viral community structure and the fate of organic matter

in soils. Similar studies can help us: (1) learn more about how microorganisms grow and die in soil and how those factors mediate soil organic matter formation, (2) predict more accurately the impacts of shifting climate conditions on carbon cycling and biosequestration in ecosystems, and (3) examine potential means of biological sequestration of carbon.

*This research is based upon work of the LLNL 'Microbes Persist' Soil Microbiome SFA, supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number SCW1632 to the Lawrence Livermore National Laboratory, and LLNL LDRDs 18-ERD-041 (S. Blazewicz) and 21-LW-060 (G. Trubl). Work at Lawrence Berkeley National Laboratory was performed under the auspices of the U.S. Department of Energy Contract No. DE- AC02-05CH11231. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.*

**Design and -omics exploration of synthetic communities**

Alexis L. Marsh[1]* (almarsh@iastate.edu), Jia Wang,[2] Dale A. Pelletier,[2] Myra B. Cohen,[1] Paul E. Abraham,[2] David J. Weston,[2] Christopher S. Henry,[3] Janaka N. Edirisinghe,[3] Paramvir S. Dehal,[4] Robert Cottingham,[2] Mitchel J. Doktycz,[2] and **Priya Ranjan**[2]

[1]Iowa State University, Ames, IA; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]Argonne National Laboratory, Lemont, IL; and [4]Lawrence Berkeley National Laboratory, Berkeley, CA

https://pmiweb.ornl.gov/ ; https://kbase.us

**Project Goals: Synthetic communities (SynComs) offer a key experimental platform for examining how environmental changes affect the structure of microbiome and host physiology and productivity. As part of this effort, we propose adding new functionality and workflows to KBase, in support of the Plant-Microbe Interfaces SFA at Oak Ridge National Laboratory. With these new tools in KBase, microbes can be identified with beneficial properties that have the ability to coexist as communities. We propose new KBase Apps to simplify various steps of microbial community design for SynCom experiments. Additionally, we will add functionality to help validate SynCom experiments.**

Synthetic communities (SynComs) can be used to obtain a deeper understanding of specific mechanisms that drive community assembly, how multiple biotic interactions take place, and the expression of functional properties that occur when microbes, plants, and the environment interact simultaneously in time and space. Furthermore, the capability to obtain correlated data on transcript, protein, and metabolite abundance from a designed synthetic community experiment facilitates linking functional processes to molecular level information and helps improve the quality of annotation and predictive models. For Plant-Microbe Interfaces Science Focus Area (PMI SFA) experiments, SynComs are crucial for taking advantage of the extensive set of plant and microbial resources assembled in order to further investigate and understand in more precise detail the productive relationship between a plant and its microbiome.

As experimental data and design complexity grow, computational modules and applications are required to support rational design of SynComs based on desired functional and phenotypic characteristics and the ability of members of the community to survive together. A more comprehensive interpretation of the resulting experimental -omics data is needed to accelerate the design, build, test, and learn cycles associated with SynCom experiments. KBase, a DOE Genomic Sciences Program funded, public, and freely accessible software and data science platform, is ideal for developing such applications as it already provides a wide range and increasing number of relevant tools for functional annotation, metabolic modeling, auxotrophy prediction, substrate utilization and production of by products, taxonomy information, and prediction of microbial traits (*e.g.*, nitrogen fixation) that are important considerations in the rational design of SynComs and its members.

We propose to add *Populus*-associated microbiome and isolate genome data generated as part of PMI SFA, and Apps to KBase that leverage existing KBase capabilities for rational SynCom design. The -omics data resulting from the designed experiments will be integrated back into KBase and stored as abundance matrices. The four Apps will: **1) Summarize functional**

**potential of genome set, 2) Generate synthetic communities functionally similar to a metagenome, 3) Design synthetic microbial communities with desired characteristics, and 4) Compare groups of genomes from SynCom experiments.** The resulting workflow will advance a more productive SynCom research effort by accelerating the design, build, test, and learn cycle in the PMI SFA, and serve the growing community that leverages this powerful research platform.

*Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKPA39.*

**Improving Candidate Gene Discovery By Combining Multiple Genetic Mapping Datasets**

Fausto Rodriguez-Zapata[1], Nirwan Tandukar[1], Jung-Ying Tzeng[2], **Rubén Rellán-Álvarez*[1]** (rrellan@ncsu.edu)

[1]Department of Molecular and Structural Biochemistry, [2]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, 27607.

**Project Goals:**

1) Perform an environmental GWAS in a panel of ~2000 sorghum accessions that have already been genotyped and georeferenced using soil phosphorus availability as the phenotype

2) Identify selection signals at the gene and pathway level using natural populations that have adapted to low phosphorus and low temperature

3) Characterize the genetic architecture of glycerolipid content during the early stages of sorghum development using the Sorghum Association Panel.

4) Test if selection for faster development in low phosphorus availability conditions reduces genome size and more phosphorus efficient varieties.

5) Develop algorithms that incorporate all the different types of information we collect (i.e., metabolite levels, GWAS candidate genes, selection signals) to improve our ability to detect signals of minor effects and increase our confidence in selecting candidate genes for further validation.

6) Validate candidate gene(s) function under low phosphorus and cold conditions.

Deciphering the number and importance of loci involved in plant local adaptation and identifying candidate genes controlling traits that are relevant for adaptation to environmental stresses are overarching goals of plant scientists across a range of fields. From plant breeders aiming to introduce beneficial genetic diversity from landraces and wild relatives into high-yielding modern varieties to evolutionary biologists striving to identify which loci are shaped by evolutionary forces and molecular biologists and physiologists seeking to understand the role of genetic variation in plant development and stress responses.

However, validating gene function at the molecular and physiological level and confirming the importance of allelic variation for a particular gene are expensive and usually only possible when candidates can be limited to those with large effects on the trait of interest. Reducing the number of candidate genes and making an informed decision on which ones should be validated is particularly challenging in the case of highly polygenic traits.

In this project, we will leverage our *a priori* knowledge of the role of metabolic pathways in plant adaptation to environmental stresses to develop a combinatorial approach to combine data

from targeted metabolic profiling GWAS, environmental GWAS, and population genetics indexes of genetic selection. With these data, we will then design mathematical methods that incorporate pathway-level data and phenotype-genotype associations to develop strategies to reduce the number of candidate genes.

Our preliminary data and the literature show that phosphorus deficiency and low temperatures can have a profound impact on lipid metabolism, particularly on the levels of plant lipids, and that the combination of population and quantitative genetics together with high-precision metabolic phenotyping can be a powerful approach to identify relevant gene/pathway–phenotype associations and ultimately elucidate the functional role of genes/pathways in plant adaptation to environmental stresses.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER) Early Career grant no. DE-SC0021889.

**Nanoparticle-Mediated Transformation of Sorghum towards the Determination of a Subcellular Metabolic Network Map**

Christopher T. Jackson[1], Destiny J. Davis[2*](destinydavis@lbl.gov), Tallyta N. Silva[2], Suryatapa Ghosh Jha[3], Gozde S. Demirer[1,#], Jason B. Thomas[3], Christine Aquino[3], David W. Ehrhardt[3], **Seung Y. Rhee[3],** Jenny C. Mortimer[2], Markita P. Landry[1]

[1]University of California, Berkeley, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Carnegie Institution for Science, Stanford, CA
[#]Current address: University of California, Davis, CA

**Project Goals: The goal of the Sorghum Metabolic Atlas (SMA) project is to create an integrated pipeline to characterize metabolic interactions and pathways at a cellular level by mapping Sorghum enzymes using a variety of experimental approaches. This pipeline is divided into three stages: a) establishing Agrobacterium- and nanotechnology-mediated transient transformation of grasses to identify subcellular location of Sorghum enzymes through high-resolution confocal imaging; b) selecting enzymes to determine their subcellular localization; and c) using experimental data to generate new compartmentalized metabolic network models as well as refining existing pathway models. This project will create a repository for subcellular locations of metabolic enzymes, yielding important insight into the location and function of metabolic networks in Sorghum.**

Understanding plant metabolic networks is essential to enable the efficient engineering of resilient and sustainable bioenergy crops. Although model species such as *Arabidopsis thaliana* have extensive resources from which to draw, there remains a lack of information in species such as *Sorghum bicolor*. Sorghum is a challenging species to work with, as it has very poor transformation efficiencies. Here, we are implementing new transformation methods using carbon nanotubes (CNTs) which will allow us to rapidly test bioinformatic predictions of enzyme subcellular locations. Initial tests using vectors with fluorescent proteins (FP) under the control of *CaMV35S* and maize *Ubiquitin* promoters have shown transient expression of FPs in sorghum leaves, indicating successful carbon nanotube-mediated transformation. However, optimization is necessary due to inconsistencies between CNT batches, leaf infiltration challenges, and validation of subcellular localization.

We found that the polymers used to load vectors onto CNTs play a key role in construct durability, DNA loading capability, and plant toxicity. To minimize hydrolysis of the polymer from the CNT surface, we tested various polymers and storage conditions. Different cationic polymers adsorbed DNA with different electrostatic strengths, largely as a function of the polymer size and structure. To identify trends in polymer structure that minimize toxicity while maintaining DNA loading capability, we measured the upregulation of *PR1*, a marker for biotic

stress, using RT-qPCR from infiltrated *Nicotiana benthamiana* leaves. With these insights, we are actively exploring new chemistries for the loading of biomolecular cargo.

A significant bottleneck in sorghum leaf transformation is the efficient infiltration of plasmid-loaded CNTs into the silica-rich leaves, additionally hindered by a thick epidermal cuticle. To enhance CNT uptake, we tested several leaf abradement and infiltration approaches. Using a combination of leaf abradement and vacuum infiltration, we successfully transformed sorghum epidermal cells. We are currently optimizing this method for reproducibility and enhanced expression of introduced genes.

Following transformation, a key component of this project lies in validation of the subcellular localizations we see in sorghum. Our primary considerations in identifying a suitable test species is the presence of prior localization studies, leaf infiltration efficiency, and close relation to sorghum. To address these criteria, we are testing several well-studied monocot species with different leaf properties in parallel, including rice, maize, and Brachypodium. As an added layer of assurance, we are also exploring alternatives to CNT-mediated leaf transformation, including Agrobacterium-mediated transformation and sorghum callus transformation.

References:


- Jha SG, Borowsky AT, Cole BJ, Fahlgren N, Farmer A, Huang SSC, Karia P, Libault M, Provart NJ, Rice SL, et al (2021) Vision, challenges and opportunities for a plant cell atlas. Elife 10: 1–27. https://doi.org/10.7554/eLife.66877.
- Jackson CT, Wang JW, González-Grandío E, Goh NS, Mun J, Krishnan S, Geyer FL, Keller H, Ebert S, Molawi K, et al (2021) Polymer-Conjugated Carbon Nanotubes for Biomolecule Loading. ACS Nano. https://doi.org/10.1021/acsnano.1c06343.
- Silva TN, Thomas JB, Dahlberg J, Rhee SY, Mortimer JC (2021) Progress and challenges in sorghum biotechnology, a multi-purpose feedstock for the bioeconomy. Journal of Experimental Botany. https://doi.org/10.1093/jxb/erab450.
- Demirer GS, Silva TN, Jackson CT, Thomas JB, Ehrhardt D, Rhee SY, Mortimer JC, Landry, MP (2021) Nanotechnology to advance CRISPR/Cas genetic engineering of plants. Nature Nanotechnology. https://doi.org/10.1038/s41565-021-00854-y.



*This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0020366*

**Systematic Identification of Subcellular Location of Sorghum Metabolic Enzymes**

Purva Karia[1,*], William Dwyer[1], Olivia MacDonald[1], Daniel Ginzburg[1], Markita P. Landry[3], Jenny C. Mortimer[2], David W. Ehrhardt[1], **Seung Y. Rhee**[1]

[1]Carnegie Institution for Science, Stanford, CA, [2]Lawrence Berkeley National Laboratory, Berkeley, CA, [3]University of California, Berkeley, CA
*pkaria@carnegiescience.edu

**Project Goals: The goal of the Sorghum Metabolic Atlas (SMA) project is to create an integrated pipeline to characterize metabolic interactions and pathways at a cellular level by mapping Sorghum enzymes using a variety of experimental approaches. This pipeline is divided into three stages: a) establishing Agrobacterium- and nanotechnology-mediated transient transformation of grasses to identify subcellular location of Sorghum enzymes through high-resolution confocal imaging; b) selecting enzymes to determine their subcellular localization; and c) using experimental data to generate new compartmentalized metabolic network models as well as refining existing pathway models. This project will create a repository for subcellular locations of metabolic enzymes, yielding important insight into the location and function of metabolic networks in Sorghum.**

Plant metabolism underpins many traits that improve plant productivity. Decoding plant metabolic networks is crucial to meet the increasing demand of crop production. Although subcellular localization of enzymes is critical to understanding metabolic networks, localization of the majority of the enzymes is unknown. Fluorescently tagged enzyme localization information is available for many enzymes from the model species *Arabidopsis thaliana* but is largely lacking in most plants, including in the DOE flagship bioenergy plants such as *Sorghum bicolor*.

Due to the challenges associated with Sorghum transformation, we used two complementary monocot systems to determine the localization of Sorghum enzymes: 1) leaves of *Egeria densa*, an aquatic grass; and 2) roots of *Brachypodium distachyon*. Sorghum enzymes were heterologously expressed in both species via an Agrobacterium-mediated transformation method and the protein subcellular localization was identified through confocal microscopy.

In our pilot experiment, we examined the expression of a fluorescent reporter under a maize ubiquitin promoter in a modified pANIC5A and pGVG monocot vectors. The vector modification added a fluorescent tag up-/down-stream of a Gateway cassette into each vector. As a proof of concept, we selected 37 genes from 27 pathways to determine subcellular localization. To quantify the degree of subcellular localization conservation between Arabidopsis and sorghum, we selected 18 sorghum enzymes whose Arabidopsis orthologs have been

experimentally localized. In addition, we selected 6 sorghum-specific genes with no Arabidopsis orthologs and 8 sorghum genes with unlocalized Arabidopsis orthologs. Subsequent enzyme datasets were selected by focusing on pathways of biological interest to maximize coverage of the metabolic network and potential downstream studies to elucidate the function and regulation of the pathways.

The organelle markers were generated using the organelle localization signal sequence tagged with a fluorescent protein[1]. The markers containing the cytosol, nucleus, chloroplast, mitochondria, and peroxisomes localizing fluorescent proteins are currently being validated. Once confirmed, these constructs will be co-expressed with fluorescently tagged enzymes to determine the organelle localization of enzymes.

Overall, this project aims to improve location to function relationship of sorghum metabolic enzymes to facilitate better strategies to bioengineer pathways to tackle challenges related to global warming and food security.

References:


- Osaki, Y and Kodama, Y (2017) Particle bombardment and subcellular protein localization analysis in the aquatic plant Egeria densa. PeerJ 5 : e3779 https://doi.org/10.7717/peerj.3779

# Phage diversity and activity associated with seasonal changes in a model montane soil ecosystem

Clement Coclet[1], Patrick O. Sorensen[2], Ulas Karaoz[2], Eoin L. Brodie[2,3], Emiley A. Eloe-Fadrosh[1], **Simon Roux**[1]* (sroux@lbl.gov)

**Institutions:** [1]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA.

**Project Goals:** This project aims at characterizing the diversity and dynamics of phage communities in a montane meadow soil across a full year, from beneath the snowpack during winter, through snowmelt and plant growth in spring, and to plant senescence in the fall. Leveraging paired metagenomes and metatranscriptomes, a holistic view of the activity and turnover of the local viral community will be assembled to explore the potential roles and impacts of different viral groups on microbiome diversity and biogeochemical cycles across ecological regimes.

**Abstract Text:** Viruses are key elements of all microbiomes on Earth, shaping microbial community composition and structure while influencing host cell metabolism during infection. Recent viral ecology studies powered by 'omics approaches have provided a thorough description and investigation of the virosphere in many ecosystems, but have been challenging to apply to the incredibly diverse soil viruses, so that the exploration of global soil viral diversity is still only getting started[1]. In addition to establishing a global census of viruses across soil types and locations, one of the major outstanding questions in soil viral ecology is the characterization of soil virus dynamics across time, both in terms of diversity and activity. At the East River, CO, watershed, soil microbiomes respond to dramatic changes in subsurface conditions through snowfall, snowmelt, plant re-emergence, and plant senescence. The snowmelt period in particular results in a large crash of microbial biomass and shifts in community composition[2]. To characterize phage diversity, dynamics, and potential ecosystem impacts during this period we leverage 44 paired metagenomes and metatranscriptomes sampled across a whole year in this montane meadow ecosystem soil. A combined assembly approach enabled by the recently developed MetaHipMer tool[3] yielded >3,800 DNA phage populations whose abundance and activity could be followed through the entire year. Overall, these DNA phages were associated with the main bacterial groups present, around one third were reliably predicted as temperate, and most were consistently detected through the year albeit with low relative abundance. Transcriptional activity could be detected for both abundant and rare DNA phages, but consisted most often in the expression of one or two individual genes, and was thus more compatible with either self-regulation of temperate phages or host cell manipulation rather than active phage replication, since the latter would be associated with widespread expression of a large fraction of the phage genes. This suggests that a substantial portion of the DNA phage community in this ecosystem may reside in their host cell for prolonged period of time without actively replicating, but maintaining a basal level of expression for key phage-encoded genes.

Meanwhile, an unexpectedly abundant and diverse community of RNA phages could be detected in the same samples totaling >3,500 distinct populations, which in contrast to the DNA phages displayed a high turnover rate between time points, and included a substantial portion which were predicted as actively replicating (~10–30%). Taken together, the picture starting to emerge from these data is one of a relatively stable DNA phage community with sporadic activity and primary impact likely stemming from host cell manipulation during latent infections, alongside a more dynamic and active RNA phage community which may be responsible for a substantial fraction of the viral lysis in these microbiomes. More broadly, these observations of contrasted diversity and activity for DNA and RNA phages across seasons in a montane soil ecosystem highlights the need to expand our understanding of the eco-evolutionary drivers of phage diversity and activity in different soil types, especially for RNA phages, and demonstrates how multi 'omics approaches can provide a unique data framework to start such characterization at both local and global scales.

**References:**


1. Williamson, K. E., Fuhrmann, J. J., Wommack, K. E. & Radosevich, M. Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annu. Rev. Virol.* **4**, 201–219 (2017).
2. Sorensen, P. O. *et al.* The Snowmelt Niche Differentiates Three Microbial Life Strategies That Influence Soil Nitrogen Availability During and After Winter. *Front. Microbiol.* **11**, 1–18 (2020).
3. Hofmeyr, S. *et al.* Terabase-scale metagenome coassembly with MetaHipMer. *Sci. Rep.* **10**, 1–11 (2020).



**Funding Statement:** This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research, Early Career Research Program awarded under UC-DOE Prime Contract DE-AC02-05CH11231.

# Cross-kingdom Comparative Genomics of Aromatic Catabolic Pathways in Fungi and Bacteria

Teeratas Kijpornyongpan and **Davinia Salvachúa*** (davinia.salvachua@nrel.gov)

Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO

**The overall goal of this project is to test the hypothesis that white-rot fungi can simultaneously depolymerize lignin extracellularly and catabolize depolymerization products intracellularly as carbon and energy sources. The results from this project will lead to improved understanding of lignin utilization by white-rot fungi, and enable identification of promising fungal strains for lignin catabolism and valorization. As part of this effort, we have conducted a comparative genomic study to investigate a diversity of aromatic catabolic pathways in the fungal and bacterial kingdoms. Furthermore, we have combined genomic and phylogenetic approaches to decipher the evolution of certain aromatic catabolic enzymes and discover new classes of enzymes, which represents a foundation for future biochemical and molecular genetic studies.**

Although lignin is the second most abundant polymer in plant biomass, its upgrading remains a major hurdle in biorefining due to its recalcitrance and structural heterogeneity. To overcome this challenge, the concept of biological funneling has emerged as an approach to convert heterogeneous mixtures of lignin-derived monomeric aromatic compounds – from various lignin deconstruction processes – to single chemicals by using engineered microbes. For this purpose, knowledge in aromatic catabolic pathways is necessary to design microbes with (1) enhanced substrate diversity utilization, (2) conversion efficiency, and (3) tailored metabolic pathways to produce the desired products (Johnson et al., 2019). Aromatic catabolic pathways have been thoroughly described in several bacteria such as *Pseudomonas putida*, *Burkholderia* sp. SJ98, *Sphingobium* sp. SYK-6, and *Rhodococcus jostii*. However, knowledge of the corresponding aromatic catabolic pathways in white-rot fungi (WRF) is quite limited, even though WRF are known to be the most efficient lignin-degrading organisms in nature (del Cerro et al., 2021).

To reveal the distribution and diversity of aromatic catabolic pathways in WRF, we performed a large-scale comparative genomic and phylogenetic study across the bacterial and fungal kingdoms. First, we selected functionally characterized enzymes from bacteria and recently validated enzymes from WRF (Kijpornyongpan et al., manuscript in preparation) involved in (1) pre-cleavage modifications of aromatic compounds (i.e. oxidative and non-oxidative decarboxylases and aromatic hydroxylases), (2) ring-cleavage (i.e. intradiol and extradiol dioxygenases), and (3) post-cleavage modifications (i.e. oxidoreductases, 3-oxoacid CoA transferases, and thiolases). Second, we examined the protein domains in each enzyme selected for this study through InterProScan, and we selected protein domains with descriptions directly related to the catabolism of aromatic compounds as proxy domains to define a potential capability to modify and/or catabolize aromatic compounds. Third, we sampled 255 bacterial genomes and 317 fungal genomes, representing different lineages and nutritional modes across

the trees of life in these two kingdoms. For each genome, we performed genome-wide protein domain searches and gene homology assessment to identify proteins with proxy domains related to the catabolism of aromatic compounds. Apart from pathway and enzyme discovery, these analyses were also designed to determine if there are distribution patterns or associations between the abundance of aromatic catabolic enzymes and bacterial and fungal lineages and/or their nutritional modes. Finally, we performed phylogenetic analyses to understand the evolution of specific aromatic catabolic enzymes.

Based on the distribution of enzyme domains, we found that fungi have a higher conservation of genes that encode for aromatic catabolic enzymes – in particular intradiol dioxygenases, phenol hydroxylases, and 3-oxoacid CoA transferases – compared to bacteria. Our association analyses indicated that microbes that are either known to utilize aromatic compounds as a carbon source or classified as plant-associated also have a higher abundance of aromatic catabolic enzymes compared to other microbes found in environment or known as animal-associated. In addition, our findings highlighted the evolution of intradiol dioxygenases: specifically, we found separate origins of catechol dioxygenases and 3,4-protocatechuate dioxygenases in bacteria and fungi. The gene tree phylogeny suggested that hydroxyquinol dioxygenases are the most common type of intradiol dioxygenases found in fungi. We also identified a novel class of intradiol dioxygenases, which likely function extracellularly based on the presence of signal peptides. Taken together, this work provides new insights related to the catabolism of aromatic compounds in different groups of bacteria and fungi which could enable the discovery of novel microbes, pathways, and enzymes to improve the 'funnel' to convert lignin-derived aromatic compounds to value-added products.

**References**

del Cerro, C., Erickson, E., Dong, T., Wong, A.R., Eder, E.K., Purvine, S.O., Mitchell, H.D., Weitz, K.K., Markillie, L.M., Burnet, M.C., Hoyt, D.W., Chu, R. K., Cheng, J.F., Ramirez, K.J., Katahira, R., Xiong, W., Himmel, M.E., Subramanian, V., Linger, J.G., Salvachúa, D. (2021). Intracellular pathways for lignin catabolism in white-rot fungi. PNAS, 118, e2017381118.

Johnson, C.W., Salvachúa, D., Rorrer, N.A., Black, B.A., Vardon, D.R., St. John, P.C., Cleveland, N.S., Dominick, G., Elmore, J.R., Grundl, N., Khanna, P., Martinez, C.R., Michener, W.E., Peterson, D.J., Ramirez, K.J., Singh, P., VanderWall, T.A., Wilson, N., Yi, X., Biddy, M.J., Bomble, Y.J., Guss, A.M., Beckham, G.T. (2019). Innovative chemicals and materials from bacterial aromatic catabolic pathways. Joule, 3, 1523-1537.

*This research is supported by the U.S. Department of Energy (DOE) Office of Science, Office of Biological and Environmental Research under the Early Career Award Program. This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308.*

# Data driven analysis of nitrogen deficit responses across a sorghum diversity panel using hyperspectral leaf reflectance

**Michael C. Tross[1,2,+], Aime V. Nishimwe[1,2,+], Guangchao Sun[1,2], Yufeng Ge2,3, and James C. Schnable[1,2*]**

[1] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, 68588 USA
[2] Center for Plant Science Innovation & Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, 68588 USA
[3] Department of Biological Systems Engineering University of Nebraska-Lincoln, Lincoln, NE, 68588 USA

[+]Authors contributed equally
[*]schnable@unl.edu

Hyperspectral reflectance data can be collected at relatively high throughput from large numbers of individual plants in controlled environments or field conditions. A growing field of research focuses on employing hyperspectral reflectance data collected from plant leaves to estimate a range of plant traits which would otherwise require slower, more labor intensive and/or more expensive protocols to directly quantify. The general procedure adopted for estimating plant traits from hyperspectral reflectance data is to collect hyperspectral reflectance data from a large set of samples and, for a subset of these samples, collect labor intensive and/or expensive trait data using ground truth protocols. Here we evaluate the potential of data driven approaches to analyzing hyperspectral reflectance data and quantify both genetic variability and genotype by environment variability in the sorghum association panel, a population representing global sorghum genetic and phenotypic diversity, across multiple environments. An ASD spectrometer was employed to quantify leaf reflectance (intensity values for 2,150 discrete wavelengths) from 1,143 plants. An autoencoder comprised of both an encoder network having an input layer, six dense layers and a sampling layer and a decoder network comprising of an input layer and 7 dense layers was also able to summarize variation within the dataset at a validation loss of 0.05 (loss = KL-divergence ($\mu$,$\sigma$ ) + reconstruction loss). Multiple principal components and latent variables exhibited significant heritability – i.e. the proportion of total variance explained by differences between plant genotypes – under stressed and non-stressed conditions. Autoencoding based dimensional reduction provides the opportunity to link additional sorghum genes to phenotypic variation via GWAS, mashr, and/or GPWAS.

# The impact of LIP36 seed-specific expression on seed and oil yields in *Camelina sativa* and its associated transcriptome and metabolome changes

*Hesham M. Abdullah[1] (abdull76@msu.edu), Bibin Paulose [1], Emily Pawlowski [1], Erich Grotewold[2], and **Danny J. Schnell**[1]

[1]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA
[2]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

URL: https://groecamelina.natsci.msu.edu/

## Project Goals:

Our research aims to sustainably increase oilseed yields in the non-food oilseed crop plant, *Camelina sativa*, thereby making it a commercially viable alternative for biofuels and bioproducts production. Camelina has shown considerable promise as a dedicated industrial oilseed crop because it requires low agronomic inputs, is naturally more resistant to both biotic and abiotic stress than other oilseed crops, and Camelina oil-based blends have been tested and approved as liquid transportation fuels[1,2]. This project was proposed to increase fixed carbon allocation by identifying metabolic bottlenecks and engineering metabolic pathways that limit or co-limit seed and oil production in Camelina. In the current project, we designed highly innovative metabolic engineering strategies to 1) increase carbon capture in photosynthetic source tissues and 2) redirect the carbon to seeds for seed oil production. We further investigated the roles of mitochondrial metabolism during seed development, seed germination, and seedling growth, through using transgenic plants.

## Abstract

A member of the mitochondrial carrier protein family and a component of algal carbon concentrating mechanisms (CCMs) in *Chlamydomonas reinhardtii* (Chlamy), designated, low-$CO_2$ inducible protein (LIP36) was identified previously in our laboratory. LIP36 is localized to mitochondria in Camelina, and its suppression by RNA interference (RNAi) suggested its essential role in growth of Chlamy under low $CO_2$ conditions[3]. To gain insight into the role of LIP36 proteins, we constitutively expressed LIP36 into the oilseed crop *Camelina sativa*. LIP36 expression has resulted in higher photosynthetic $CO_2$ assimilation (30-46%), under limiting conditions, relative to the wildtype (WT) controls. LIP36 lines showed a more favorable redox status, higher water and nitrogen use efficiency, and increased seed and oil yields. The $^{13}C$ labeling experiments in Camelina leaves have also suggested that LIP36 improves the

capacity of anaplerotic pathways, thereby maintaining plant productivity under non-ideal growth conditions. Based on these preliminary results, we aim to confirm the impact of LIP36 on the flux through central carbon metabolism in both source and sink (seed) tissues through expressing it under seed-specific oleosin promoter, and to emphasize whether its increased activity in seeds could lead to enhanced carbon assimilation, and therefore further increased seed and oil yields. The transformation of Oleosin :: LIP36 construct into Camelina was successful and homozygous T4 seed lines were generated and screened/evaluated for seed attributes (seed size, seed weight, and % oil content). The elite lines expressing LIP36 were also subjected to RNA-Seq analysis to investigate the global changes in Camelina transcriptome in response to LIP36 expression in seeds, and to identify candidate genes/enzymes for further improving seed biomass and yield traits in Camelina. The steady-state metabolite profiling and the metabolic flux analysis (MFA) is currently ongoing in order to monitor the global metabolic changes in LIP36 transgenics in order to define the impact of LIP36 on the flux through central carbon metabolism, and to further integrate LIP36 with other yield traits to enhance water use efficiency and improve $CO_2$ assimilation under limiting environmental conditions.

## References

1. United States Department of Agriculture, F.A.S. https://apps.fas.usda.gov/psdonline/app/index.html#/app/downloads (2017).

2. Rosillo-Calle, F., Pelkmans, L. & Walter, A. A global overview of vegetable oils, with reference to biodiesel. A Report for the IEA Bioenergy Task 40 (2009).

3. Pollock SV, Prout DL, Godfrey AC, Lemaire SD, Moroney JV. The Chlamydomonas reinhardtii proteins Ccp1 and Ccp2 are required for long-term growth, but are not necessary for efficient photosynthesis, in a low-CO2 environment. Plant Mol Biol. 2004 Sep;56(1):125-32. doi: 10.1007/s11103-004-2650-4. PMID: 15604732.

**Acknowledgment:** This research is funded by the Department of Energy (ARPAe DE-0000200 & BER DE-SC0018269)

# Transcriptional regulation of lipid metabolism in *Camelina sativa*

Fabio Gomez-Cano[1], Allison M. Hurlock[1], Yi-Hsuan Chu[1], Mariel Cruz-Gomez[1], Suresh Kumar Gupta[1*] (guptasu6@msu.edu), Hesham M. Abdullah[2], Emily Pawlowski[2], Danny Schnell[2], **Erich Grotewold[1]**
[1]Department of Biochemistry and Molecular Biology, and [2]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA.

https://camregbase.org/

**Project Goals**

The main goals of this project element of this large project is to define the transcriptional networks that control tissue-specific and whole plant carbon allocation in Camelina (this correspond to Project Element 3 of the entire project) and the specific goals are:

a) Develop a High-Resolution Transcriptome for Camelina Seed Development
b) Characterize and Identify Targets of Camelina Seed Oil TFs
c) Manipulate TFs and Metabolic Genes for Pathway Discovery and Seed Oil Metabolic Engineering

Camelina is an annual oilseed plant that is gaining momentum as a biofuel winter cover crop. Understanding gene regulatory networks (GRN) is essential in deciphering plant metabolic pathways, such as lipid metabolism. In camelina, a collection of gene expression datasets enabled characterizing the expression of >4,600 TFs in >130 RNA-seq samples[1]. We collected lipid-related genes (LRGs) and expression data from CamRegBase[1]. The LRGs were classified according to the information provided by AraLip[2], and the expression data was used to identify genes highly co-expressed with TF genes. By combining these two analyses, we identified 350 TFs highly co-expressed with LRGs. We ranked these TFs based on their expression in seeds and on the number of LRGs co-expressed and selected the top 22 for further validations after discarding low expressed paralogous genes. Potential target genes for these Camelina TFs potentially involved in the control of lipid metabolism were further studied by DNA affinity purification sequencing (DAP-seq). Enrichment analyses of targets supported the co-expression predictions and predicted TFs associated with seed fatty acid elongation, and synthesis and degradation of triacylglycerols (TAGs). We expressed these 22 TFs in Camelina transgenic plants driven by the soybean glycinin promoter, and the analysis of oil by NMR of $T_2$ seeds provided some preliminary evidence of effects on fatty acid accumulation. The results to be presented highlight the importance of the integration of multiple layers of information to guide predictive biology toward the discovery of TFs regulators of metabolic processes.

**References**

1. Gomez-Cano, F., Carey, L., Lucas, K., García Navarrete, T., Mukundi, E., Lundback, S., Schnell, D., and Grotewold, E. (2020). CamRegBase: A gene regulation database for the biofuel crop, *Camelina sativa*. Database 2020.
2. Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D., Baud, S., Bird, D., DeBono, A., and Durrett, T.P. (2013). Acyl-lipid metabolism. The Arabidopsis Book/American Society of Plant Biologists 11.

*This research was funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0018269 to E.G. and D.S.*

# High-Throughput Detection of T-DNA Insertion Sites for Multiple Transgenes in Complex Genomes

Brianne Edwards,[1] Heike Sederoff,[1*](hwsedero@ncsu.edu) Nathan Wilson,[1] Eli Hornstein,[1] **Danny Schnell** [2]

[1]North Carolina State University, Dept Plant and Microbial Biology, Raleigh, NC
[2]Michigan State University, Department of Plant Biology, East Lansing, MI

**Website URL:** https://groecamelina.natsci.msu.edu/

**Project Goals:** Integrating multiple transgenes into elite lines can be a time consuming and labor-intensive process. Generation of stable homozygous lines often requires multiple generations and screening of large numbers of plants especially in polyploid species. Our goal was to develop a fast, high-throughput method to identify transgene insertion sites in the T1 generation that accelerates identification of stable lines with the desired numbers of transgene copies.

**Abstract:** Genetic engineering of crop plants has been successful in transferring traits into elite lines that could not be achieved with breeding techniques. Introduction of transgenes originating from other species have conferred resistance to biotic and abiotic stresses, increased efficiencies and modified developmental programs. Many of these traits focus on herbicide and insecticide resistance and are widely used as single gene traits while genetic engineering of new pathways often require multiple genes and regulators. In recent years, the development of resistance to pesticides in wild species has driven companies to combine transgenes for different pesticide resistances into crops via gene stacking. However, generating stable homozygous lines with multiple transgenes requires selection over several segregating generations and therefore is time consuming and labor intensive, especially if the crop is polyploid. Insertion site effects and transgene copy numbers are important metrics for commercialization and trait efficiency.

We have developed a method based on a genome-walking PCR approach and demonstrate how it can be easily adapted for high-throughput screening of multiple lines and transgenes at a time using a short-read sequencing platform. Sites identified via HT-sequencing could then be used to design screening primers to test for zygosity of subsequent generations at every locus and also enable segregation of irregular or unidentified insertions. The overall process of identification of transgene insertion sites is accessible, as it involves commonly used laboratory techniques.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018269.

# A mitochondrial carrier protein from *Chlamydomonas* alter the root architecture in Camelina.

Rajneesh Singhal[1]*(singha24@msu.edu), Bibin Paulose[1] Erich Grotewold[2], and **Danny Schnell**[1]

[1]Department of Plant Biology, Michigan State University, East Lansing-48824-MI.
2Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing-48824-MI

URL: https://groecamelina.natsci.msu.edu/

## Project goals:

Our research aims to sustainably increase oilseed yield in the non-food oil seed crop plants, *Camelina sativa*, thereby making it a commercially viable alternative for biofuels and bioproducts production. Camelina has been has shown considerable promise as a dedicated industrial oilseed crop because it requires low agronomic inputs, is naturally more resistant to both biotic and abiotic stress than other oilseed crops, and Camelina oil-based blends have been tested and approved as liquid transportation fuels[1,2]. The limitation in commercial adoption of camelina as an oilseed crop is its modest oil yield than its relative *Brassica napus*. The overall goal of this project is to boost the seed oil production of Camelina while retaining its low-input and stress resistance advantage. In this project, we seek to identify, optimize and combine traits that contributes to the flow of carbon from source to sink. Here we describe the impact of expression of LIP36 gene ( Low $CO_2$ induced protein) on roots, which are the main sink organs during early stages of plant development.

## Abstract:

The expression of *LIP36* (Low $CO_2$ induced protein) gene, a mitochondrial carrier protein from *Chlamydomonas reinhardtii* carbon concentrating mechanism, under constitutive promoter (35S) in Camelina has resulted in 40-70% increase in seed oil yield per hectare relative to control plants in field trials. *LIP36* is proposed to act as dicarboxylate transporter in mitochondria thereby promoting the cellular redox balance, photorespiratory flux and anapleurotic metabolism. This increases the flux through metabolic pathways that facilitates the fixation of $CO_2$ released during respiration and photorespiration. The expression of this gene promotes the non-cyclic TCA pathway and provides the carbon skeleton for N-fixation. One of the project goals was to

identify the positive traits that can result in better seedling establishment thereby contributing towards plant productivity. In the current work we further characterize the phenotype of the *LIP36* expressing Camelina plants. During early stages of development, the transgenic camelina plants has longer root length and higher rate of root elongation as compared to that of WT plants. The metabolite analysis on roots showed that the overexpressing lines has lower sucrose levels and alteration in mitochondrial metabolism. Genes related energy metabolism such as those involved in TCA cycle, oxidative phosphorylation and glucose metabolism are differentially regulated suggesting that the energy requirements in the root mitochondria have been altered. Thus, during early stages of the plant growth, the roots may be the active sink organs in the plants expressing LIP36 gene.

## References


1. United States Department of Agriculture, F.A.S. https://apps.fas.usda.gov/psdonline/app/index.html#/app/downloads (2017).
2. Rosillo-Calle, F., Pelkmans, L. & Walter, A. A global overview of vegetable oils, with reference to biodiesel. A report for the IEA Bioenergy Task 40 (2009)


## Funding Statement:


This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DOE-DE-SC0018269.

**Title:** Expanding the Utility of Integrases for Genome Editing and Stabilizing Gene Modules in Target Bacteria

**Authors:** Lauren G. Clark[2] (lauren.clark@northwestern.edu), Jesse Cahill[1], Catherine M. Mageeney[1], Grant Rybnicky[2], Michael Jewett[2], Joseph Schoeniger[1], **Kelly P. Williams[1]**

**Institutions:** [1] Sandia National Laboratories, Livermore, CA; [2] Northwestern University, Evanston, IL

**Project Goals:**

The overall goal for the Intrinsic Control for Genome and Transcriptome Editing in Communities (InCoGenTEC) project funded under the Secure Biosystems Design initiative is to expand our mechanistic and practical understanding of horizontal gene transfer mechanisms in bacterial communities, and to harness mobile elements to create and deliver constructs to transform, control and detect the genetic and biochemical state of bacteria. Improved ability to engineer genomes of both isolatable and non-isolatable species will enable better scientific understanding of bacterial communities and facilitate biotechnology applications that promote the growth of the bioeconomy. However, biocontainment and biosafety concerns must be addressed. We are creating modular synthetic genetic elements (SGEs) that can sense, and control altered states of microorganisms. These synthetic genetic elements can be used to transform community members *in situ* provided selective delivery and transformation mechanisms if delivery vectors can be created in an agile manner for new target strains. We have developed software that can precisely identify genomic islands in bacteria, which has yielded both thousands of unique DNA integrase insertion sites, and hundreds of thousands of prophages These phages can be recovered and used to deliver modular SGEs to target species within a community, even non-isolable members.

Integrases are key enzymes that can be used in phage-based or conjugative delivery systems for the site-specific insertion of delivered SGE modules into target microbe genomes, or for excising SGEs. Our goals for this project include developing a better understanding of integrase biology and building a broad set of tools that can be used for genome manipulation in a wide range of microbial species including many that are challenging to modify. We will accomplish this through the more immediate goal of conducting a high throughput screen of integrases and their att sites identified by the TIGER and Islander genome mining tools developed at Sandia National Labs. Next, a comprehensive mutational analysis of the natural attB sites will enable targeting of variant attB sites in target organism genomes in addition to providing a syntax of attB sites for the two families of integrase proteins. These goals will enable integrases, delivered via in vitro assembled phage, to be used in the larger framework of the project to edit the genomes of specific targeted non-model bacterial organisms within mixed bacterial communities. They will also inform the degree to which integrase sequence specificity requirements exclude or permit horizontal gene transfer between bacterial genomes.

**Abstract Text:** Integrases catalyze the splicing of large multi-gene DNAs (genomic islands) into specific sites (attBs) in chromosomes and have thus proven useful in genome editing. Expanding the set of attBs that integrases target will further increase their utility. We have applied our

software[1,2] for finding genomic islands to hundreds of thousands of bacterial genomes, while speeding the software with species-tailored databases and extending its reach to cross-contig islands. Because mapping is precise, each island links an integrase enzyme to the chromosomal site (attB) that it targets. We have thus developed a very large database of integrase/attB pairs. The main organizing principle to apply is a phylogenetic tree of the integrases, which will help us enumerate the number of truly different site-specificities in our database. The tree is challenging because it is large and a typical aid to its assessment - the bacterial tree of life - is not applicable, since integrase genes are on mobile DNAs that effectively jump from branch to branch on the bacterial tree. We have chosen an initial set of integrases to study biochemically, based on their occurrence among the study organisms of our project, and with a balance between the two main integrase protein families. Both in vitro (transcription/translation of integrase genes in cell-free extracts) and in vivo approaches are being explored to demonstrate integrase activity and application, with the goal of characterizing a large number of candidate integrase/attB pairs and conducting a comprehensive mutational analysis on a selected subset of the natural attB sites, to enable the targeting of much larger numbers of variant attB sites present in target organism genomes.

**References/Publications**


1. C. M. Mageeney *et al.*, New candidates for regulated gene integrity revealed through precise mapping of integrative genetic elements. *Nucleic Acids Res* **48**, 4052-4065 (2020).
2. C. M. Hudson, B. Y. Lau, K. P. Williams, Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res* **43**, D48-53 (2015).



**Funding Statement:** *This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under the Secure Biosystems Design Initiative. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND number: SAND2022-0188 A*

**Repurposing Cas13 for Precise Translational Inhibition and Activation**

Brady F. Cress[1,2*] (Cress@Berkeley.EDU), Peter B. Otoupal[3,4,5]* (PeterOtoupal@lbl.GOV), Muntathar Al-Shimary,[1,2*] Kate Miller,[1,2] Emeric J. Charles,[1,2] David F. Savage,[1,2] Jennifer A. Doudna,[1,2,6,7,8,9,10] and **Joseph S. Schoeniger**[11] (jsschoe@sandia.GOV)

[1]Innovative Genomics Institute, University of California, Berkeley; [2]Department of Molecular and Cell Biology, University of California, Berkeley; [3]Biomanufacturing and Biomaterials Department, Sandia National Laboratories, Livermore, CA; [4]Deconstruction Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA; [5]Agile BioFoundry, Department of Energy, Emeryville, CA; [6]California Institute for Quantitative Biosciences, University of California, Berkeley; [7]Department of Chemistry, University of California, Berkeley; [8]Howard Hughes Medical Institute, University of California, Berkeley; [9]Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [10]Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, [11]Systems Biology Department, Sandia National Laboratories, Livermore, CA

https://genomicscience.energy.gov/research/sfas/snlsecurebiodesign.shtml

**Project Goals:** The Intrinsic Control for Genome and Transcriptome Editing in Communities (InCoGenTEC) project (funded under the Secure Biosystems Design initiative) seeks to develop technologies that foster the growth of our burgeoning bioeconomy, while maximizing biosafety, biosecurity, and biocontainment. We aim to expand our mechanistic and understanding of horizontal gene transfer within bacterial communities, and to harness synthetic mobile elements (SGEs) to create and deliver agile tools for transforming, controlling and detecting the genetic and biochemical state of bacteria. By improving our ability to engineer specific species (even non-isolatable ones) within bacterial communities in situ, we will enable better scientific understanding of microbial consortia and facilitate biotechnology applications that promote the growth of the bioeconomy while maintaining a paramount commitment to biosafety. We are harnessing the modular and programmable nature of CRISPR systems to create SGEs that enable more advanced manipulation of bacterial genomes, particularly in how they are expressed. Our team will then employ software to identify species-specific bacteriophages from near-neighbors and exploit them to deliver these modular SGEs into target species within a community.

**Abstract:**

Tools for synthetically controlling gene expression levels are a fundamental cornerstone of genetic engineering endeavors. While CRISPRi and CRISPRa technologies have been applied extensively towards control over transcription, comparatively little has been done in extending this control towards the translational level. Here, we employ CRISPR to modulate the rate of mRNA translation.

In the model prokaryote *E. coli*, approximately half of its genes are expressed in operons, where their transcription levels are tied to the surrounding genes encoded on the same mRNA. For this reason, state-of-the-art synthetic regulators like dCas9 and dCas12 are unable to selectively knock down a gene's expression without repressing genes downstream in the operon, prohibiting independent regulation of a large fraction of prokaryotic genes. We hypothesized that Cas13, an RNA-guided, RNA-targeting protein, might selectively block translation of a single gene on an mRNA molecule without impacting the expression of its operonic neighbors. To this end, our labs recently demonstrated that targeting catalytically inactive Cas13d (dCasRx) to the ribosome binding site of a fluorescent protein-encoding mRNA efficiently inhibits its translation in *E. coli*. We show here that dCasRx excels at targeted knockdown of genes in operons, selectively repressing individual fluorescent proteins in a three-gene synthetic reporter operon. We have also developed a genome-wide dCasRx guide RNA library in *E. coli*, consisting of ~150k pooled guides, to systematically determine design rules for efficient inhibition of translation and of non-coding RNA function. dCasRx represents the first application of a programmable RNA-binding protein for precise microbial gene regulation, holding great promise for functional genomics and synthetic gene regulation.

In tandem, we also demonstrate the application of Cas13 for enhancing translation rates. By targeting dCasRx to the start of the 5' untranslated region of RFP mRNA in *E. coli*, we demonstrate a 6.6-fold enhancement of gene expression. We further develop Cas13 as a translational activator by creating a novel variant dCasRx linked to the translation initiation factor IF3. We demonstrate dCasRx-IF3's ability to further enhance gene expression 16.0-fold above basal levels. Activation of translation is location-dependent, and we show dCasRx-IF3 maintains the ability to repress translation 7.4-fold when targeted to the RBS. Tuning experimental parameters resulted in 15.5-fold activation of RFP expression between dCasRx-IF3 and dCasRx strains, providing an insight into how dCasRx-IF3 can be further engineered and applied to enhance translation rates. We also demonstrate targeted upregulation of native β-galactosidase gene LacZ, highlighting dCasRx-IF3's versatility in application. Furthermore, we show that dCasRx-IF3 exhibits minimal fitness impacts and requires no additional host modification to influence gene expression.

Taken as a whole, this work outlines novel approaches for applying CRISPR for post-transcriptional control of mRNA translation rates to both inhibit and activate gene expression.

**Funding Statement:** This material is based upon work supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under the Secure Biosystems Design Initiative. Sandia National Laboratories is managed by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy under contract no. DE-NA-0003525.

**Title: Phage Factory: Creating a phage for any bacterial species**

**Authors:** Catherine M. Mageeney[1#*] (cmmagee@sandia.gov), Grant Rybnicky[2#], Michael Jewett[2], Kelly P. Williams[1], **Joseph Schoeniger[1]**

**Institutions:** [1] Sandia National Laboratories, Livermore, CA; [2] Northwestern University, Evanston, IL [#]equal contribution

**Project Goals:** The overall goal for the Intrinsic Control for Genome and Transcriptome Editing in Communities (InCoGenTEC) project funded under the Secure Biosystems Design initiative is to expand our mechanistic and practical understanding of horizontal gene transfer mechanisms in bacterial communities, and to harness mobile elements to create and deliver constructs to transform, control and detect the genetic and biochemical state of bacteria. Improved ability to engineer genomes of both isolatable and non-isolatable species will enable better scientific understanding of bacterial communities, and facilitate biotechnology applications that promote the growth of the bioeconomy. However, biocontainment and biosafety concerns must be addressed. We are creating modular synthetic genetic elements (SGEs) that can sense, and control altered states of microorganisms. These synthetic genetic elements can be used to transform community members *in situ* provided selective delivery and transformation mechanisms if delivery vectors can be created in an agile manner for new target strains. We have developed software that can identify numerous bacteriophages in near-neighbors of a target strain. These phages can be recovered and used to deliver modular SGEs to target species within a community, even non-isolable members.

**Abstract Text:**

In natural settings bacteriophages (phages) are extremely diverse, efficient shuttles of genetic information, and can infect all groups of bacteria currently discovered. We have developed a Phage Factory platform to computationally predict prophages precisely (phages integrated into bacterial genomes) and experimentally turn predicted phage sequences into productive virions capable of killing their host, or into phage-vectors, capable of cargo delivery.

We have computationally predicted 185,535 prophages in 207,083 bacterial and archaeal genomes obtained from NCBI, using our TIGER [1] and Islander [2] software. This creates the largest database of phage sequences to date, and at least one host of each phage is known. We have selected to initially focus on *Pseudomonas, Streptomyces*, and *Rhodococcus* species to produce phage using our Phage Factory. Our Phage Factory begins by mining our prophage database for prophages in close relatives of our target strain and annotating these prophage genomes. We experimentally verify the prophages through PCR-based methods and a computational method Juxtaposer [3]. We then create synthetic phage genomes through methods such as long PCR and Gibson assembly, but a key challenge is rebooting, or reconstituting the phage genomes into infectious phage particles [4]. Rebooting has been accomplished through four main methods: transformation, conjugation, L-form transformation, and cell-free protein synthesis. We will present results from initial work on applying the Phage Factory to *Pseudomonas,* a genome with many targets relevant to biomanufacturing or the rhizosphere, and results on developing cell-free phage rebooting methods.

We have identified 14 prophages in six *P. putida* strains of interest, with two clusters and 10 singletons. We verified 9 of these 14 prophages to be active after mitomycin C induction; however, no phage plaques or clearing were visible on any strain tested, indicating these are non-permissive hosts for these 9 active prophages. We further analyzed the genomic relationships

using fastANI to measure the average nucleotide identity (ANI) between our six strains. This revealed these six strains are phylogenetically diverse, with the closest sharing only 98% ANI. This indicated the genetic similarity for permissive prophage hosts is more than 98%.

We are now in the process of creating synthetic genomes for prophages harbored in close relatives of *P. putida* S12. We have two strategies to obtain synthetic phage genomes: 1) obtain the strain of the close relative, long PCR prophage fragments and Gibson Assemble, or 2) Assemble gBlocks for the prophage of interest and Gibson Assemble the fragments.

In order to convert these natural or synthetic phage genomes into infectious virions, we are pursuing cell-free rebooting techniques [5], which we will compare to established approaches such as electroporation. Cell-free protein synthesis uses cell extracts to perform transcription and translation *in vitro*, which can be directly manipulated to optimize rebooting viral genomes. Through screening a variety of biochemical parameters, we have found that adding crowding agents to the reaction and inhibiting exonuclease activity in the extract is essential to producing viable phage particles by cell-free protein synthesis. Likewise, the concentration of phage genome in a cell-free rebooting reaction positively impacts the final titer of phage particles produced beyond the stoichiometric amount that would be expected.

Using this biochemically optimized cell-free protein synthesis reaction, we intend to investigate the host range over which *E. coli* cell-free extracts can reboot phages as well as elucidate the phage-encoded determinants of cell-free rebooting. In determining extract host range, we will apply cell-free rebooting to multiple dsDNA phages per host organism in a phylogenetic walk-out strategy. Starting with *Pseudomonas*, we will evaluate the ability of phages to reboot in cell-free protein synthesis reactions and move further from *E. coli*, eventually to select *Actinobacteria* and *Cyanobacteria*. In parallel, we will test whether doping unoptimized cell-free extracts from the host organism is a viable strategy to extend the host range of *E. coli*-based cell-free rebooting reactions. To determine characteristics of phages that make them amenable to cell-free rebooting, we will assay a wide variety of *Pseudomonas aeruginosa* phages for their ability to reboot in cell-free reactions. Using *P. aeruginosa* phages gives sufficient diversity in genome molecule, genome length, particle morphology, and gene content such that we should be able to draw generalizable patterns from the dataset. Likewise, *P. aeruginosa* phages have been demonstrated to reboot in *E. coli* through electroporation previously. Through these investigations, we will inform the limits of cell-free phage rebooting and anticipate extending the range of bacteria that the Phage Factory can produce active phage against.

**References/Publications**


[1] C. M. Mageeney, B. Y. Lau, J. M. Wagner, C. M. Hudson, J. S. Schoeniger, R. Krishnakumar, and K. P. Williams, Nucleic Acids Res **48**, 4052 (2020).

[2] C. M. Hudson, B. Y. Lau, and K. P. Williams, Nucleic Acids Res **43**, D48 (2015).

[3] J. S. Schoeniger, C. M. Hudson, Z. W. Bent, A. Sinha, and K. P. Williams, Nucleic Acids Res **44**, 6830 (2016).

[4] C. M. Mageeney, A. Sinha, R. A. Mosesso, D. L. Medlin, B. Y. Lau, A. B. Rokes, T. W. Lane, S. S. Branda, and K. P. Williams, mSystems **5** (2020).

[5] M. Rustad, A. Eastlund, P. Jardine, and V. Noireaux, Synth Biol (Oxf) **3**, ysy002 (2018).



**Funding Statement:** *This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under the Secure Biosystems Design Initiative. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND number: SAND2022-0152 A.*

**Design, delivery and expression of synthetic genetic elements in diverse microorganisms**

**Jaymin Patel, Laura Quinto, Shenqi Wang, Katie Mageeney, Joe Schoeniger, Jason Crawford, Farren Isaacs**

The overall goal for the Intrinsic Control for Genome and Transcriptome Editing in Communities (InCoGenTEC) project funded under the Secure Biosystems Design initiative is to expand our mechanistic and practical understanding of horizontal gene transfer mechanisms in bacterial communities, and to harness mobile elements to create and deliver constructs to transform, control and detect the genetic and biochemical state of bacteria. Improved ability to engineer genomes of both isolatable and non-isolatable species will enable better scientific understanding of bacterial communities, and facilitate biotechnology applications that promote the growth of the bioeconomy.

A major challenge is in mobilizing, stabilizing and selectively activating functional genetic programs such as biosynthetic pathways in undomesticated environmental strains or into specific organisms *in situ* in intact microbiomes. Here, we describe an integrated computational—experimental technology to decouple biosynthetic capacity from host-range constraints that enabled pathway-targeted metabolite analyses in a diverse set of prokaryotic and eukaryotic hosts to harness their innate metabolic adaptability and activate silenced pathways. By placing these pathways under transcriptional control mechanisms orthogonal to the bacterial host, we can standardize design, maximize reuse of components, minimize unexpected behavior, and ensure that genetic subfragments of our heterologous synthetic pathways are non-functional, thereby improving both versatility of use and biocontainment.

Specifically, we developed a computational algorithm to redesign genes and their regulatory regions to adopt hybrid elements for cross-species expression of synthetic genetic elements (SGEs) in gram- and gram+ bacteria and eukaryotes. These algorithms have been converted to a web interface to enable impact in the broader scientific community. To transfer SGEs into diverse hosts, we also developed a mobilization strategy by engineering conjugation, transposition, and site-specific recombination, establishing feasibility on a validated violacein pathway. Further, we are developing redesigned pathways for intrinsic bioluminescence, which can enable facile labeling of individual strains within a complex community. We are currently expanding the capabilities of SGE mobilization to broaden host range, enable intracellular and environmental biosensing, introduce mechanisms for strain targeting, and multiplexing of genetic cargo. Jointly with other parts of the InCoGenTEC project, we are also exploring the incorporation of our SGEs into transforming bacteriophage vectors as an alternative, strain-specific delivery and transformation mechanism. These technologies establish a general strategy for highly-controlled expression and mobilization of genetic elements in diverse organisms and communities.

**The Impact of Waterlogging on Pennycress Morphology and Yield and Modeling Effective Water Availability of Pennycress Natural Populations**

Rachel Combs-Giroir[1*] (combs-giroir.1@osu.edu), John Lagergren[2*] (lagergrenjh@ornl.gov), Jared Streich[2], Daniel Jacobson[2], Andrea R. Gschwend[1], and **John Sedbrook**[3]

[1]The Ohio State University, Columbus, OH; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]Illinois State University, Normal, IL

https://www.pennycressresilience.org/

**Project Goals: This project employs evolutionary and computational genomic approaches to identify key genetic variants that have enabled *Thlaspi arvense* L. (Field Pennycress; pennycress) to locally adapt and colonize all temperate regions of the world. This, in combination with knowledge of metabolic and cellular networks derived from first principles, is guiding precise laboratory efforts to create and select high-resilience lines, both from arrays of random mutagenesis and by employing cutting-edge CRISPR genome editing techniques. This project will deliver speed-breeding methods and high-resilience mutants inspired by natural adaptations and newly formulated biological principles, to be introduced into a wide range of commercial pennycress varieties to precisely adapt them to the desired local environments.**

Field pennycress (*Thlaspi arvense* L.; pennycress) is a winter annual oilseed cover crop with extreme cold hardiness and a short life cycle enabling off-season integration into corn and soybean rotations across the Midwest. While undergoing rapid domestication over the past several years, pennycress has become a promising bioenergy crop due to high seed oil content with a fatty acid composition suitable for biofuel. Pennycress fields are susceptible to winter snow melt and spring rainfall, leading to waterlogged soils where the plants' roots are submerged under water. Waterlogging has been reported to cause yield loss and negative impacts on oil quality in *Brassica napus*, a close-relative of pennycress. The objective of this research was to determine if waterlogging had a significant effect on morphology and yield of two pennycress reference lines (Spring 32-10 and MN106) and to develop a model to predict the effective water availability for natural pennycress populations. One week of waterlogging during the reproductive stage of development caused several negative phenotypes, including early senescence, aborted silicles, and a reduction in root tissue compared to controls. Additionally, seed count and seed weight in Spring 32-10 waterlogged plants were significantly decreased compared to the controls, implying yield loss in pennycress fields that experience waterlogging from heavy spring rainfall events. The Spring 32-10 and MN106 accessions showed differences in morphology and seed yield in response to waterlogging, demonstrating natural variation in pennycress accessions can contribute to waterlogging resilience. To identify the climate-divergent natural pennycress accessions for waterlogging resilience, we also developed a mathematical model that leverages ORNL's climatype modeling framework to predict effective water availability at high geospatial resolution. The model integrates multiple layers of pre-processed climatic information (e.g., soil water concentration, precipitation, evapotranspiration,

temperature, solar radiation, aridity, and slope) in a weighted linear combination to produce a distribution of effective water availability scores. A score is computed for each of the 500 natural accession locations in our current collection (https://www.pennycressresilience.org/sample-collection) and then visualized on a map to facilitate the selection of genotypes of interest. The contribution of each climatic layer in the score was optimized by systematically varying the weight of each layer and observing the changes in the effective water availability distribution. Future work involves screening the accessions identified by our model for waterlogging resilience, as well as analyzing pennycress waterlogging transcriptome data, to identify genetic and phenotypic variation contributing to waterlogging resilience. We plan to apply our model to additional climate variables to identify natural pennycress populations with resilience to diverse abiotic conditions. As domesticated pennycress (aka Covercress) is adopted as a crop across the Midwest, this research will aid the development of elite varieties with enhanced tolerance to waterlogging.

*This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021286.*

**Root adaptive responses for improvement of abiotic stress tolerance in Pennycress**

Marcus Griffiths[1*] (mgriffiths@danforthcenter.org), Vannessica Jawahir[1], G Cody Bagnall[1], Alexander Liu[1], Kong Wong[1], Tyler Dowd[1], Dmitri Nusinow[1], Christopher N Topp[1], and **John Sedbrook**[2]

[1]Donald Danforth Plant Science Center, St. Louis, MO 63132, United States; [2]Illinois State University, Normal, IL

https://www.pennycressresilience.org/

**Project Goals: This project will employ evolutionary and computational genomic approaches to identify key genetic variants that have enabled *Thlaspi arvense* L. (Field Pennycress; pennycress) to locally adapt and colonize all temperate regions of the world. This, in combination with knowledge of metabolic and cellular networks derived from first principles, will guide precise laboratory efforts to create and select high-resilience lines, both from arrays of random mutagenesis and by employing cutting-edge CRISPR genome editing techniques. This project will deliver speed-breeding methods and high-resilience mutants inspired by natural adaptations and newly formulated biological principles, to be introduced into a wide range of commercial pennycress varieties to precisely adapt them to the desired local environments.**

Abstract: Roots are the interface between the plant and the soil and play a central role in multiple ecosystem processes. With intensification of agricultural practices, rhizosphere processes are being disrupted and are causing degradation of the physical, chemical, and biotic properties of soil. Improvement of ecosystem service performance is rarely considered as a breeding trait due to the complexities and challenges of belowground evaluation. Advancements in root phenotyping and genetic tools are critical in accelerating ecosystem service improvement in cover crops. Here I will present root phenotyping approaches for assessing ecosystem service in a prospective cash cover crop; pennycress (*Thlaspi arvense* L.). In development is a large format mesocosm system that will allow 3D root system architecture analysis of multiple plants. Using this system, we will be assessing how variation in pennycress root system architecture can affect ecosystem service and abiotic stress tolerance with the plant to scale from single plant to canopy level traits.

*This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021286.*

**Tolerance to Heat Stress in Natural Variants and CRISPR Gene-Edited High Oleic Acid Lines of the Oilseed Plant Pennycress (*Thlaspi arvense*).**

Nikhil Jaikumar[1*] (njaikum@ilstu.edu), Vanessica Jawahir,[2] Brice Jarvis,[1] Maliheh Esfahanian,[1] Dmitri A. Nusinow,[2] and **John Sedbrook**[1] (jcsedbr@ilstu.edu)

[1]Illinois State University, Normal, IL; [2]Donald Danforth Plant Science Center, St. Louis, MO

https://www.pennycressresilience.org/

**Project Goals: This project employs evolutionary and computational genomic approaches to identify key genetic variants that have enabled *Thlaspi arvense* L. (Field Pennycress; pennycress) to locally adapt and colonize all temperate regions of the world. This, combined with knowledge of metabolic and cellular networks derived from first principles, guides precise laboratory efforts to create and select high-resilience lines, both from arrays of random mutagenesis and by employing cutting-edge CRISPR genome editing techniques. This project will deliver speed-breeding methods and high-resilience mutants inspired by natural adaptations and newly formulated biological principles into a wide range of commercial pennycress varieties to precisely adapt them to the desired local environments.**

Abstract: Pennycress (*Thlaspi arvense* L.; Field Pennycress) is a cold-temperate and boreal-zone winter annual currently being rapidly domesticated for use as an oilseed crop with winter cover crop benefits. In contrast to its extreme cold tolerance, pennycress is susceptible to heat stress, particularly during flowering. To assess pennycress responses to heat stress across developmental stages, we tested the responses of seedlings and flowering plants to elevated temperatures being experienced more frequently due to climate change. We found that pennycress seedlings elongate hypocotyls at 28 °C, displaying variation between natural populations originally collected from different latitudes and altitudes. This natural variation to heat stress was also observed in plants during reproductive growth as Minnesota-collected MN106 was significantly more tolerant to heat stress than Montana-collected Spring32. In addition, we tested pennycress natural populations (wild-type) and lines CRISPR-edited to produce high-oleic triacylglycerides for lipid peroxidation and seed yield following a heat stress treatment, as well as pollen viability at a series of progressively increasing temperatures. We hypothesized natural lines from low latitudes as well as high-oleic lines could show improved tolerance to heat stress, the latter due to greater oxidative stability and reduced membrane fluidity. This hypothesis appeared correct; for example, high-oleic pennycress showed improved pollen viability at 28 – 30 °C compared to the corresponding wild type, but not at lower or higher temperatures. High-oleic pennycress lines also exhibited higher seed yields following a one-week heat stress at 32 °C. This effect was most likely not due to changes in lipid peroxidation levels, which did not differ between the lines. In a separate experiment, high-temperature stress was found to affect both male and female fertility, as well as seed fitness in the next generation. These results indicate that natural variants can be investigated to discover heat-tolerant varieties and that lines with reduced fatty acid desaturation, in addition to improved seed oil quality, could confer improved heat stress tolerance as an additional benefit. The causative variation is being explored.

*This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021286.*

**Pennycress as an Emerging Bioenergy Crop: How Does the Microbiome Impact Performance and Resilience Factors in the inland Pacific Northwest?**

Tarah S. Sullivan[1], Karen A. Sanguinet[1*](karen.sanguinet@wsu.edu), Isaac Madsen[1], Shyam Solanki[1,2], Pubudu Handakumbura[3] and **John Sedbrook[4]**

[1]Washington State University, Pullman, WA;[2]South Dakota State University, Brookings, SD; [3]DOE Pacific Northwest National Laboratory EMSL, Richland, WA; and [4]Illinois State University, Normal, IL

https://www.pennycressresilience.org/

**Project Goals: This project employs evolutionary and computational genomic approaches to identify key genetic variants that have enabled *Thlaspi arvense* L. (Field Pennycress; pennycress) to locally adapt and colonize all temperate regions of the world. This, in combination with knowledge of metabolic and cellular networks derived from first principles, is guiding precise laboratory efforts to create and select high-resilience lines, both from arrays of random mutagenesis and by employing cutting-edge CRISPR genome editing techniques. This project will deliver speed-breeding methods and high-resilience mutants inspired by natural adaptations and newly formulated biological principles, to be introduced into a wide range of commercial pennycress varieties to precisely adapt them to the desired local environments.**

Abstract: Pennycress (*Thlaspi arvense*; field pennycress) is under development as a winter annual oilseed bioenergy crop for the 80 million-acre U.S. Midwest Corn Belt and other temperate regions including the Pacific Northwest. Pennycress has unique attributes such as extreme cold tolerance and rapid spring growth. Off-season integration of domesticated pennycress varieties into existing corn and soybean acres would extend the growing season on established croplands, avoid displacement of food crops, and yield up to 3 billion gallons of seed oil annually. Pennycress oil has a fatty acid composition well-suited for conversion to biodiesel and biojet fuel. Academic, governmental, and industrial stakeholders are working closely to commercialize domesticated pennycress varieties that can yield over 1,500 pounds per acre of seeds producing 65 gallons of oil per acre annually; the first commercial planting of Covercress (domesticated pennycress) occurred this past fall, 2021. However, the adaptability of first-generation Covercress varieties to and resilience against environmental challenges including drought, heat, and flooding is extremely limited. Therefore, crucial work remains to identify mechanisms conferring stress tolerance for incorporation of next generation elite pennycress varieties into current agronomic systems. While the rhizosphere microbiome is known to confer many plant growth-promoting characteristics to most crops, very little is known regarding pennycress adaptations and reciprocal dependence on the rhizosphere assemblage for stress resistance and tolerance. We will present our initial findings detailing and seeking to understand the pennycress rhizosphere microbiome. We believe dynamic interactions of the soil microbiome with the pennycress metabolome form crucial relationships to support agronomic development of

this oilseed bioenergy crop in the inland Pacific Northwest and Midwest. Key findings from this work will be translatable to improving other Brassica crops important for bioenergy including camelina, carinata, rapeseed, and canola.

*This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021286.*

**Title: Structural and redox requirements for dioxygen generation coupled to metal reduction by methanobactins: implications for greenhouse gas emissions**

**Authors:** Phillip Dershwitz[1*](phild@iastate.edu), **Jeremy D. Semrau[2],** [1]Alan A. DiSpirito

**Institutions:** [1]Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011-3260, USA, [2]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109-2125, USA

**Website URL:** https://emmb.engin.umich.edu/

**Project Goals:** It is the general goal of this project to determine how microbial competition for copper affects net greenhouse gas emissions, specifically methane and nitrous oxide

**Abstract Text:** Methanobactins (MBs) are low molecular mass (<1,300Da), high-potential ($E_m$ of 483 – 745mV), ribosomally synthesized posttranslationally modified peptides (RiPPs) and represent the extracellular component of a copper acquisition system in a variety of methanotrophs. MBs are characterized by a unique pair of heterocyclic groups with an associated thioamides and coordinate Cu via a N from each heterocyclic group and a S from each thioamide in an $N_2S_2$ ligand set. Structurally, MBs are divided into two groups with Group I MBs, represented by the MB from *Methylosinus trichosporium* OB3b (MB-OB3b), have a dicyclic structure while Group II MBs, represented by the MB from *Methylocystis* strain SB2 (MB-SB2), has a hairpin-like structure.

In addition to copper, MBs will bind a variety of metals and reduce some but not all following binding. In MB-OB3b, metals such as $Cu^{2+}$ are bond and reduced to $Cu^{+}$ using both oxazolone groups with associated thioamides. In the presence of a molar excess of $Cu^{2+}$ MB-OB3b will also catalytically reduce $Cu^{2+}$ to $Cu^{1+}$. MB-OB3b binds other metals, such as $Fe^{3+}$, as a dimer using the N-terminal oxazolone group and thioamide from each MB-OB3b. Both $Cu^{2+}$ and $Fe^{3+}$ bound by MB-OB3b show a similar $N_2S_2$ coordinated, however, only metals bound as monomers are reduced. In contrast to MB-OB3b, MB-SB2 coordinates all metals as a monomers and oxidized metals, including $Fe^{3+}$, are reduced following binding. As observed with MB-OB3b. in the presence of a molar excess of metal, MB-SB2 will catalytically reduce metals at a rate of approximately 1 electron • $min^{-1}$ • MB-SB2. Following the loss of 4-5 electrons to metal reduction, MBs will oxidize $2H_2\mathbf{O}$ to $4H^+ + \mathbf{O_2}$ demonstrating a mechanism by which methanotrophs expressing MB can oxidize methane under anoxic conditions through "self-generation" of dioxygen required for the initial oxidation of methane to methanol ($CH_4 + \mathbf{O_2} \rightarrow CH_3\mathbf{O}H + H_2\mathbf{O}$). To initiate studies on the mechanism of water oxidation by MBs, generation of dioxygen following metal reduction is examined in holo-MBs as well as MBs minus the N-terminal or C-terminal heterocyclic groups. The results provide insights into the structural elements necessary to catalyze water oxidation by MBs.

**Publications Resulting from this Project**

1. Dershwiz P, Bandow NL, Yang J, Semaru JD, McEllistrem MT, Heinze RA, Fonseca M, Ledesma JC, Jennett JR, DiSpirito AM, Athwal NS, Hargrove MS, Bobik TA, Zischka H, DiSpirito AA. 2021. Oxygen generation via water splitting by a novel biogenic metal ion binding compound. Appl Environ Microbiol 87:in press.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0020174.

**Title: Microbial competition for copper can enhance greenhouse gas emissions**

**Authors:** Peng Peng[1*] (pepeng@umich,edu), Jin Chang[1], Alan A. DiSpirito[1], **Jeremy D. Semrau[1]**

**Institutions:** [1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109-2125, USA, [2]Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA 50011, USA.

**Website URL:** https://emmb.engin.umich.edu/

**Project Goals:** It is the general goal of this project to determine how microbial competition for copper affects net greenhouse gas emissions, specifically methane and nitrous oxide.

**Abstract Text:** Copper is an important trace nutrient for both methanotrophs and denitrifiers as it controls the activity and expression of key metabolic enzymes involved in methane consumption and nitrous oxide production, i.e., the methane monooxygenase (MMO) and nitrous oxide ($N_2O$) reductase (NosZ). Considering that methanotrophs and denitrifiers can co-exist in many environments, significant competition for copper can occur between these microbes. To collect copper, some methanotrophs secrete a copper-binding compound named methanobactin (MB) that has very high affinity for copper. MB is secreted into the environment, and after binding copper, is re-internalized via a specific TonB-dependent transporter (TBDT). *Methylosinus trichosporium* OB3b has two such MB uptake systems - one for uptake of its own MB (MB-OB3b), and another for uptake of MB from other methanotrophs, e.g., MB of *Methylocystis* sp. SB2 (MB-SB2). Interestingly, expression of the TBDT for MB-SB2 uptake in *M. trichosporium* OB3b is induced by MB-SB2, suggesting that some methanotrophs actively "steal" MB from other microbes. Moreover, both TBDTs are involved in regulating the expression of alternative MMOs in *M. trichosporium* OB3b. Active uptake of both forms of MB and the corresponding metabolic regulation thus exerts not only strong control on overall methane consumption by *M. trichosporium* OB3b, but also the activity of other methanotrophs. Interestingly, however, copper bound to MB is unavailable to denitrifiers as they lack the TBDT required for uptake. As a result, in the presence of MB or MB-expressing methanotrophs, complete conversion of nitrate to dinitrogen by denitrifying bacteria does not occur. Rather, as copper is unavailable to these microbes in the presence of MB, NosZ expression and activity is substantially reduced, significantly increasing $N_2O$ emissions. Interestingly, MB-OB3b and MB-SB2 differentially affected copper uptake, *nosZ* expression and $N_2O$ production by multiple denitrifiers. This study provides critical information for enhanced understanding of microbial copper competition that are important for the development of better predictive models of net greenhouse gas emission that are significantly controlled by microbial activity.

**References/Publications**


1. "Two TonB-dependent Transporters in *Methylosinus trichosporium* OB3b are Responsible for Uptake of Different Forms of Methanobactin and are Involved in the Canonical 'Copper Switch'". P. Peng, C. S. Kang, J. Chang, W. Gu, A. A. DiSpirito, J. D. Semrau. *Applied & Environmental Microbiology*. https://doi.org/10.1128/AEM.01793-21
2. "Evidence for Methanobactin "Theft" And Novel Chalkophore Production in Methanotrophs: Impact on Methanotrophic-Mediated Methylmercury Degradation". C. S. Kang-Yun, X. Liang, P. Dershwitz, W. Gu, A. Schepers, A. Flatley, J. Lichtmannegger, H. Zischka, L. Zhang, X. Lu, B. Gu, J. C. Ledesma, D. J. Pelger, A. A. DiSpirito, and J. D. Semrau. *The ISME Journal*. https://doi.org/10.1038/s41396-021-01062-1.
3. "Differential Stimulation of Nitrous Oxide Production from Denitrifying Bacteria Exposed to Methanobactin From Either *Methylocystis* sp. strain SB2 or *Methylosinus trichosporium* OB3b". J. Chang, P. Peng, A. A. DiSpirito, and J. D. Semrau. Submitted to *Applied & Environmental Microbiology*.



**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0020174.

**Genome-scale structural prediction of protein sequences and complexes with deep learning**

**Authors:** Mu Gao[1]* (mu.gao@gatech.edu), Davi N. An[1], Mark Coletti[2], Russell B. Davidson[2], Jerry M. Parks[2], Jianlin Cheng[3], Ada Sedova[2], and **Jeffrey Skolnick**[1]

**Institutions:** [1]Georgia Institute of Technology, Atlanta, GA; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]University of Missouri, Columbia, MO.

**Project Goals:** With the advances in next generation sequencing technologies, the number of sequenced genomes is growing exponentially. This has resulted in a bottleneck for the translation of sequence information into functional hypotheses about each gene. Current gene annotation technologies are primarily based on evolutionary inference by sequence comparison; however, many proteins in a proteome remain uncharacterized. To address this challenge, this collaborative team is currently developing a suite of novel high-performance-computing (HPC), deep-learning methods that predict protein structures at unprecedented accuracy, making use of the Summit supercomputer at the DOE leadership computing facility at the Oak Ridge National Laboratory. The combination of deep learning, HPC, and structural-based analysis will help break the gene annotation bottleneck and enable rapid, accurate prediction of gene function on a genomic scale.

**Abstract text:** One key aspect of protein annotation is the atomic structure encoded in a protein sequence. The release of AlphaFold2 in July 2021 has provided a powerful deep learning based computational method to decode protein sequences by predicting high confidence structural models of individual proteins. Taking advantage of this advance, our team has developed a genome-scale protein structural modeling and analysis pipeline using AlphaFold2 and deployed this workflow successfully on several full proteomes on Summit. This workflow has been applied to a few bacterial and plant species of interests to DOE's Office of Biological and Environmental Science. One of them is *Smagellanicum*, whose proteome consists of about 25,227 protein sequences with 11 million total amino acids. We have modeled 25,134 (99.7%) of the proteome and about 57% of sequences have at least one high confidence model.

Moreover, interactions between proteins are vital to the understanding of their functions. Excitingly, we have developed AF2Complex, a generalization of AlphaFold2 for predicting physical interactions between different proteins via deep docking, i.e., by exploring physically favorable structural models of a putative multimeric protein complex with the same deep learning neural networks originally developed for modeling a single protein sequence. By incorporating a new evaluation metric and optimizing input data, AF2Complex can effectively predict if a set of protein sequences interact, and if so, then provide high-confidence models for the predicted protein complex. In a benchmark test on dimeric protein pairs, it achieves higher accuracy than strategies that combine AlphaFold2 and protein-protein docking. It also achieved significantly better performance than DeepMind's AF-multimer when the same set of deep learning models of AF-multimer are used. Importantly, going beyond most other approaches that focus on dimeric protein pairs, we demonstrate that a protein complex consisting of multiple proteins can be accurately

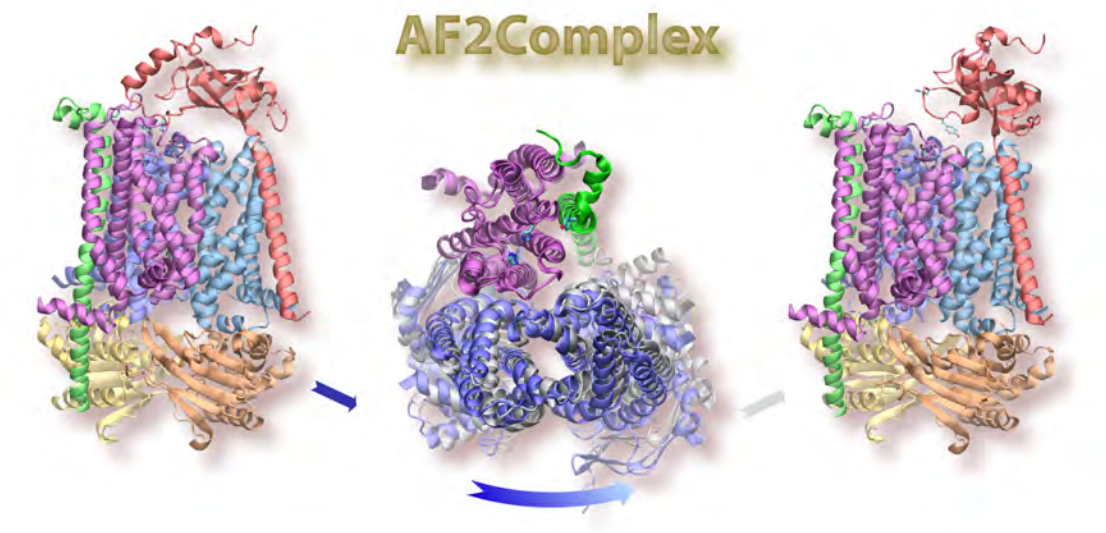



Figure 1. AF2Complex models of the cytochrome c maturation system $CcmA_2B_2CD$ from E. coli. Conformational changes relevant to the function of the system are observed in the superimposition (center) top two models (left and right). The superimposition of the two models uses one component (CcmE in purple) as the reference and is viewed from a perspective that is above the two individual models. For clarity, CcmF (red) is not shown in the superposition.

modeled using this deep learning approach. AF2Complex was successfully validated on some challenging CASP14 multimeric targets, a small but appropriate benchmark set, and the *E. coli* proteome. In a practical application, using the cytochrome *c* biogenesis system as an example, we predict high-confidence models of three sought-after assemblies formed by eight members of this system (the predicted models of one complex assembly are shown in Figure 1 above). To the best of our knowledge, this is the first time that a deep learning algorithm has been applied to model the atomic structures of a whole biomolecular system.

**References/Publications**

1. Gao, M., Lund-Andersen P, Morehead A, Mahmud S, Chen C, Chen X, Giri N, Roy R S, Quadir F, Effler T C, Prout R, Abraham S, Skolnick J, Cheng J, Sedova A. *High-Performance Deep Learning Toolbox for Genome-Scale Prediction of Protein Structure and Function.* In Machine Learning in HPC Environments, held in conjunction with The International Conference for High Performance Computing, Networking, Storage, and Analysis. 2021. St. Louis.
2. Gao, M, An D, Parks, P, and J. Skolnick. *Predicting direct physical interactions in multimeric proteins.* bioRxiv, 2021: p. 2021.11.09.467949.

**Funding statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021303.

**Title: Identifying microbial drivers in biological phenotypes with a Bayesian Network Regression model**

**Authors:** Samuel Ozminkowski[1]* (ozminkowski@wisc.edu) and **Claudia Solis-Lemus**[1]

**Institutions:** [1]University of Wisconsin-Madison, Madison, WI

**Website:** https://github.com/samozm/BayesianNetworkRegression.jl

**Project Goals: Short statement of goals. (Limit to 1,000 characters)**

1. Assess the applicability of a novel Bayesian Network Regression framework to microbiome applications.
2. Implement a fast and efficient sampling algorithm to sample posterior conditional distributions for the model.
3. Release an open-source and publicly available package in the Julia programming language so that domain scientists can utilize this model on their data.

**Abstract text:** Please limit such that entire document does not exceed 2 pages.

Microbial communities are among the main driving forces of biogeochemical processes in the biosphere. Understanding the composition of microbial communities and how these compositions shape specific biological phenotypes is crucial to comprehend complex biological processes in soil, plants and humans alike. Standard approaches to study the connection between microbial communities and biological phenotypes rely on abundance matrices to represent the microbial compositions. Different experimental settings are defined and then microbial compositions are measured (as abundances) on each experimental setting. Next, the abundance matrices are used as input in a regression-type (or machine-learning) analysis to relate the microbial community to phenotypes of interest.

Given that relative abundances only provide a snapshot of the composition of the community at the specific time of sampling and do not account for correlations between microbes, microbial interaction networks have been recently preferred to represent microbial communities. Yet models to connect a microbial network to a biological phenotype remain unknown. There has only been a handful of new methods that aim to identify associations between network predictors and a phenotype via a regression framework. However, these methods have only been studied for brain connectome networks which, unlike microbial networks, are intrinsically dense. In conclusion, methods to find associations between a sample of microbial networks and a biological phenotype remain unknown.

In this work, we introduce a Bayesian Network Regression (BNR) model that uses the microbial network as the predictor of a biological phenotype. This model intrinsically accounts for the interactions among microbes and is able to identify influential edges (interactions) and influential

nodes (microbes) that drive the phenotypic variability. While the model itself is not new, it has only been studied for brain connectome networks, and thus, its applicability to microbial networks which are inherently more high-dimensional and sparser has not been studied. Furthermore, unlike in brain connectome research, in microbiome research, it is usually expected that the presence of microbes have an effect on the response (main effects), not just the interactions. Here, we develop the first thorough investigation of whether Bayesian Network Regression models are suitable for microbial datasets on a variety of synthetic data that was generated under realistic biological scenarios. We test whether the Bayesian Network Regression model that accounts only for interaction effects (edges in the network) is able to identify key drivers in phenotypic variability (microbes). We show that this model is able to identify influential nodes and edges in the microbial networks that drive changes in the phenotype for most biological settings, but we also identify scenarios where this method performs poorly which allows us to provide practical advice for domain scientists aiming to apply these tools to their datasets. In addition, we implement the method in an open-source publicly available and easy-to-use new Julia package (BayesianNetworkRegression.jl) with online documentation and step-by-step tutorial which will allow scientists to easily apply this model on their own data.

This research directly addresses the DOE SC program goals of developing computational approaches that can integrate large omics data types from multiple and heterogeneous sources, such as those used in the Genomic Science program. Our open-source software will be incorporated into the DOE Systems Biology Knowledgebase, an open-source software platform that serves the systems biology research community.

**References/Publications**

1. Ozminkowski, Samuel and Claudia Solis-Lemus “Identifying microbial drivers in biological phenotypes with a Bayesian Network Regression model” (in preparation).

**Funding statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021016.

**Developing anaerobic fungal tools for efficient upgrading of lignocellulsic feedstocks**

Casey Hooker[1,3*] (hookerc@udel.edu), Radwa Hanafy[3], Ethan Hillman[1,2†], Javier Muñoz Briones[1,2], and **Kevin Solomon[1,2,3]**

[1]Agricultural & Biological Engineering, Purdue University, West Lafayette, IN; [2]Purdue University Interdisciplinary Life Sciences (PULSe) Program, Purdue University, West Lafayette, IN; [3]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE;

[†]Present Address: Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI

http://solomonlab.weebly.com

**Project Goals: This project develops genetic and epigenetic tools for emerging model anaerobic fungi to identify the genomic determinants of their powerful biomass-degrading capabilities, facilitate their study, and enable direct fungal conversion of untreated lignocellulose to bioproducts.**

Deconstruction of plant cell wall biomass is a significant bottleneck to the production of affordable biofuels and bioproducts. Anaerobic fungi (*Neocallimastigomycota*) from the digestive tracts of large herbivores, however, have evolved unique abilities to degrade untreated fiber-rich plant biomass by combining hydrolytic strategies from the bacterial and fungal kingdoms[1]. Anaerobic fungi secrete the largest known diversity of lignocellulolytic carbohydrate active enzymes (CAZymes) in the fungal kingdom (>300 CAZymes), which unaided can degrade up to 60% of the ingested plant material within the animal digestive tract[2,3]. Unlike many other fungal systems, these CAZymes are tightly regulated and assembled in fungal cellulosomes to synergistically degrade plant material, including untreated agricultural residues, bioenergy crops, and woody biomass, with comparable efficiency regardless of composition[1,4–6]. Our efforts to characterize gut fungal CAZymes reveal industrially relevant properties such as remarkable stability and activity towards untreated plant biomass[6-7]. Gut fungal CAZymes liberate sugars from cellulosic substrates for timeframes beyond 192 hours after inoculation. These accumulated sugars can be fed to model bioproduction hosts (e.g. *K. marxianus*) to create coculture systems capable of upgrading sugars from plant biomass to high value products (e.g. ethyl-acetate and 2-phenylethanol)[7]. Similarly, anaerobic fungal biosynthetic enzymes possess unique cofactor substrate preferences that support higher catalytic efficiencies, which are easily overlooked via heterologous expression due to the extremely high AT content (~83%) of gut fungal genomes and biased codon preferences[7]. Thus there is an unmet need to build genetic tools and methods to study these enzymes natively in anaerobic fungi.

As a first step towards genetic tool development, we sequenced the genomes of three novel specimens of anaerobic fungi representing two genera of Neocallimastigocota. Previous anaerobic fungal genomes were highly fragmented into as many as 30,000 scaffolds thus limiting efforts to mine for basic genetic parts (e.g. promoters and terminators). By leveraging high quality genome isolations, long-read sequencing, and Hi-C (chromosomal conformation capture) sequencing, we

have improved genome assembly by an order of magnitude to generate the first genomes for this phylum of organisms with chromosomal resolution. Our assemblies incorporate more than 99% of the genome into 12-25 chromosomes with N50 < 10. These high quality reference genomes show extensive genome duplication and horizontal gene transfer events that partially contribute to the ability of gut fungi to robustly degrade crude plant biomass. We are investigating the epigenetic regulation of these genes by monitoring expression through transcriptomic and proteomic approaches supported through the JGI-FICUS program. These studies show epigenetic control, induced in part by microbial competition, mediates enzyme expression and substrate preference, and thus which parts of plant biomass are preferentially degraded. Thus, these reference genomes and datasets reveal a wealth of regulatory sequences and CAZymes for study. We have synthesized more than a dozen promoters and terminators as well as nearly 20 codon optimized reporter proteins through the JGI-BERSS program for evaluation along with reporter proteins and other putative parts for gene expression. By combining these regulatory sequences with selectable markers we have verified functionality of 2 fluorescent reporters via flow cytometry. Current efforts are aimed at improving transformation efficiencies and expression of selectable markers by mining for centromere binding sequences, Kozak sequences, and autonomously replicating sequences to further enhance plasmid stability and maintenance. Together, our rich high quality reference genomes and datasets are expediting efforts to mine for key regulatory sequences and parts needed to build a stable episomal plasmid.

In summary, the ongoing work has harnessed anaerobic fungal isolates and genes for bioproduction, begun to identify parts for a genetic toolbox and provides an atlas of anaerobic fungal genomes complete with targets for study and genomic regions to avoid. Once validated, these parts will form foundational tools to generate a deeper systems-level understanding of anaerobic fungal physiology while establishing fundamental knowledge about regulation of gut fungal CAZymes. Ultimately, we enable predictive biology in anaerobic fungi and derive insight into microbial plant deconstruction to advance the development of economical biofuels and bioproducts.

## References


**1** Haitjema, C. H. *et al. Nature Microbiology* 2, 17087 (2017)
**2** Seppälä, S. *et al. Metabolic Engineering* 44, 45–59 (2017)
**3** Youssef, N. H. *et al. Appl. Environ. Microbiol.* 79, 4620–4634 (2013)
**4** Solomon, K. V. *et al. Science* 351, 1192–1195 (2016)
**5** Solomon, K. V. *et al. Fungal Genetics and Biology* 121, 1–9 (2018)
**6** Hooker, C. A. *et al. Biotechnology for Biofuels* 11, 293 (2018)
**7** Hillman, E. T. *et al. Biotechnology Progress* 37, e3172 (2021)
**8** Hillman, E. T. *et al. Microorganisms* 9, 1986 (2021)



*This project is supported by the Office of Biological and Environmental Research through the DOE Office of Science under Award No DE-SC0022206. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL).*

**Plastic degradation and upcycling by the gut microbiome of yellow mealworms**

Lummy Maria Oliveira Monteiro[1,2*] (lummy@udel.edu), Jyoti Singh[1,2], Mark Blenner[1,2], **Kevin Solomon**[1,2]

[1]Delaware Biotechnology Institute, [2]Chemical & Biomolecular Engineering, University of Delaware, Newark, DE 19716

http://solomonlab.weebly.com
https://www.blennerlab.com/

**Project goals: This project discovers and reconstructs the plastic degradation pathways distributed across the gut microbiome of yellow mealworms (larvae of *Tenebrio molitor*) to develop enhanced capabilities for biologically based polymer recycling.**

Plastics, initially selected for their durability and environmental resiliency, pose a significant environmental challenge for modern economies. Polystyrene (PS), high- and low-density polyethylene (HDPE and LDPE), and polypropylene (PP) are produced at a rate of more than 228 million tonnes globally each year. However, none currently have robust infrastructures for mechanical or chemical recycling and ultimately become polluting waste streams. To address this need, we pursue biological strategies for plastics depolymerization. We focus on the microbiomes of insect larvae (colloquially called worms) as they degrade plastics more rapidly than microbial isolates and do not require clean plastics or pretreatment. In particular, the microbiome of yellow mealworms is unique in that its host does not appear to contribute to degradation of a wide range of plastics. While bacterial community members have been identified, the specific pathways responsible for biodegradation remain to be elucidated and the potential contributions of fungal members are unexamined. Additionally, emerging evidence suggests that nutrient supplementation enhances plastic metabolism up to 70% and gives rise to a gut community structure distinct from that without additional nutrients. However, it is unclear if nutrient supplementation induces microbes to participate in in the plastic degradation or if it supports an optimal community composition for function.

As a first step to address these gaps, we characterized the consumption rates of PS, LDPE, HDPE, and PP via *T. molitor* larvae in the presence and absence of co-fed oats as a nutritional supplement. The consumption rates of PS, LDPE, and HDPE were 20.4, 12, and 1.1, mg (100 larvae)$^{-1}$d$^{-1}$, respectively, in agreement with established studies. However, oat supplementation enhanced plastics consumption by 158.8, 60, and 232.1 %, respectively. These studies establish the use of oats as a potent supplement for enhancement of PS and LDPE consumption rates, up to double that obtained with established supplements, and validated HDPE consumption by *T. molitor*.

Antibiotic and antifungal selection studies supported the role of both fungal and bacterial populations in plastics consumption. Mealworms treated with penicillin/streptomycin to remove their bacterial population consumed PS 15% faster (60.6 mg (100 larvae)$^{-1}$d$^{-1}$) than untreated mealworms. Similarly, amphotericin B antifungal treatment selecting for bacteria enhanced PS consumption by 4.1% (55 mg (100 larvae)$^{-1}$d$^{-1}$). These results suggest that fungal communities in the mealworm gut microbiome are likely to play an important role in the plastic consumption and that the inter-kingdom relationship between bacteria and fungi may be antagonistic.

Worm-consumed plastics were chemically modified beyond simple mechanical degradation validating biological mechanisms for plastics depolymerization. Fourier transform infrared spectroscopy (FTIR) analysis of frass (excrement) from mealworms fed PS revealed incorporation of oxygen not found in untreated controls. Moreover, benzene ring cleavage was observed for treated PS samples. Similarly, FTIR spectra of the frass from mealworms fed LDPE revealed the incorporation of carbonyl and alcohol groups. Thermogravimetric analyses (TGA) of frass also confirmed the changes in physical properties, supporting the biodegradation of PS and LDPE via *T. molitor* microbiomes. Finally, gel permeation chromatography (GPC) of the frass of *T.molitor* larvae fed PS confirmed a decrease in polymer molecular weight and an increase in polydispersity. Taken together, these results demonstrate that the plastics being ingested by the larvae are being depolymerized.

Microbiome community analysis via 16s and ITS sequencing revealed a rich consortium of bacteria and fungi. The bacterial community was more diverse than the fungal community with observed taxa belonging to the bacterial phyla Firmicutes, Tenericutes, Proteobacteria, Actinobacteria, Spirochaetes, Bacteroidetes, and Fusobacteria, and fungal Ascomycota, Basidiomycota and Mucoromycota. As expected, mealworm diet led to unique community structures adapted to degradation of the fed plastic substrate. However, oats co-supplementation frequently selected for taxa that were not observed in plastics only or oats only controls suggesting currently unrecognized interactions. Despite these unique community structures, microcosms of communities in planktonic culture selected for with LDPE, HDPE, PS, and PP diet were all able to grow on LDPE as a sole-carbon source. Finally, our community analyses revealed obligate anaerobic genera such as *Selbadella* associated with PP degradation, suggesting potentially novel oxygen-independent pathways for plastics depolymerization.

In summary, our ongoing work has characterized plastic consumption rates in *T. molitor* microbiomes, revealing novel strategies to structure gut microbial populations for enhanced degradation. Plastics were noted to be metabolized and not only mechanically degraded by both bacterial and fungal communities that contribute to plastic degradation even independent of the host mealworm. Future studies will probe the contributions of individual taxa within these communities and generate systems-level insight into their metabolic pathways to develop consortia enriched in plastic degradation activity, and identify novel enzymes from community members.

*This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0022018. A portion of this research will be performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative and use resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL).*

**Improving Bioprocess Robustness by Cellular Noise Engineering**

Constantinos Katsimpouras,[1*] (ckatsimp@mit.edu), **Andreas E. Vasdekis**,[2] **Gregory N. Stephanopoulos**[1], (gregstep@mit.edu)

[1]Massachusetts Institute of Technology, Cambridge, MA, US; [2]University of Idaho, Moscow, ID, US

**Project Goals: This project sets out to develop an integrative workflow combining genome-wide editing, multi-omic and single-cell analyses, and computational models to enable cellular noise engineering and, thus, the design of robust biological systems that maintain function beyond standard laboratory conditions. In this context, cellular noise engineering bestows a non-genetic division of labor, where distinct subpopulations express pathways that maximize function under key environmental conditions. Specifically, the project focuses on improving the robustness of cellulosic oil and alkane production in *Yarrowia lipolytica* under genuine industrial conditions of time-varying toxic lignocellulosic hydrolysate inhibitor concentration and temperatures. To this end, our approach involves *Y. lipolytica* evolution, multi-omics and single-cell analyses, and construction of genome-wide kinetic models that will guide the noise engineering efforts.**

Growing global energy demands, environmental concerns, and the need to reduce dependency on decreasing fossil fuel reserves drive the pursuit of identifying alternative energy sources. The trucking, shipping, and aviation industries require heavy liquid fuels that can presently only be supplied by fossil fuels. Alternative forms of heavy fuels, such as microbial oils and alkanes, are gathering increasing interest; however, despite the increasing efficiency of microorganisms, such as *Y. lipolytica*, in converting sugars to microbial oils, the cost of glucose feedstocks is still very high. Lignocellulosic plant matter represents an inexpensive feedstock with an ample supply that does not compete with food resources. However, lignocellulosic biomass pretreatment increases not only the amount of released sugars but also the levels of toxic byproducts. Specialized methods to reduce the resulting toxicity levels come at high costs and processing complexity. Recently, we reported a more cost-effective and scalable approach to reduce hydrolysate toxicity by engineering the fermenting microorganism to reduce furans to furan alcohols[1], for which tolerance strategies are already available[2].

Our project aims to develop a similar phenotype in *Y. lipolytica* that can tolerate common lignocellulosic hydrolysate inhibitors, as well as efficiently produce cellulosic oils and alkanes by harnessing all available hydrolysate sugars. Our approach involves evolving *Y. lipolytica* to obtain variants that can tolerate, grow, and efficiently synthesize biofuel precursors under steady-state, albeit stressful conditions pertaining to both toxic hydrolysates and high temperatures. Multi-omics analyses of the evolved strains will inform about the dynamic control of key genes required for coping with varying stressful conditions and aid the construction of genome-wide metabolic kinetic models. These models will guide the selection of gene targets for engineering noise, which we will accomplish by replacing the native promoters of these genes with synthetic

ones that confer specific noise levels. Testing in programmable microfluidics will validate the effects of noise on robustness before undertaking similar validation efforts in bioreactors under industrially relevant conditions for producing cellulosic oils and, after further adaptation, cellulosic alkanes[3]. Overall, we anticipate that strains exhibiting optimal levels of cellular noise will also demonstrate robustness that maintains production under time-varying stresses and both laboratory and simulated industrial-scale conditions.

**References**


1. Lam, F. H. *et al.* Engineered yeast tolerance enables efficient production from toxified lignocellulosic feedstocks. *Sci. Adv.* **7**, (2021).
2. Lam, F. H., Ghaderi, A., Fink, G. R. & Stephanopoulos, G. Engineering alcohol tolerance in yeast. *Science (80).* **346**, 71–75 (2014).
3. Li, J. *et al.* Synthesis of high-titer alka(e)nes in *Yarrowia lipolytica* is enabled by a discovered mechanism. *Nat. Commun.* **11**, 6198 (2020).



*This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), Genomic Science Program, grant no. DE-SC0022016.*

**Title:** Dynamic *Phaeodactylum tricornutum* Exometabolites Shape Surrounding Bacterial Communities

**Authors:** Vanessa Brisson[1]* (brisson2@llnl.gov), Courtney Swink,[1] Ty Samo,[1] Xavier Mayali,[1] Trent Northen,[2,3] and **Rhona Stuart[1]**

**Institutions:** [1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]The DOE Joint Genome Institute, Berkeley, CA.

**Website URL:** http://bio-sfa.llnl.gov

**Project Goals: Algal and plant systems have the unrivaled advantage of converting solar energy and $CO_2$ into useful organic molecules. Their growth and efficiency are largely shaped by the microbial communities in and around them. The µBiospheres SFA seeks to understand phototroph-heterotroph interactions that shape productivity, robustness, the balance of resource fluxes, and the functionality of the surrounding microbiome. We hypothesize that different microbial associates not only have differential effects on host productivity but can change an entire system's resource economy. Our approach encompasses single cell analyses, quantitative isotope tracing of elemental exchanges, 'omics measurements, and multi-scale modeling to characterize microscale impacts on system-scale processes. We aim to uncover cross-cutting principles that regulate these interactions and their resource allocation consequences to develop a general predictive framework for system-level impacts of microbial partnerships.**

**Abstract Text:** Algal-bacterial interactions in phycosphere microbial communities have important implications for the stability and productivity of algal biofuel systems. However, the role of exometabolites in mediating these interactions and establishing microbial community structure are not well understood. We set out to characterize exometabolites from the model diatom *Phaeodactylum tricornutum* and investigate the role of those exometabolites in driving the composition of surrounding microbial communities.

First, to characterize the impacts of algal exudates on microbial community composition, we compared the compositions of bacterial communities originating from an algal enrichment (1) grown under three conditions: with alga (*P. tricornutum*) present, with algal exudates (spent medium from axenic *P. tricornutum* growth), and without any algal exudates (alga free controls). Using amplicon sequencing based microbial community analysis, we found that communities grown with algal spent medium were intermediate in composition between those grown with the alga present and algae free controls. Both algal spent medium and algal presence drove shifts in relative abundances compared to alga free controls. Thus, although the communities grown on algal spent medium and with algal were distinct, our results indicate that algal exudates, and the exometabolites present in them, are important for shaping surrounding bacterial communities.

To identify metabolites that could be driving microbial community composition, we used an LC-MS/MS metabolomics analysis to characterize the metabolite composition of the algal spent medium at different growth stages. We identified 58 metabolites produced by *P. tricornutum*

including organic acids, vitamins, amino acids, and nucleotides and derivatives of several of these. Exometabolite production was dynamic, with different metabolites accumulating during different algal growth phases. To investigate the potential role of specific metabolites in shaping microbial communities, we chose 14 identified exometabolites and tested their ability to selectively support the growth of 12 individual bacterial isolates from algal associated enrichment communities (1) and seawater. Of the 14 metabolites tested, three (4-hydroxybenzoic acid, shikimic acid, and hydroxyphenylacetic acid) supported growth of at least one of the bacterial isolates. However, each metabolite supported growth of a subset of the isolates, suggesting that increased levels of these metabolites could selectively increase growth of some bacteria and not others within a community.

To better understand the roles of specific exometabolites in the context of a complex microbial community, we extended the microbial community experiment above to compare community compositions in response to addition of two specific algal exometabolites: one which supported the growth of 4 out of 12 bacterial isolates (4-hydroxybenzoic acid), and another, lumichrome, a vitamin derivative which did not support detectable growth but which has been shown to affect *P. tricornutum* growth (2). Exogenous additions of both metabolites had significant impacts on bacterial community compositions, both with and without *P. tricornutum* present. Addition of 4-hydroxybenzoic acid led to substantial increases in relative abundances of two amplicon sequence variants (ASVs) which were identical to the 16S-V4 sequences of bacterial isolates that grew on 4-hydroxybenzoic acid in the experiment above. The addition of lumichrome also impacted the relative abundances of ASVs, but the responses differed, and were generally lower in magnitude than those to 4-hydroxybenzoic acid. Our work demonstrates the importance of specific algal exometabolites in driving microbial community composition and suggests that algal exometabolites have distinct effects on microbial communities, in part as a function of their bacterial nutritional capability.

**References/Publications**

1. Kimbrel JA, Samo TJ, Ward C, Nilson D, Thelen MP, Siccardi A, Zimba P, Lane TW, Mayali X. 2019. Host selection and stochastic effects influence bacterial community assembly on the microalgal phycosphere. *Algal Research* 40, 101489.

2. Brisson VB, Mayali X, Bowen B, Golini A, Thelen M, Stuart RK, Northen TR. 2021. Identification of effector metabolites using exometabolite profiling of diverse microalgae. *mSystems* 6, e00835-21.

**Funding Statement:** *This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL microBiospheres SFA, FWP SCW1039. LLNL-ABS-830258.*

**Title:** Siderophore-Producing Phycosphere Bacteria Alleviate Iron Limitation Stress in *Phaeodactylum tricornutum*

**Authors:** Coffey NR[1*] (coffeyni@oregonstate.edu), Rodriguez MCA[1], Rolison KA[2], Mayali X[2], Chu RK[3], Kew WR[3], Kimbrel JA[2], **Boiteau, RM[1,3], Stuart RK[2]**

**Institutions:** [1]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR; [2]Lawrence Livermore National Laboratory, Livermore, CA; [3]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA.

**Website URL:** https://bio-sfa.llnl.gov/

**Project Goals: Algal and plant systems have the unrivaled advantage of converting solar energy and $CO_2$ into useful organic molecules. Their growth and efficiency are largely shaped by the microbial communities in and around them. The µBiospheres SFA seeks to understand phototroph-heterotroph interactions that shape productivity, robustness, the balance of resource fluxes, and the functionality of the surrounding microbiome. We hypothesize that different microbial associates not only have differential effects on host productivity but can change an entire system's resource economy. Our approach encompasses single cell analyses, quantitative isotope tracing of elemental exchanges, 'omics measurements, and multi-scale modeling to characterize microscale impacts on system-scale processes. We aim to uncover cross-cutting principles that regulate these interactions and their resource allocation consequences to develop a general predictive framework for system-level impacts of microbial partnerships.**

**Abstract Text:** *Phaeodactylum tricornutum* (Pt) is a saltwater, low iron (Fe)-tolerant diatom ideal for use in studies probing the regulation of micronutrient availability by algae and their microbiome. While Pt's physiological response to Fe limitation has been fairly well-studied, its interactions with microbial partners is not well constrained. Microbial symbionts in Pt's phycosphere (a microscale zone of influence at/near the diatom's surface) may employ mutualistic strategies to enhance the growth of both the bacteria and the algae, possibly via the solubilization and subsequent bioavailability increase of micronutrient Fe. Through genome mining of bacteria isolated from Pt's phycosphere, three heterotrophic strains were identified with putative biosynthesis gene clusters for siderophores (small molecules that are secreted in response to Fe deficiency to solubilize and bind Fe to facilitate uptake). Initial results from co-culture experiments reveal enhanced growth of Pt in co-culture with each of these strains under low-Fe conditions relative to an axenic Pt control. Mass spectrometry based metallomic analyses of spent media reveal the production of metal chelators at low Fe concentrations within Pt-microbial cocultures but not in the axenic control. Notably, in response to Fe limitation stress, a novel siderophore was secreted by a Pt co-culture with bacterial isolate Stappia sp., a member of the Rhodobacteriaceae commonly found associated with microalgae in nature. This siderophore is believed to be produced via a nonribosomal peptide synthetase pathway by Stappia sp.. These results suggest that microbial partners may facilitate algal uptake of Fe when it may otherwise

limit algal growth and production. Since siderophore-bound Fe can only be utilized by microbes that can actively take up those specific molecules, this strategy may also play a broader role in managing Fe bioavailability by promoting the growth of certain taxa within a complex community. Metabolomic studies such as these are a promising avenue for understanding how host-microbe interactions manage resource allocation and contribute to algal biofuel pond community composition and resiliency.

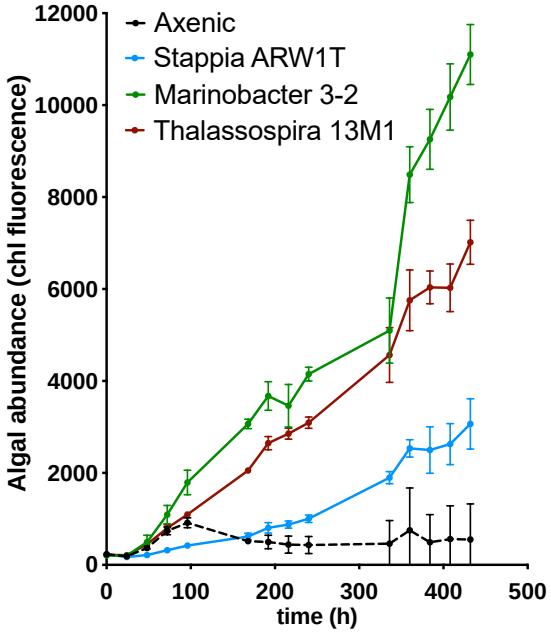



Figure 1: *P. tricornutum* growth response in 0.01 µM Fe in coculture with various single phycosphere taxa (Stappia, Marinobacter or Thalassospira) or axenic. Error bars represent standard deviation between 4 biological replicates.

**Funding Statement:** *This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL microBiospheres SFA, FWP SCW1039. A portion of this research was performed on a project award from the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830.*

**Title: Mycorrhizal Fungi Mediate Plant and Bacterial Response to Water Limitation in a Marginal Soil**

**Authors:** Rachel Hestrin[1*] (hestrin1@llnl.gov), Megan Kan[1], Rebecca Ju[1], Jeffrey Kimbrel[1], Christina Ramon[1], Prasun Ray[2], Rina Estera-Molina[3], Steven Blazewicz[1], **Kelly Craven[2], Mary Firestone[3], Peter Weber[1], Rhona Stuart[1], Erin Nuccio[1], Jennifer Pett-Ridge[1]**

**Institutions:** [1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Noble Research Institute, Ardmore, OK; [3]University of California, Berkeley, CA

**Website:** https://bio-sfa.llnl.gov/

**Project Goals: Algal and plant systems have the unrivaled advantage of converting solar energy and $CO_2$ into useful organic molecules. Their growth and efficiency are largely shaped by the microbial communities in and around them. The µBiospheres SFA seeks to understand phototroph-heterotroph interactions that shape productivity, robustness, the balance of resource fluxes, and the functionality of the surrounding microbiome. We hypothesize that different microbial associates not only have differential effects on host productivity but can change an entire system's resource economy. Our approach encompasses single cell analyses, quantitative isotope tracing of elemental exchanges, 'omics measurements, and multi-scale modeling to characterize microscale impacts on system-scale processes. We aim to uncover cross-cutting principles that regulate these interactions and their resource allocation consequences to develop a general predictive framework for system-level impacts of microbial partnerships.**

**Abstract text:**

Background: Multipartite mutualisms between plants, mycorrhizal fungi, and other soil biota can enhance plant productivity, resilience to stress, and carbon (C) allocation below ground. A better understanding of these relationships can inform more sustainable management of cellulosic bioenergy crops, such as switchgrass (*Panicum virgatum* L.), a C4 perennial grass championed for its high biomass yields and tolerance to a broad spectrum of climatic conditions and soils unsuitable for intensive agriculture. We are investigating context-dependent multipartite relationships between *Panicum hallii*—a model species closely related to switchgrass, two mycorrhizal lineages with different genomic repertoires—the arbuscular mycorrhizal fungus *Rhizophagus irregularis* and the Sebacinales fungus *Serendipita bescii*, and microbial communities in a marginal soil. We hypothesize that both mycorrhizal lineages mediate microbial community composition and function, plant and microbial tolerance to resource limitation, and soil organic matter dynamics. However, we anticipate that due to their different genomic repertoires, each fungus will elicit a distinct biotic response.

Experimental design: We grew *P. hallii* with and without *R. irregularis* or *S. bescii* in microcosms containing 'live' soil harvested from a marginal Oklahoma pasture. Each microcosm contained a hyphal ingrowth core that excluded plant roots. This allowed us to investigate rhizosphere and hyphosphere processes separately. We maintained half of the microcosms under water-replete conditions and the other half under water-limited conditions. We also maintained half of the microcosms in a $^{13}CO_2$ atmosphere, which allowed us to track plant- and mycorrhizal-derived $^{13}C$ into other soil C pools (live and dead microorganisms, metabolites, mineral associated organic matter, particulate organic matter) and C fluxes ($CO_2$, dissolved C, volatile

compounds). We harvested the microcosms at 5, 8, and 12 weeks after the onset of $^{13}CO_2$ labeling. After 12 weeks, we also conducted quantitative stable isotope probing (qSIP) with $H_2{}^{18}O$ to assess microbial growth rates in hyphosphere soil. Although mycorrhizal fungi were present in the native soil microbial community, qPCR analyses show that the *R. irregularis* and *S. bescii* inoculant strains were more abundant in roots and soils harvested from respectively inoculated microcosms. Thus, differences observed in mycorrhizal-inoculated microcosms represent the additional effect of these inoculants despite the presence of a background microbial community.

Results: Soil moisture regime and mycorrhizal inocula altered plant and microbial activity, with significant effects on soil C cycling. In uninoculated microcosms, rhizosphere soil $^{13}C$ content was similar under both moisture regimes. In mycorrhizal-inoculated microcosms, rhizosphere soil $^{13}C$ was lower in water-replete compared to water-limited soil. This suggests that under water-replete conditions, plants inoculated with *R. irregularis* and *S. bescii* either allocate less C below ground (possibly due to a reduced need to invest in resource acquisition) or that the C that they allocate below ground is consumed and respired more quickly than under water-limited conditions. Following water limitation, bacterial growth rates, microbial growth efficiency, and the diversity of the active bacterial community were suppressed in uninoculated soils, but not in soils inoculated with either fungus. Several of the bacterial taxa that responded positively to mycorrhizal inocula in water-limited soil belong to lineages that are considered drought-susceptible. Although both fungi supported bacterial communities exposed to water limitation, *R. irregularis* elicited a stronger positive response than *S. bescii.* Together, our findings demonstrate that context-dependent mycorrhizal relationships influence biotic response to resource limitation.

**Funding Statement:** *This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL Biofuels SFA, SCW1039 and Award Number DE-SC0014079 to UC Berkeley, Noble Research Institute, University of Oklahoma, Lawrence Livermore National Laboratory and Lawrence Berkeley National Laboratory. LLNL-ABS-830506*

**Title:** Microscale Characterization Tools for Algal-Bacterial Interactions and Cell Sinking

**Authors:** Hyungseok Kim[1]* (hskimm@mit.edu), Megan Morris[2], Ty Samo[2], Silvia Vidal Melgosa[3], **Rhona Stuart**[2], **Cullen R. Buie**[1], and **Xavier Mayali**[2]

**Institutions:** [1]Massachusetts Institute of Technology, Cambridge, MA; [2]Lawrence Livermore National Laboratory, Livermore, CA, [3]MPI for Marine Microbiology, Bremen, Germany

**Website URL:** https://bio-sfa.llnl.gov

**Project Goals:** Algal and plant systems have the unrivaled advantage of converting solar energy and $CO_2$ into useful organic molecules. Their growth and efficiency are largely shaped by the microbial communities in and around them. The μBiospheres SFA seeks to understand phototroph-heterotroph interactions that shape productivity, robustness, the balance of resource fluxes, and the functionality of the surrounding microbiome. We hypothesize that different microbial associates not only have differential effects on host productivity but can change an entire system's resource economy. Our approach encompasses single cell analyses, quantitative isotope tracing of elemental exchanges, 'omics measurements, and multi-scale modeling to characterize microscale impacts on system-scale processes. We aim to uncover cross-cutting principles that regulate these interactions and their resource allocation consequences to develop a general predictive framework for system-level impacts of microbial partnerships.

**Abstract Text:** Previous studies have documented that different algal associated bacteria influence the aggregation of algal cells in culture [1], which has implications for harvesting large-scale algal cultures for biofuels. Sinking aggregates, consisting of microbial cells including photosynthetic algae and heterotrophic bacteria, also are major contributors to biological carbon sequestration in aquatic environments [2], which could be used as an engineering solution to combat climate change [3]. Cells within these particles are physically connected via an exuded, sticky polymer called extracellular polymeric substances (EPS). For decades, it has been hypothesized that algal-bacterial interactions modulate the biosynthesis of EPS and leads to a change in formation and sinking rate of the aggregates [4]; however, its detailed mechanism remains unanswered due to a limited number of tools that characterize this microscale interaction.

Here we report a new phenomenon where the sinking rate of algal cells is influenced by the presence of a host-associated bacterial strain. Based on our observation that a cell culture stage (log or stationary phase) is the main driver that leads to different sinking rates, we develop a hypothesis that algal EPS production increases due to the presence of bacteria and this facilitates algal aggregation and increases cell sinking. To test this hypothesis, we employ several imaging-based characterization methods. First, we establish an optical setup that tracks vertical movement of single cells and compare sinking rates between different culture conditions. Second, we measure the surface charge of individual algal cells via electrokinetic microfluidic tools [5] and test if host-associated bacteria can influence algal surface charge. Third, we examine EPS

production from algae at both single-cell (using conjugate fluorescein that binds to a cell surface) and bulk (using polysaccharide extraction). Finally, as a proxy of nutrient exchange we report differential algal incorporation rate of carbon and nitrogen across bacterial taxa based on NanoSIMS [6] and compare it to the algal sinking speed under bacterial co-culture. Taken together, these characterization tools will provide a fundamental understanding on the biophysical role of microbes in microalgal aggregate dynamics critical for biofuel production and carbon sequestration.

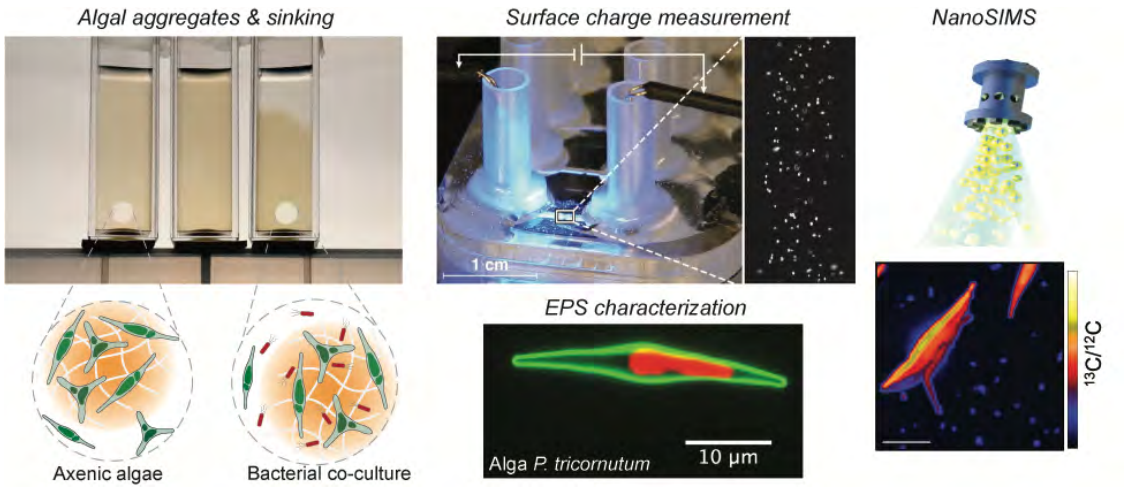



Figure. Imaging-based methods to characterize microscale interaction between algae and bacteria and their sinking.

**References/Publications**


[1] Gärdes, A., et al. (2011). "Diatom-associated bacteria are required for aggregation of Thalassiosira weissflogii." The ISME Journal 5(3): 436-445.

[2] Kim, H., et al. (2021). "Bacterial response to spatial gradients of algal-derived nutrients in a porous microplate." The ISME Journal.

[3] Raven, J. A. (2017). "The possible roles of algae in restricting the increase in atmospheric CO2 and global temperature." European Journal of Phycology 52(4): 506-522.

[4] Azam, F. and D. C. Smith (1991). Bacterial Influence on the Variability in the Ocean's Biogeochemical State: A Mechanistic View. Particle Analysis in Oceanography. S. Demers. Berlin, Heidelberg, Springer Berlin Heidelberg: 213-236.

[5] Wang, Q., et al. (2019). "Microfluidic dielectrophoresis illuminates the relationship between microbial cell envelope polarizability and electrochemical activity." Science Advances 5(1): eaat5664.

[6] Mayali, X. (2020). "NanoSIMS: Microscale Quantification of Biogeochemical Activity with Large-Scale Impacts." Annual Review of Marine Science 12: 449-467.



**Funding Statement:** *Part of this work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL microBiospheres SFA, FWP SCW1039.*

**Title:** A novel algicidal bacterium threatens diatom productivity and incorporates algal-derived carbon and nitrogen

**Authors:** Megan Morris (morris81@llnl.gov)[1]*, Jeffrey Kimbrel[1], Helena van Tol[1], Genevieve Parkey[2], Diana Morales[3], Patrik D'haeseleer[1], Ali Navid[1], **Rhona Stuart**[1], Peter K. Weber[1], and **Xavier Mayali**[1]

**Institutions:** [1] Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, CA, USA
[2] Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC, USA
[3] The Division of Physics, Mathematics and Astronomy, Caltech, Pasadena, CA, USA
* presenting author

**Website URL:** https://bio-sfa.llnl.gov/

**Project Goals: Algal and plant systems have the unrivaled advantage of converting solar energy and $CO_2$ into useful organic molecules. Their growth and efficiency are largely shaped by the microbial communities in and around them. The µBiospheres SFA seeks to understand phototroph-heterotroph interactions that shape productivity, robustness, the balance of resource fluxes, and the functionality of the surrounding microbiome. We hypothesize that different microbial associates not only have differential effects on host productivity but can change an entire system's resource economy. Our approach encompasses single cell analyses, quantitative isotope tracing of elemental exchanges, 'omics measurements, and multi-scale modeling to characterize microscale impacts on system-scale processes. We aim to uncover cross-cutting principles that regulate these interactions and their resource allocation consequences to develop a general predictive framework for system-level impacts of microbial partnerships.**

**Abstract Text:**

Photosynthetic algae, including phytoplankton, are globally significant $CO_2$ sinks and have high biotechnological potential for renewable biofuels and sustainable carbon sequestration. The heterotrophic bacteria that associate with microalgae in aquatic environments mediate biogeochemical cycling of algal-derived carbon and nitrogen, affecting the flow of nutrients to surrounding microbial cohorts and higher trophic levels. Through complex interactions and exchanges of secondary metabolites, bacteria can also influence algal productivity, either positively via mutualism or negatively via antagonism. Antagonistic algicidal bacteria lyse microalgal cells, diminishing populations and compromising or halting algal productivity. In the process, algal-sequestered nutrients are released that stimulate the metabolism of the surrounding microbial community, interfering with the system resource pool and economy. Catastrophic crashes within susceptible microalgal biofuel ponds are fairly common, and while these may be

attributed to biological entities such as algicidal bacteria, the causative agent is rarely investigated.

Here, we document a novel algicidal bacterium that lyses the biofuel-relevant diatom *Phaeodactylum tricornutum*, crashing laboratory cultures in days. While this novel bacterium is unculturable using traditional microbiological laboratory techniques, we use unique resources and molecular techniques to characterize the identity, strategy, and mechanism of this bacterium. Based on phylogeny of both a partial sequence obtained from amplicon sequencing and assembled 16S contig from a metagenome-assembled-genome (MAG), this is the first documented algicidal bacterium belonging to the taxonomic order Rickettsiales, which we name as a Rickettsia-like diatom killer, or "RLDK". Through super-resolution confocal microscopy and fluorescence in situ hybridizations using custom probes, we identified RLDK within a laboratory enrichment of a mixed community. Unlike other Rickettsiales pathogens, RLDK appears to have a free-living life stage and is capable of survival outside of host cells; however, most RLDK cells were found associated with the diatom cells, especially just prior to algal death. Metabolic predictions from the MAG also confirmed RLDK's genomic capacity for aggregation and attachment to host cells using pili. Further, the RLDK genome encodes a full type IV secretion system that may transport virulence factors and effector molecules extracellularly. In conjunction with peptidoglycanase-encoding genes, we hypothesize that RLDK's mechanism of attack is a series of events where bacterial cells attach to *P. tricornutum* and secretes host cell-degrading peptidoglycanases.

RLDK is unable to survive in pure culture free of the diatom host or other bacteria. Additionally, this bacterium has a reduced genome with numerous auxotrophies. Based on these two pieces of evidence, we hypothesized that RLDK is likely dependent on other entities for its growth and metabolism. To test this, we conducted a NanoSIP experiment and found that diatom-attached bacteria (RLDK) show enrichment of both isotopically labeled $^{13}C$ and $^{15}N$, neither of which RLDK can directly uptake or metabolize based on predictions from a genome-scale metabolic model. These results suggest that RLDK incorporates algal-derived carbon and nitrogen. Genome and metabolic model predictions will be tested using metatranscriptomes of both the bacterial parasite RLDK and algal host *P. tricornutum*. These data contribute to our understanding of poorly characterized algicidal bacteria in the environment and more broad understanding of algal-bacterial interactions. This research has implications for applied sciences, as our insights into the strategy and mechanism of algicidal bacterial activity may be leveraged to potentially prevent future catastrophic algal biofuel crash events.

*This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL Biofuels SFA, FWP SCW1039.*

**Title:** System-Level Analysis of Metabolism in a Novel Algicidal Bacterium

**Authors:** Helena van Tol,[1] Genevieve Parkey[2], Megan Morris[1], Jeffrey Kimbrel[1], **Rhona Stuart**[1], Patrik D'haeseleer[1], Xavier Mayali[1], **Ali Navid[1]*** (navid1@llnl.gov)

**Institutions:** [1]Physics & Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Georgetown University, Washington, DC

**Website:** https://bio-sfa.llnl.gov/

**Project Goals: Algal and plant systems have the unrivaled advantage of converting solar energy and $CO_2$ into useful organic molecules. Their growth and efficiency are largely shaped by the microbial communities in and around them. The µBiospheres SFA seeks to understand phototroph-heterotroph interactions that shape productivity, robustness, the balance of resource fluxes, and the functionality of the surrounding microbiome. We hypothesize that different microbial associates not only have differential effects on host productivity but can change an entire system's resource economy. Our approach encompasses single cell analyses, quantitative isotope tracing of elemental exchanges, 'omics measurements, and multi-scale modeling to characterize microscale impacts on system-scale processes. We aim to uncover cross-cutting principles that regulate these interactions and their resource allocation consequences to develop a general predictive framework for system-level impacts of microbial partnerships.**

Planktonic algae play an outsized role in regulating the dynamics of earth's ecosystem. Their activity affects the global oxygen supply, the food chain, and climate. Due to their ecological importance, as well as the industrial interest in using algal biomass for production of renewable biofuels, understanding the factors that control algal population dynamics are imperative for devising strategies to develop renewable fuels and lessen the harmful effects of climate change. Interactions of algae with bacteria are a major factor in the fate of algal populations.

As part of our system-level examination of microbial interactions with bioenergy-relevant phototrophs, we found that a previously uncharacterized Rickettsia-like diatom killing (RLDK) bacterium was crashing diatom *Phaeodactylum tricornutum* cultures while increasing in relative abundance in conjunction with algal decline. Unfortunately, RLDK cannot be isolated in pure culture. Additionally, our experimental and bioinformatic analyses have indicated that its mechanism of killing is different from other well studied algicidal organisms like *Kordia algicida*.

In cases where direct lab examination of an organism is infeasible, in silico analyses serve as a powerful alternative method of study. So, to examine the novel algicidal mechanism of RLDK, particularly the role that its metabolic characteristics might play in this transient behavior, we characterized its metabolism by developing a genome-scale model of its metabolism using a high-quality, near-complete metagenome-assembled genome (MAG).

We used several steps to generate a relevant genome-scale model of this novel bacterium. We initiated the model development process by using the new apps developed by members of our team for the DOE KBase platform to import and merge annotations of the RLDK MAG from a number of different sources. This because we have previously shown that combining annotations from multiple sources will provide us with a more complete annotation and subsequently metabolic network reconstruction[1]. We then used the KBase to generate a draft genome-scale model. TranSyt (a KBase app used to annotate and add transport reactions)[2] was used to identify new gene-protein-reaction associations and many new transport reactions were added.

Despite constraint-based system-level analyses showing that RLDK's algicidal activity does not stem from a need to scavenge any specific organic compound from algal remains, preliminary simulations and bioinformatic analyses point to several interesting metabolic characteristics. For example: 1) examination of the RLDK genome shows that it contains a Type IV secretion system, as well as a surprising number of peptidoglycan degrading enzymes which could point to the means by which it attacks the algae; 2) flux balance analysis of RLDK metabolism indicates that it uses a non-orthodox pathway for production of thiamine phosphates which results in production and export 4-hydroxybenzyl alcohol, a reduced form 4-hydroxybenzoate, a metabolite that our metabolomic studies indicate could have a significant role in algal-bacterial interactions; 3) RLDK metabolism of some compounds can result in production of superoxides, a practice of heterotrophic bacteria when they need to acquire trace metals or to prevent viral infections[3]. This suggests that the impetus for RLDK's algicidal activity may be collection of inorganic nutrients instead of organic resources. We are currently analyzing proteomics and transcriptomic data to ascertain RLDK metabolic changes that facilitate its diatom killing activity. This research has provided a suite of new hypotheses to guide our experiments, illustrating the value of systems-level in silico analysis to support experimental work, and uncovered new potential resources and metabolites which may govern our algal-bacterial interactions.

**References/Publications**


1 Griesemer, M., Kimbrel, J. A., Zhou, C. E., Navid, A. & D'haeseleer, P. Combining multiple functional annotation tools increases coverage of metabolic annotation. *BMC genomics* **19**, 948 (2018).

2 Lagoa, D. *et al.* TranSyT, the Transport Systems Tracker. *bioRxiv*, 2021.2004.2029.441738, doi:10.1101/2021.04.29.441738 (2021).

3 Diaz Julia, M. *et al.* Widespread Production of Extracellular Superoxide by Heterotrophic Bacteria. *Science* **340**, 1223-1226, doi:10.1126/science.1237331 (2013).



**Funding Statement:** *This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL μBiospheres SFA, FWP SCW1039.*

**Title: Nutrient Limitation Drives Dynamics of Host-virus Interactions**

**Authors:** Roya AminiTabrizi,[1*] (royaaminit@email.arizona.edu), Cristina Howard-Varona,[2] Natalie Solonenko,[2] Jane Fudyma,[1] Melissa Duhaime,[3] **Matthew Sullivan[2]**, and **Malak Tfaily[1,4]**
**Institutions:** [1]The University of Arizona, Tucson; [2]The Ohio State University, Columbus; [3]University of Michigan, Ann Arbor; and [4]Pacific Northwest National Laboratory, Richland, WA

**Project Goals:**

The overarching goal of this project is to establish ecological paradigms for how viruses alter soil microbiomes and nutrient cycles by developing foundational (eco)systems biology approaches for soil viruses. Here, we used a multi-omics approach to investigate phage-specific metabolic reprogramming in virus-infected cells (virocells) to build critically needed model systems and in-silico resources and tools extending to new soil model phage-host systems. Together, these efforts will establish a foundational ecological understanding for the soil microbiome while also developing essential tools and resources for transformative soil viral ecology advances. The development and application of –omics approaches will further help investigate microbial community processes involved in biogeochemical nutrient cycling in terrestrial ecosystems.

**Abstract:**

Viruses utilize infection to control bacteria that perform vital planetary functions by changing them into new entities called virocells that are reprogrammed to obtain energy and resources differently from uninfected cells[1]. Since microbial metabolic outputs dictate ecosystem-level biogeochemical processes, and virocells are fundamentally reprogrammed metabolically, their interaction with the surrounding environment is expected to be different. Here we used a known, ecologically relevant bacterium (*Pseudoalateromonas*) and two unrelated infecting phages (HP1, a podophage, 45.06 Kbp dsDNA, and HS2, a siphophage, 37.72 Kbp dsDNA)[2] under phosphorus (P) rich and poor conditions to see whether infection dynamics are the same or different under the different nutrient types and to develop foundational approaches to studying soil viruses. Using an integrated multi-omics approach, we found that under nutrient-rich conditions, differences in the phage mechanism of action were driven by the metabolic reprogramming of the host, resulting in two different virocells with different energy and resource acquisition strategies[3]. In contrast, under nutrient limitation, we found that the environment quite strongly drove the metabolic reprogramming of the host, where the common strategy across different phages was survival under stress. These findings contribute to a better biological knowledge of phage-host interactions and the effects of nutrient constraints on their dynamics, shedding light on the environment-specific impacts on virocells that could significantly affect the metabolic reprogramming of the host and thus reshape the biogeochemical cycles in various ecosystems. Our research reveals significant insights into phage-host interactions and offers methods for new soil model phage-host systems.

**References/Publications**

1. Correa, A. M., Howard-Varona, C., Coy, S. R., Buchan, A., Sullivan, M. B., & Weitz, J. S. (2021). Revisiting the rules of life for viruses of microorganisms. Nature Reviews Microbiology, 1-13.
2. Duhaime, M. B., Solonenko, N., Roux, S., Verberkmoes, N. C., Wichels, A., & Sullivan, M. B. (2017). Comparative omics and trait analyses of marine Pseudoalteromonas phages advance the phage OTU concept. *Frontiers in microbiology*, *8*, 1241.

3. Howard-Varona, C., Lindback, M. M., Bastien, G. E., Solonenko, N., Zayed, A. A., Jang, H., ... & Duhaime, M. B. (2020). Phage-specific metabolic reprogramming of virocells. *The ISME journal*, *14*(4), 881-895.

**Funding statement:**
This work was funded by the DOE Genomic Science Program award number DE-SC0020173. Portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative through separate awards to MT and MS and used resources at the Environmental Molecular Sciences Laboratory and Joint Genome Institute, two DOE Office of Science User Facilities.

# Viral ecogenomics across a permafrost thaw gradient

Christine L. Sun[1,2,3*], Akbar Adjie Pratama[1,2,3*] (pratama.3@osu.edu), Maria Consuelo Gazitúa[4], Ahmed A. Zayed[1,2,3], Dylan Cronin[1,2,3], Benjamin Bolduc[1,2,3], Bridget B. McGivern[3,5], Lindsey Solden[1], Dean R. Vik[1,2,], IsoGenie Project Field Teams 2010-2017, IsoGenie Project Coordinators, Kelly C. Wrighton[3,5], Virginia I. Rich[1,2,3], Matthew B. Sullivan[1,2,3]

[1]The Ohio State University, Columbus, Ohio, USA; [2]Center of Microbiome Science Columbus, Ohio, USA; [3]EMERGE Biology Integration Institute, Columbus, Ohio, USA; [4]Viromica Consulting, Santiago, Chile. [5]Colorado State University, Fort Collins, CO, USA; [*]Authors contributed equally to this work.

**Project goals:** The overarching goal of this project is to establish ecological paradigms for how viruses alter soil microbiomes and nutrient cycles by developing foundational (eco)systems biology approaches for soil viruses. Within this overall project, we examined viruses from metagenomic datasets to investigate viral ecology across a nearly decadal time series. Such work is critical for establishing baseline ecological understanding of viruses in soils, as well as elucidating the role of viruses in terrestrial carbon and nutrient cycling.

Since permafrost accounts for 30% - 50% of global soil carbon (C), it is important to understand how thawing permafrost will impact the release of greenhouse gasses. Microorganisms are known to play critical roles in the terrestrial C cycle, but their viruses are less well-understood. In marine systems, viruses impact C cycling by controlling hosts via predation, transferring genes between hosts, and metabolically reprogramming hosts via encoding auxiliary metabolic genes (AMGs). Here we identify and examine viruses from a metagenomic dataset composed of peat samples collected between 2010 - 2017 across the active layer of a permafrost thaw gradient at a northern Sweden long-term ecological research site. This augmented known viruses at this already intensively studied site approximately 2-fold. The vast majority of viruses were novel, even when compared to other viruses from terrestrial ecosystems. Viral communities differentiated based on thaw stage (palsa, bog, and fen) and on depth, but not by time (year). Leveraging co-sampled metagenome-assembled genomes (MAGs) to *in silico* predict microbial hosts, we found that the majority of predicted hosts had genome-encoded functions involving the degradation of plant-derived C polysaccharide (such as cellulose, fructose, and xylose), which implicated viruses in soil C cycling. Further support for this came from identifying a myriad of virus-encoded AMGs, including those involved in central C metabolism and C degradation (e.g. glycoside hydrolases). Together, these results provide a growing baseline ecological understanding of soil viruses and increasingly implicate them in permafrost C cycling.

***Funding statement**: This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Numbers DE-SC0020173 (to MBS), DE-SC0004632, DE-SC0010580, and DE-SC0016440 (to VIR), as well as the National Science Foundation #1758974 (to MBS) and #2022070 (to VIR).*

# Corrinoids, a Class of Model Metabolites, Differentially Impact Isolation of Soil Bacteria

Zoila I. Alvarez-Aponte[1*] (zoila.alvarez@berkeley.edu), Zachary F. Hallberg,[1] Alexa M. Nicolas,[1] Myka A. Green,[1] Eoin L. Brodie,[2] Mary K. Firestone,[1]Hans K. Carlson,[1,2] and **Michiko E. Taga[1]**

[1] University of California, Berkeley; [2] Lawrence Berkeley National Laboratory, Berkeley, CA

https://tagalab.berkeley.edu

**Project Goals: The overall goal of this project is to gain a deeper understanding of the microbial interactions that drive soil community structure. This research leverages a model group of metabolites related to vitamin $B_{12}$, known as corrinoids, to investigate microbial interactions. Corrinoids are a structurally diverse nutrient class shared among microbes, as they are produced by only a subset of the bacteria that require them. Based on the inherent preferences that bacteria display for specific corrinoids, the hypothesis driving this work is that corrinoids are keystone nutrients and can shape soil microbial communities. To test this hypothesis, we are examining the effects of corrinoid addition on community composition and function across multiple levels of complexity. By investigating the impact of a key nutrient in soil at levels spanning the whole community to individual isolates, this work will reveal an unprecedented view of metabolic interactions in soil microbial communities.**

Soil microbial communities impact numerous global processes, from the global carbon cycle to agriculture and human health. The effect of microbes on the processes they modulate is greatly dependent on the composition of microbial communities and nutrient availability. Thus, to understand the global role of microbes, we must take a closer look at the interactions that determine community structure, function, and dynamics. However, the physicochemical and taxonomic complexity of the soil microbiome poses a challenge to disentangling these interactions. By focusing on corrinoids as model metabolites, we aim to mechanistically study complex metabolic interactions between soil microbes.

In this work, we isolated bacteria from soil on media amended with different corrinoids to address two hypotheses. First, because bacteria are known to have distinct preferences for specific corrinoids, we hypothesize that corrinoids influence culturability, and we expect to isolate distinct taxa on each corrinoid condition. Second, because corrinoids other than $B_{12}$ have never been used in growth media, species that prefer them have likely remained uncultured, leading to a pervasive bias across microbiology.

By using seven different corrinoid conditions, we have isolated an even and diverse collection of 243 isolates, including representatives from seven phyla. We predict at least 21 of

these isolates to be novel based on full 16S rRNA amplicon sequence similarity, and are in the process of characterizing these isolates further. Preliminary data suggest that one of the novel isolates prefers [5-OHBza]Cba, a corrinoid known for being produced by methanogens, over $B_{12}$. This preference has rarely been observed in bacteria. This observation is part of the diversity of corrinoid metabolisms we have observed in this set of isolates, which include corrinoid production, auxotrophy, and independence. This isolate collection and the methods we are developing will be crucial for future steps in this research, which include testing corrinoid sharing interactions in co-cultures.

*This research is funded by the DOE Genomic Sciences Program Grant DE-SC0020155 (to MET).*

# Impact of Corrinoids on Soil Community Assembly in a Grassland Soil

Zachary F. Hallberg (zhallber@berkeley.edu),[1] Zoila I. Alvarez-Aponte,[1] Alexa M. Nicolas,[1] Mary K. Firestone,[1,2] Hans K. Carlson,[2] and **Michiko E. Taga[1]**

[1]University of California, Berkeley
[2]Lawrence Berkeley National Laboratory, Berkeley, CA

http://tagalab.berkeley.edu

**Project Goals: The overall goal of this research is to gain a deeper understanding of the microbial interactions that drive soil community structure. This research leverages a model group of key metabolites related to vitamin $B_{12}$, known as corrinoids, to investigate microbial interactions. Corrinoids are a structurally diverse nutrient class shared between different bacterial species, as they are produced by only a subset of the bacteria that use them. Based on the inherent specificity of bacteria for particular corrinoids, the hypothesis driving this work is that corrinoids are keystone nutrients in shaping soil microbial communities. To test this hypothesis, we are examining the effects of corrinoid addition on community composition and function across multiple levels of complexity. By investigating cycling of a key nutrient in soil at levels spanning the whole community to individual isolates, this work will reveal an unprecedented view of metabolic interactions in a soil microbial community.**

Nutritional interactions in the soil microbiome drive critical environmental and ecological processes including biogeochemical cycling and plant growth. Identifying key nutrients that could predictably modulate soil communities is an attractive way to impact these processes. Nevertheless, the complexity of the soil milieu and diverse microbial community composition make it impossible to experimentally analyze all microbe-microbe interactions simultaneously using current analytical methods. We aim to identify key microbe-microbe interactions by unraveling the nutrient network produced by corrinoids, a group of structurally diverse metabolites used by a majority of the community. Exemplified by its flagship member, Vitamin B12, corrinoids are essential cofactors produced only by a fraction of the bacteria that use them, and thus are shared metabolites. Unlike other shared nutrients, corrinoids are structurally diverse: at least 16 distinct corrinoids have been described. Our group and others have shown that bacteria that play central roles in bioremediation, plant-bacteria symbiosis, and elemental cycling possess preferences for distinct corrinoids. Thus, corrinoids appear to be an ideal nutrient by which we can pinpoint specific microbial interactions.

Here, we test the hypothesis that corrinoids are a key nutrient family that shape soil microbial ecosystems, by determining the impact of corrinoids on soil-derived enrichment community assembly. We find that vitamin B12 is the primary corrinoid present in grassland

soils. Reflecting this, B12 addition causes marked changes to enrichment culture assembly over extended time periods. Further inquiry demonstrates that inoculation conditions drastically alter the corrinoid dependence of enrichment cultures. Enrichment communities at high inoculation levels remain stable to nutrient shifts, whereas low inoculation levels begin to impart stochastic effects. Our work identifies a subset of corrinoid-limited taxa that can be leveraged to understand nutritional networks in soils, and empowers the future study of *in vitro* soil-derived enrichment communities to further understand nutrient sharing.

*This research is funded by the DOE Genomic Sciences Program Grant DE-SC0020155 (to MET).*

**Microbial controls on biogeochemical cycles in permafrost ecosystems**

**Neslihan Taş***, Marla DeKlotz, Nancy Conejo, Megan Dillon, Yaoming Li and Hoi-Ying Holman

Lawrence Berkeley National Laboratory

**Project Goals: The permafrost carbon reservoirs are currently protected from microbial decomposition by frozen conditions. Upon permafrost thaw microbial metabolism leads to decomposition of soil organic matter, substantially impacting the cycling of nutrients and significantly affecting the Arctic landscape. This project use state of the art molecular techniques to resolve complex microbial processes governing the biogeochemical cycles in arctic soils and permafrost to better inform efforts to access uncertainties surrounding ecosystem responses.**

Permafrost microbial communities are complex, diverse, and active at subzero temperatures. While carbon turnover at depth is proposed to be slower than surface, especially the fate of carbon in deep permafrost, which is currently protected from the warming climate, is uncertain. Permafrost microbiome is a seed bank of mostly novel organisms that have a diverse and broad metabolic potential. The microbial response to thaw in arctic environments is not uniform and the relationship between permafrost microbiomes and greenhouse gas emissions is not well understood. Following thaw, redistribution of water is a key event that conditions the permafrost for microbial decomposition. We initiated batch-scale permafrost incubation experiments dry, natural, and saturated moisture states and under microaerophilic or anaerobic headspaces. We couple omics methods with analysis of soil chemistry via synchrotron Fourier transform infrared spectral imaging at the Berkeley Infrared Structural Biology beamline of the Advanced Light Source. Analysis showed that variety of organic compounds and metabolites were accumulated in thawed permafrost soils. Especially under saturated conditions while carbohydrates were depleted, dry soils accumulated aliphatic compounds. We found strong trends that under dry conditions *Dormibacteria* and *Chloroflexi* were able to survive over multiple years and retain carbon stocks as carbohydrates and microbial biomass. In contrast saturated conditions gave rise to *Actinobacteria*. This project use field observations, laboratory manipulations, and multi-omics approaches to examine how microbial processes and hydrology interact during permafrost thaw Alaska in order to determine how these factors drive biogeochemical cycles in different arctic soils.

*This research is supported by DOE Early Career Program by the Office of Biological and Environmental Research in the DOE Office of Science.*

**MetaboDirect: An analytical pipeline for FT-ICR mass spectrometry data**

Christian Ayala-Ortiz,[1] Nathalia Graf-Grachet,[1] Viviana Freire-Zapata,[1] Jane Fudyma,[1] Gina Hildebrand,[1] Roya AminiTabrizi,[1] Cristina Howard-Varona,[2] Yuri E. Corilo,[3] Nancy Hess,[3] Melissa Duhaime,[4] Matthew Sullivan,[2,5] **Malak Tfaily**,[1,3*] (tfaily@arizona.edu)

[1] Department of Environmental Science; University of Arizona, Tucson, AZ, 85721, USA

[2] Department of Microbiology, and Center of Microbiome Science; Ohio State University, Columbus, OH, 43210, USA

[3] Environmental Molecular Sciences Laboratory; Pacific Northwest National Laboratory, Richland, WA, 99354, USA

[4] Department of Ecology and Evolutionary Biology; University of Michigan, Ann Arbor, MI, 48109, USA

https://github.com/Coayala/MetaboDirect

**Project Goals**: Advances in different -omics technologies have revolutionized biological research by enabling high-throughput monitoring of biological processes at the molecular level and their responses to environmental perturbation. Metabolomics is a fast-emerging technology in systems biology that aims to profile small compounds within a biological system that are often end products of complex biochemical cascades. Thus, metabolomics can enable discovery of the genetic basis of metabolic variation by linking the genotype to the phenotype. Despite increasing accessibility of multi-omics technologies, integration of multi-omics data in analysis pipelines remains a challenge especially in the environmental field. In addition, there are still many associated bottlenecks to overcome in metabolomics before measurements will be considered robust. The overarching goal of this proposal is to optimize the analysis of complex and heterogeneous biological and environmental datasets by developing a user-friendly, open-source metabolomics data analysis pipeline that is integrable with other multi-omics data sets.

Microbiomes are now recognized as driving ecosystem functions ranging from the oceans and soils to humans and bioreactors. However, a grand challenge in microbiome science is to characterize and quantify the chemical currencies of organic matter that microbes respond to, produce and alter. Critical to this has been the development of Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), which has drastically increased molecular characterization of complex organic matter samples, but challenges users with hundreds of millions of data points where readily available, user-friendly, and customizable software tools are lacking. Here we build on years of analytical experience with diverse sample types to develop MetaboDirect, a comprehensive, open-source, command-line based pipeline for direct injection FT-ICR MS data analysis, as well as visualizing the resulting molecular formula assignments. MetaboDirect requires only to run a single line of code that launches a fully automated framework for the generation and visualization of organic matter transformation networks within few minutes. For more experienced users (python savvy), MetaboDirect allows users to customize plots, pipeline outputs, and analyses. Application of MetaboDirect to FT-ICR MS experimental data – from each a marine phage-bacterial infection and a *Sphagnum* leachate microbiome incubation – showcase the data discovery and exploration capabilities of the pipeline

that we hope will enable the FT-ICR MS research community to evaluate and interpret their data in greater depth and in less time.

**Funding Statement**: This work was funded by the DOE SC program in Biological and Environmental Research (BER) award number DE-SC0021349. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative through an award to MT and used resources at the Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility.

**Understanding Robustness of *Yarrowia lipolytica* for Undetoxified Biomass Hydrolysate Utilization**

Seunghyun Ryu[1*] ( sryu2@utk.edu), Caleb Walker,[1] Stephanie Thompson[2], Richard J. Giannone[3], Patricia J. Slininger[2], Bruce S. Dien[2*], (bruce.dien@usda.gov) and **Cong T. Trinh**[1*] (ctrinh@utk.edu)

[1]University of Tennessee, Knoxville; [2]National Center for Agricultural Utilization Research, USDA-ARS, Peoria, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN;

**Project Goals: To elucidate and harness the exceptional robustness of novel, undomesticated *Y. lipolytica* isolates from genetic diversity screening as a bioenergy-relevant microbial platform for efficient conversion of undetoxified biomass hydrolysates into designer bioesters with continuous recovery using solvent extraction.**

**Abstract text.** Robustness is an important phenotype for bioenergy microbes to acquire but is difficult to engineer. The oleaginous yeast, *Yarrowia lipolytica*, is an exceptionally robust microbe that can naturally tolerate stressful environments, assimilate a wide range of substrates, and produce high-value chemicals. In this study, we aim to understand and harness the robust characteristics of *Y. lipolytica* for utilizing mixed C5 and C6 sugars in undetoxified switchgrass biomass hydrolysates. From a screen of 57 undomesticated *Y. lipolytica* isolates, we selected top-performing strains exhibiting robust growth and lipid accumulation in biomass hydrolysates and subjected them to comprehensive growth, lipid production, and multi-omics characterization. We then engineered these top performing strains through a combination of targeted pathway engineering followed by adaptive laboratory evolution to further enhance their inherent robustness. Multi-omics profiling of these robust *Yarrowia* strains revealed novel genotype, regulation and cellular processes that positively impact their utilization of biomass hydrolysate sugars. These foundational studies provide key insight into the genetic mechanisms responsible for robustness in *Yarrowia* strains, enabling targeted engineering strategies to be deployed for enhanced production of biofuels and bioproducts from lignocellulosic biomass.

**References**

1 Walker, C., Dien, B., Giannone, R.J., Slininger, P., Thompson, S.R., **Trinh, C.T.***, 2021. Exploring Proteomes of Robust *Yarrowia lipolytica* Isolates Cultivated in Biomass Hydrolysate Reveal Key Processes Impacting Mixed Sugar Utilization, Lipid Accumulation, and Degradation, ***mSystems***, **6**(4): e00443-21.

2. Walker, C., Ryu, S. & Trinh, C. T. Exceptional solvent tolerance in *Yarrowia lipolytica* is enhanced by sterols. *Metabolic Engineering* **54**, 83-95, doi:https://doi.org/10.1016/j.ymben.2019.03.003 (2019).

3. Ryu, S. & Trinh, C. T. Understanding Functional Roles of Native Pentose-Specific Transporters for Activating Dormant Pentose Metabolism in *Yarrowia lipolytica*. *Applied and Environmental Microbiology* **84**, doi:10.1128/aem.02146-17 (2018).
4. Walker, C. *et al.* Draft Genome Assemblies of Five Robust Yarrowia lipolytica Strains Exhibiting High Lipid Production, Pentose Sugar Utilization, and Sugar Alcohol Secretion from Undetoxified Lignocellulosic Biomass Hydrolysates. *Microbiol Res Announc* **7**, e01040-01018 (2018).
5. Quarterman, J., Slininger, P. J., Kurtzman, C. P., Thompson, S. R. & Dien, B. S. A survey of yeast from the Yarrowia clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate. *Applied microbiology and biotechnology* **101**, 3319-3334, doi:10.1007/s00253-016-8062-y (2017).
6. Ryu, S., Labbé, N. & Trinh, C. T. Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of α-ketoglutaric acid by Yarrowia lipolytica. *Applied microbiology and biotechnology* **99**, 4237-4244, doi:10.1007/s00253-015-6521-5 (2015).
7. Ryu, S., Hipp, J. & Trinh, C. T. Activating and Elucidating Metabolism of Complex Sugars in *Yarrowia lipolytica*. *Applied and Environmental Microbiology* **82**, 1334-1345, doi:10.1128/aem.03582-15 (2016).

*Funding statement. This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DESC0019412.*

**Controlling Selectivity of Modular Microbial Biosynthesis of Designer Acetate Esters through Proteome Reallocation**

Authors: Hyeongmin Seo[1], Jong-Won Lee[1], Seunghyun Ryu[1], Richard J. Giannone[2], Bruce S. Dien[3], and Cong T. Trinh[1] (ctrinh@utk.edu)

**Institutions:** [1]University of Tennessee, Knoxville, TN, USA; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]National Center for Agricultural Utilization Research, USDA-ARS, Peoria, IL

**Project Goals:** To fundamentally understand and redirect metabolism and regulation of thermophilic *Bacillus coagulans* for the efficient conversion of undetoxified lignocellulosic biomass hydrolysates into designer bioesters.

**Abstract text:** Short-chain esters have broad utility as ingredients for flavors, fragrances, solvents, and drop-in biofuels. Biologically, these esters are derived from the condensation of acyl CoAs and alcohols from cellular metabolism, resulting in a large portfolio of ester molecules. However, controlling the selectivity of microbial ester biosynthesis has remained challenging from a metabolic engineering standpoint. Here, we present a generalizable framework for the *de novo* biosynthesis of short-chain designer bioesters (i.e., n-butyl acetate, isobutyl acetate, and isoamyl acetate) through microbial fermentation with controllable selectivity from renewable feedstocks. Using modular design principles, we propose to design and package efficient ester production pathways into exchangeable ester modules compatible with an engineered chassis cell. Successful implementation of our strategy required manipulation of replication, transcription, (post)translation, and enzyme specificity to control proteome reallocation for designer ester biosynthesis. By coupling the exchangeable ester modules with the chassis cell(s), we demonstrated the assembled production strains exhibited enhanced production of target esters with high selectivity.

**References/Publications**

1. Lee, J., **Trinh, C.T.***, 2021. Controlling Selectivity of Modular Microbial Biosynthesis of Butyryl-CoA-Derived Designer Esters**,** ***Metab Eng****,* **69:** 262-274.
2. Lee, J., Seo, H., Young, C.[§], **Trinh, C.T***, 2021. Probing Specificities of Alcohol Acyltransferases for Designer Ester Biosynthesis with a High-Throughput Microbial Screening Platform, ***Biotechnol Bioeng***, 1-13.
3. Seo, H., Lee, J., Dunlap, N.[§], Giannone, R., **Trinh, C.T**.*, 2021. Repurposing chloramphenicol acetyltransferase for a robust and efficient designer ester biosynthesis platform, ***Metab Eng***, **66:** 179-190.
4. Lee, J., **Trinh, C.T.***, 2020. Towards renewable flavors, fragrances, and beyond, ***Curr Opin Biotechnol,*** **61**: 168-180.
5. Lee, J., **Trinh, C.T.***, 2019. Microbial Biosynthesis of Lactate Esters, ***Biotechnol Biofuels***, **12**(1):226.

**Funding statement:** *This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0022226.*

**Title:** Employing Bacterial Microcompartments To Create Privileged Redox Pools for Biofuel Production

**Authors:** Carolyn E. Mills[1*] (cemills@northwestern.edu), Charlotte H. Abrahamson[1] (cabrahamson@u.northwestern.edu), Sasha Shirman[1] (aleksandra.shirman@northwestern.edu), Md Aminul Islam Prodhan[1] (md.prodhan@northwestern.edu), Svetlana P. Ikonomova[1,] Andre G. Archer[1*], Taylor M. Nichols[1], Brett J. Palmero[1], Keith E.J. Tyo,[1] Niall Mangan,[1] and **Danielle Tullman-Ercek**[1]

**Institutions:** [1]Northwestern University

**Website URL:** https://dtelab.northwestern.edu/research/#

**Project Goals:** To compartmentalize metabolic pathways along with enzyme cofactor recycling pathways to increase the yield and efficiency of bioproduction processes

**Abstract Text:** Metabolic engineering holds great promise for creating efficient, competitive routes for the production of biofuels and biochemicals without the necessity for harsh chemicals and hazardous byproducts. Successes in biochemical engineering include Dupont's Sorona fiber, which is made using bacterially-produced 1,3-propanediol from glucose. However, roadblocks to biosynthesis prevent many biochemicals from being produced biologically given current technology. Nature uses compartmentalization (e.g. in organelles in eukaryotes and in bacterial microcompartments in prokaryotes) to solve issues such as intermediate leakage, toxicity, and byproduct formation. We proposed to deploy compartmentalization as a strategy to overcome a critical roadblock: the requirement for redox cofactor recycling. We set out to directly demonstrate the redox recycling feature of a bacterial microcompartment (MCP) for the first time, and apply it to 1,3-propanediol production from glycerol. With this poster, we will describe how we have leveraged modeling to both inform strain design and provide fundamental insight into MCP function.

We first studied the native 1,2-propanediol (1,2-PD) utilization (Pdu) pathway to develop understanding of the principles that govern cofactor recycling in MCPs. We investigated the cofactor recycling thought to occur in the native 1,2-PD degradation pathway by running *in vitro* assays on purified Pdu MCPs. We found that while external cofactor addition is required for Pdu pathway activity, cofactor recycling enabled the pathway production to exceed stoichiometric cofactor concentrations. These findings suggest that the native system benefits from cofactor recycling but is also somewhat permeable to cofactors. Mathematical modeling of the native pathway further informed the conditions tested in these *in vitro* experiments, allowing precise fitting for parameters that govern MCP performance, such as MCP permeability. Lastly, alternative morphologies of the Pdu MCP were investigated using the native pathway *in vivo* and *in silico* for use as alternative scaffold structures.

Our insights from the native system were used to guide modeling and strain development to advance our goal of encapsulating the 1,3-propanediol (1,3-PD) production pathway. To optimize stoichiometry for 1,3-PD enzymes, we built a model to predict toxic intermediate build-up and product formation in the pathway. We quantified parametric uncertainty based on existing data, and optimized enzyme ratio in compartments based on the resulting range of intermediate and product dynamics. The expression and encapsulation of the 1,3-PD pathway enzymes *in vivo* were tuned to achieve the optimal stoichiometry predicted by the mathematical model. Combined, this work will facilitate future use of MCPs for addressing issues with pathways limited by imbalances in redox cofactors.

**Funding Statement:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019337.

**Engineering bacterial microcompartments in *Clostridium autoethanogenum* to overcome bottlenecks in sustainable production of synthetic rubber**

Alexander P. Mueller,[1] Fungmin (Eric) Liew,[1] Shilpa Nagaraju,[1] Michael Köpke,[1] and **Danielle Tullman-Ercek**[2]

[1]LanzaTech, Skokie, Illinois
[2]Northwestern University, Evanston, Illinois

https://dtelab.northwestern.edu/research/#nanobioreactors

**Project Goals: To investigate bacterial microcompartments in *C. autoethanogenum* and engineer them to compartmentalize synthetic metabolic pathways.**

One promising route to sustainable bioproduction of fuels and chemicals is the engineering of organisms such as acetogens to efficiently convert abundant and low-cost carbon monoxide (CO) or carbon dioxide ($CO_2$) and hydrogen ($H_2$) containing gases to desirable products at high efficiency and low cost. This approach not only provides an avenue for repurposing greenhouse gases (GHG), but also minimizes the necessity for harsh chemicals and hazardous byproducts common in petroleum-based processes. However, many biochemicals are not yet produced biologically due to roadblocks in the cellular biosynthesis process. These roadblocks can include toxicity of intermediates, redox imbalances, and/or loss of product to off-pathway reactions. Nature uses spatial organization strategies, such as sequestration in organelles, to alleviate these issues. In bacteria, organization occurs in protein organelle-like structures known as bacterial microcompartments.

We will investigate the native regulation, assembly, and function of microcompartments in the industrially relevant non-model host *Clostridium autoethanogenum*. In the *C. autoethanogenum* genome, two unique clusters (Pdu and Cut) that resemble MCP operons, including the presence of putative hexamers, trimers, pentamers, and enzyme encapsulation sequences have been identified. RNAseq data showed functional genes in the Pdu cluster to be significantly upregulated ($p<0.001$) in the presence of specific substrates compared to control conditions. These findings are corroborated by electron microscopy of *C. autoethanogenum* grown in the same conditions, which shows distinctive polyhedral shapes within the cells indicative of MCP formation.

Our goal is to sequester key biosynthesis enzymes from two distinct metabolic pathways into microcompartments to make compounds involved in rubber production to showcase the power of the strategy for reducing toxicity and product losses due to side reactions. We will also couple modeling with experiments to understand the native system and identify the most promising targets for compartmentalization. If successful, this work would 1) provide insight into the native function of these structures in this organism, 2) be the first direct demonstration of this feature of a bacterial microcompartment in a non-model organism, and 3) would provide a detailed method for repeating this success in other organisms and with other pathways. Ultimately, this will lead to the cost-efficient production of chemicals that are currently derived from petroleum.

*This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0022180.*

# Modeling and NMR Methods to Probe Spatial Arrangements in Biomolecules: Towards predictive models of plant cell wall structure

Bennett Addison*[1,2] (bennett.addison@nrel.gov), Vivek Bharadwaj[1,2], Anne E. Harman-Ware[1,2], Yannick Bomble[1,2] and **Gerald A. Tuskan**[2]

**Institutions:** [1]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN.

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

The lignified plant secondary cell wall is comprised of a complex and heterogeneous framework of three major biopolymers; cellulose, hemicelluloses (xylan and glucomannan) and polyaromatic lignin. The intrinsic physical properties of these polymers and the complicated, yet poorly understood, interplay between them, gives rise to a lignocellulosic material that is not only structurally and mechanically sound but also frustratingly recalcitrant to enzymatic and chemical deconstruction methods. Due to inherent limitations of many common analytical techniques applied to heterogeneous materials, detailed molecular-level information on secondary cell wall (SCW) structure and architecture, especially regarding interactions between the constituent polymers, remains scarce. As a result, even the best available SCW models in literature are limited to being mere conceptual renderings rather than acting as frameworks for enhancing our scientific understanding of the role played by molecular-level actors in contributing to emergent properties.

A combined experimental and computational approach is being developed to help bridge this gap. Experimentally, we employ solid state Nuclear Magnetic Resonance (ssNMR) techniques, including methods developed in house [1], to probe polymer-polymer interactions within the secondary cell wall of $^{13}$C-enriched poplar wood in great detail. The major ssNMR observables that informed the construction and validation of the SCW atomistic molecular models of poplar include the following: 1) cellulose microfibrils are on the order of 3 nm in diameter with tightly bound water and possibly some acetylated xylan trapped between; 2) spin-diffusion rate constants and loose inter-polymer distance estimates confirm that xylan is “in the middle” between lignin and cellulose and that xylan interacts with cellulose in a 2-fold arrangement with its decorations pointing away from the cellulose surface; 3) the inter-polymer distance between xylan acetyl groups and the lignin surface is ~0.3 – 0.5 nm; 4) ~80% of all lignin and ~40% of cellulose chains are within ~1 nm of xylan acetyl carbons; 5) ~60% of all xylan and ~20% of all cellulose carbons are within ~1 nm of lignin ring carbons; and 6) lignin and xylan are most likely of linear morphology, not globular, and lignin/xylan heterogeneities are generally not larger than ~1 nm in size.

Even for a known polymer composition (*e.g.,* 50% cellulose, 25% lignin, 25% hemicelluloses), the construction of molecular models of these complex systems entails the consideration of a number of variable factors such as the relative locations of xylan, lignin and water with respect to cellulose. Quantitative observables from molecular dynamics (MD) simulations (*e.g.*: radial distribution functions, polymer-polymer distances and conformational analyses) of models built with varying arrangements of these components enables the corroboration of ssNMR inferences and lays the foundations for the development of realistic plant cell wall architectures with atomistic details. Here we demonstrate how ssNMR data has enabled the iterative development of these atomistic SCW models leading to the most detailed molecular picture of the plant cell wall architecture to date. Our approach of building atomistic models with varied spatial arrangements of constituent biopolymers, benchmarked with ssNMR data presents a robust protocol for the development of realistic and holistic plant cell wall models. These ssNMR-informed and experimentally-validated atomistic models lay the foundations for myriad future *insilico* explorations to gain insights into the molecular level determinants of the emergent properties of naturally occurring, treated and engineered biomass.

**References/Publications**

1. Addison, B., et al., *Selective One-Dimensional 13C-13C Spin-Diffusion Solid-State Nuclear Magnetic Resonance Methods to Probe Spatial Arrangements in Biopolymers Including Plant Cell Walls, Peptides, and Spider Silk.* The Journal of Physical Chemistry B, 2020: p. acs.jpcb.0c07759.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Techno-economic analysis and life cycle assessment of a biorefinery utilizing reductive catalytic fractionation

Andrew W. Bartling*[1,2] (andrew.bartling@NREL.gov), Michael L. Stone[2,3], Rebecca J. Hanes[1,2], Mary J. Biddy[1,2] Ryan Davis[1], Jacob S. Kruger[1], Nicholas E. Thornburg[1], Yuriy Román-Leshkov[2,3], Gregg T. Beckham[1,2] and **Gerald A. Tuskan**[2]

[1]National Renewable Energy Laboratory, Golden CO; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN; [3]Massachusetts Institute of Technology, Cambridge MA.

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

The aromatic polymer, lignin, which comprises a major fraction of carbon in plants, remains a challenge to cost effectively and sustainably convert to valuable products, despite a century of microbial, genomic, and catalytic research. While potential bench-scale routes to convert lignin-derived compounds into both fuels and valuable chemicals exist, industrially viable methods to depolymerize lignin into usable fragments conducive to these valorization routes remains a subject of ongoing research and development. Reductive catalytic fractionation (RCF) is one promising approach to fractionate lignocellulosic biomass and convert lignin into a narrow slate of products amenable to upgrading[1]. To quantitatively guide research towards critical areas for commercialization, cost and sustainability must be considered. To that end, we report a detailed techno-economic analysis (TEA) and life cycle assessment (LCA) of the RCF process, wherein biomass carbohydrates are converted to ethanol and the lignin-rich RCF oil is the lignin-derived product. We first evaluate a process configuration using methanol as a solvent and $H_2$ as a hydrogen source which predicts a minimum crude RCF oil selling price of $0.51/lb when the coproduct ethanol is sold at $2.50/gallon of gasoline equivalent. When normalized to just the lignin content of the RCF oil, the minimum selling price of the lignin fraction is $0.79/lb. Analysis of additional cases using different solvents and an *in situ* hydrogen donor from hemicellulose-derived compounds revealed that limiting reactor pressure using solvents with lower vapor pressure could greatly reduce capital expenses while still maximizing lignin yields and exhibiting promising economics and environmental impacts. Process configurations that reduce the energy demand for solvent separation also improve both global warming potential (GWP) and cumulative energy demand (CED) through reducing natural gas demand. This study suggests prioritization of research

that can reduce capital expenses and environmental impacts by lowering RCF operating pressure, minimizing solvent loading to reduce reactor size and energy required for solvent recovery, implementing condensed-phase separations for solvent recovery, and utilizing the entirety of RCF oil for value-added products.

**References/Publications**


1. W. Schutyser, T. Renders, S. Van den Bosch, S.-F. Koelewijn, G.T. Beckham, and B.F. Sels. "Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerization, and upgrading." *Chem Society Rev.* (2018), **47:** 852-908.



***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Deploying docking calculations and resource balance allocation modeling alongside kinetic model parameterization to elucidate mechanisms controlling metabolism in *Clostridium thermocellum*

Veda Sheersh Boorla*[1,2] (vqb5186@psu.edu), Kevin Glass[1,2], Debolina Sarkar[1,2], Charles J. Foster[1,2], Costas D. Maranas[1,2], and **Gerald A. Tuskan**[2]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; [2]Center of Bioenergy Innovation, Oak Ridge National Laboratory, TN.

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

A wide range of natively produced value chemicals and a high native cellulase activity make *Clostridium thermocellum* an attractive bioproduction platform especially for CBP. Despite significant efforts, achieved yields and titers fall below industrially relevant targets. One of the reasons is a limited understanding of enzymatic, regulatory, and/or possible thermodynamic bottlenecks that might be at play in *C. thermocellum's* metabolism. To bridge some of these knowledge gaps and propose engineering strategies for improving yields of desired products, we deploy computational tools to (1) postulate regulatory mechanism active in *C. thermocellum* metabolism, and (2) quantify the enzyme cost of glycolysis and explore the trade-off between more efficient glycolysis and ATP production. We have developed an ensemble docking workflow to compare structural energetics of enzyme-ligand interactions for identifying novel substrate-level regulatory mechanisms with those of experimentally characterized protein structures. By using a kinetic model of *C. thermocellum's* core metabolism as the basis, we hypothesized the presence of several substrate-level enzyme inhibitions based on the improvement in kinetic model's fitness upon their addition. We then used the developed docking workflow to evaluate these enzymatic inhibitions using docking studies of regulatory molecules with enzyme structures involved. We thus flagged a total of 67 substrate-level inhibitions across central carbon metabolism supported by both kinetic formalism and docking analysis. To explore the enzyme cost of metabolism, we constructed a genome-scale resource allocation model of *C. thermocellum* metabolism (i.e., cthRAM), using flux-force efficacy (FFE)[1] constraints to ensure enzyme usage consistent glycolysis operating near thermodynamic equilibrium[2].

References/Publications

1] E. Noor et al., *PLoS Comput. Biol*., 2014, doi: 10.1371/journal.pcbi.1003483.

2] T. Jacobson, et al., *m-Systems*, 2020, doi: 10.1128/mSystems.00736-19.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Candidate Genes for Lignin Structure Identified Though Genome Wide Association of Naturally Variant *Populus*

Nathan Bryant*[1] (nbryant3@vols.utk.edu), Jin Zhang[2], Kai Feng[2], Jin-Gui Chen[2], Wellington Muchero[2], Chang Geun Yoo[3,] Timothy J. Tschaplinski[2], Yunqiao Pu[2], Arthur J. Ragauskas[1,2], and **Gerald A. Tuskan**[2]

[1]Department of Chemical and Biomolecular Engineering, University of Tennessee, [2]Center for Bioenergy Innovation, [3]Department of Chemical Engineering, State University of New York College of Environmental Science and Forestry

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Understanding and controlling lignin variation will be a key technology in implementing biorefineries[1]. Further, genetic attribution of plant phenotypes is a barrier to the engineering and improvement on plant performance. To address these barriers, a genome wide association study (GWAS) was conducted to identify the underlying genetic basis of lignin variation. In this study, nuclear magnetic resonance (NMR) was utilized to elucidate twelve structural phenotypes of lignin across over 400 unique genotypes of three-year-old, naturally variant *Populus trichocarpa*. Population phenotyping revealed an average S/G ratio of 2.70 ± 0.45 and an average β-O-4 linkage content of 62.0% ± 3.18%. The population distribution of p-hydroxybenzoate (PB) content was unique in that it is best described by a lognormal distribution, whereas other phenotypes were better modeled by a normal distribution. PB content was also tended to be higher and trend with S/G ratio in samples with a low (<2.70) S/G ratio. The genotype-to-phenotype analysis identified 80 promising candidate genes strongly associated with at least one lignin phenotype. The candidate genes identified by this GWAS analysis can serve as guidance for future transgenesis targets. Several of these identified candidate genes have not previously been associated with lignin biosynthesis. Among the identified candidate genes include several from the 26S proteasome/ubiquitin pathway, including PBD1 (20S proteasome beta subunit D1), ATL13 (RING/U-box superfamily protein), and XBAT32 (RING-type E3 ligase). One candidate gene identified by this analysis, dirigent protein DIR23, was selected for overexpression in transgenic *Populus*. Another dirigent protein was shown to influence stereo- and regioselectivity of monolignol binding in lignans. Preliminary analysis of DIR23 overexpression lines indicates increased β-O-4 linkage content and lower molecular weight compared to wild-type. While additional analyses are required to confirm these results, these early findings serve as a promising indication that this GWAS analysis has identified novel candidate genes associated with lignin

phenotypes. The functional characterization of candidate genes identified in this study could provide new targets for the control of lignin biosynthesis.

**Reference**

1. U.S. Dept. of Energy (2016). *Biorefinery Optimization Workshop Summary Report*.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Identification and validation of a key gene controlling differential flowering time between switchgrass ecotypes

Soyeon Choi*[1,3] (Soyeon.Choi@uga.edu), Ratul Chowdhury[2,3], Peng Qi[1,3], Tom H. Pendergast IV[1,3], Katrien M. Devos[1,3], and **Gerald A. Tuskan**[3]

[1]University of Georgia, Athens, GA; [2]The Pennsylvania State University, University Park, PA; [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN.

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Flowering time in switchgrass, a crucial determinant of biomass yield, differs between upland and lowland ecotypes. Lowland switchgrass flowers later than uplands but key genes contributing to the delayed flowering in lowlands have not been elucidated. In this research, *PvHd1,* orthologous to rice *Hd1,* has been identified and validated as a causal gene to the differential flowering time between switchgrass ecotypes. This provides the first example in switchgrass of QTL data advancing to validated underlying gene. Two alleles of *PvHd1* encoded proteins that differed by an ecotype-specific amino acid in B-Box domain 1 and the protein variants were predicted to be folded differentially due to the single amino acid substitution (p.S35G). The effect of the amino acid variant was assessed in *CONSTANS* (*CO*, orthologous to rice *Hd1*)-null *Arabidopsis* plants, showing that the p.S35G substitution is causal to the delayed flowering time. The findings from this research will help control the onset of flowering in switchgrass and contribute to acquiring higher biomass from the bioenergy crop across a wider cultivation area.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Catalytic Upgrading of Bioderived Alcohols to Sustainable Aviation Fuel (SAF)

Meijun Li[1], Stephen Purdy[1], Aibolat Koishybay[1], Zhenglong Li[1], Brian Davison[1] (davisonbh@ornl.gov), Andrew Sutton[1] (suttonad@ornl.gov) and **Gerald A. Tuskan**[1]

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Upgrading small bioderived alcohols to Sustainable Jet Fuel (SAF) is one promising alcohol-to-jet pathway as the olefin intermediates can be readily converted to jet-range hydrocarbons with high selectivity. We have shown that using the zeolite H-ZSM-5 we can convert butanol to butenes with near quantitative conversion at 225 °C and ambient pressure, where 1-butene is the dominating fraction among butene isomers.[1] The olefins were further oligomerized into longer-chain hydrocarbons over Amberlyst-36 catalyst at 150 °C and 15 bar, leading to 70% of jet, 12% of gasoline and 10% of diesel fractions. Recently we have modified the catalyst by incorporating copper with zinc[2] and lanthanide metals which allows us to tailor the olefin composition and conversion efficiencies to enable us to determine the optimal conditions to maximize the carbon we can incorporate in the resultant SAF. This also inhibits ethylene formation which is not conducive to higher olefin formation and allows for > 90% selectivity to $C_{4+}$ olefins. Preliminary catalyst stability has been evaluated with over 200 hours on stream and regeneration of the catalyst performed successfully over this period. We have also investigated the relationship between Lewis acid sites and alcohol dehydrogenation and C-C bond formation activity. We are currently determining the role of metal distribution on deactivation and subsequent regeneration and how these effects are influenced by the choice of lanthanide metal.

**References/Publications**

1. Yuk S., Lee M-S., Collinge G., Zhang J., Padmaperuma A. B., Li Z., Polo-Garzon F., Wu Z., Glezakou V-A., Rousseau R., *J. Phys. Chem. C.,* 2020, 124, 52, 28437
2. Zhang J., Wegener E. C., Samad N. R., Harris J. W., Unocic K. A., Allard L. F., Purdy S., Adhikari S., Cordon M. J., Miller J. T., Krause T. R., Cheng S., Liu D., Li M., Jiang X., Wu Z., Li Z., *ACS Catal*., 2021, 11, 15, 9885

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Dynamic Control of Aromatic Catabolism, *In Situ* Efflux Pump Engineering, and High-Throughput Functional Genomics in *P. putdia* KT2440 Enabled by CRISPR-Cas9

Jacob Fenster*,[1,3] (jacob.fenster@colorado.edu), Allison Werner,[2,3] Emily Freed[1,3], Jian Wei Tay[1], Erica Prates[3], Andrew Levitt[1], Audrey Watson[1], Jeff Cameron[1], Carrie Eckert[1,2,3], and **Gerald A. Tuskan**[3]

[1]University of Colorado, Boulder [2]National Renewable Energy Laboratory, Golden, CO; and [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

cbi.ornl.gov

**The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI has identified key barriers for the current bioeconomy in (1) high-yielding, robust feedstocks, (2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and (3) methods to create valuable byproducts from the lignin residues. CBI will identify and utilize key plant genes for growth, yield, composition and sustainability traits as a means of achieving lower feedstock costs, focusing on the perennial feedstocks - poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and esters) using CBP at high rates, titers and yield. CBP will be combined with cotreatment or pretreatment. CBI will maximize process and product value by *in planta* modifications of lignin and biological funneling of lignin to value-added chemicals. Techniques for rapid domestication of non-model microbes and plants will be improved.**

*Pseudomonas putida* KT2440 is a promising chassis for the renewable conversion of lignin into commodity chemicals. While this organism has a wealth of genetic tools available, high-throughput (HTP) experiments that enable forward engineering in *P. putida* are in their infancy. As current microbial engineering efforts rely on the Design, Build, Test cycle to generate desirable production phenotypes, it is critical to deploy HTP capabilities in this organism to domesticate this host. Towards this goal, this work seeks to leverage CRISPR-Cas9 technologies to demonstrate HTP capabilities on various biotechnologically relevant targets.

First, CRISPR-interference was optimized by screening inducible promoters expressing dCas9 on various targets. The arabinose inducible promoter was shown to be valuable for targeting essential genes such as *ftsZ* and *rpoD,* where a tightly regulated off-state allowed for the construction of inducible knockdown mutants. The lactose inducible system, with its higher induction strength was shown to be applicable for targeting metabolic proteins such as *pcaIJ* and *catB*. *PcaIJ* was knocked down during growth shunting 76% of the fed p-coumarte to beta-ketoadipate, producing a titer of 4.4 g/L after 36 hours. This initial study has laid the foundation for the study of a CRISPRi guide RNA (gRNA) library, constructed through a collaboration with JGI, targeting every coding region in the KT2440 genome. Successful pilot experiments have established the protocols for studying the full 87,000 gRNA library that will profile knockdown mutants in biotechnologically relevant growth and stress conditions.

Second, the CRISPR-Cas9 genome editing toolset first developed by Sun et al. (2018) has been optimized for library scale mutagenesis. The minimum homology length requirements for efficiently introducing both deletion and single nucleotide polymorphism (SNP) mutations was determined and the transformation protocol was optimized to increase the number of edited cells 100-fold while keeping the editing efficiency at 100%. This updated design was used to introduce SNPs in the TolC-like multidrug efflux pump, ttgABC, that were discovered during adaptive laboratory evolution. Efforts are underway to conduct deep scanning mutagenesis of this efflux pump in HTP to discover novel mutations that determine substrate binding and recognition important to both antibiotic resistance and tolerance to high lignin concentrations. Last, to make future genome editing experiments a plug and play effort, the activity of 150,000 gRNAs, constructed through a collaboration with JGI, will be assayed in order to catalogue efficient gRNAs and to develop a predictive gRNA model.

**References/Publications**

1. Sun J, Wang Q, Jiang Y, Wen Z, Yang L, Wu J, et al. Genome editing and transcriptional repression in Pseudomonas putida KT2440 via the type II CRISPR system. Microb Cell Fact. 2018 Mar 13;17(1):41

*The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.*

# Declining Carbohydrate Solubilization with Increasing Solids Loading During Fermentation of Cellulosic Feedstocks by *Clostridium thermocellum*: Documentation and Diagnostic Tests

Matthew R. Kubis[1,2], Evert K. Holwerda*[1,2] (evert.k.holwerda@dartmouth.edu), Lee R. Lynd[1,2] and **Gerald A. Tuskan**[2]

[1]Thayer School of Engineering, Dartmouth College, 14 Engineering Drive, Hanover, NH; [2]The Center for Bioenergy Innovation, Oak Ridge National Laboratory

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

The cellulolytic thermophilic anaerobe *Clostridium thermocellum* is one of the most effective biocatalysts for solubilization of carbohydrate harbored in lignocellulose. Pursuant to major goal #2 for CBI (above), we aim to demonstrate values of key performance indicators supporting the economic viability of CBP for conversion of cellulosic feedstocks to small molecules – e.g. ethanol, butanol, or 2,3-butane diol (BDO) - that can be upgraded to sustainable aviation fuels. As part of this effort, we seek to document, diagnose, and remediate factors that limit biomass deconstruction at high solids. This study aims to document the solubilization performance of *Clostridium thermocellum* at increasing solids concentrations for two lignocellulosic feedstocks, corn stover and switchgrass, and explore potential effectors of solubilization performance. Diminishing fractional carbohydrate solubilization with increasing substrate loading was observed for *C. thermocellum* mediated-solubilization and fermentation of unpretreated lignocellulose feedstocks. Results of experiments involving spent broth addition do not support a major role for inhibitors present in the liquid phase. Mid-fermentation addition experiments of cells (fresh biocatalyst), cellulose (model insoluble substrate) and cellobiose (model soluble substrate) confirm that *C. thermocellum* and its enzymes remain capable of converting model substrates during the middle of high solids lignocellulose fermentation. An increase in fractional carbohydrate solubilization was made possible by 1) mid-fermentation solid loading dilutions and 2) coculturing *C. thermocellum* with *T. thermosaccharolyticum*, which ferments solubilized hemicellulose. Incomplete utilization of solubilized carbohydrates suggests that a small fraction of the carbohydrates is unaffected by the extracellular carbohydrate active enzymes present in the culture.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

**Transient delivery of Cas9 using Agrobacterium for genome editing**

Timothy M. Chappell[1,4], Eudald Illa-Berenguer*[2,4] (eillaberenger@uga.edu), Sueme Ueno[2], Wayne A. Parrott[1,2,3,4], and **Gerald A. Tuskan**[4]

[1]Institute of Plant Breeding, Genetics & Genomics, University of Georgia, Athens, GA; [2]Center for Applied Genetic Technologies, University of Georgia, Athens, GA; [3]Department of Crop and Soil Sciences, University of Georgia, Athens, GA; [4]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Improving our knowledge on plant genes is essential to achieve domestication of non-model plant species important for bioenergy. Great advances in this front have been done in recent years, but often genetic engineering and genome editing rely on stable integrations of the DNA of interest. These DNA fragments usually encode genome editing reagents, selectable markers, or novel traits and once integrated in the genome DNA can pose regulatory issues. As such, removal of the integrated DNA is required and achieved via downstream breeding. However, this is not a feasible option in the case of switchgrass and poplar, as downstream breeding and segregation cannot be performed without losing the genotype.

To overcome the limitations imposed by current methods for plant engineering, we generated a system in Agrobacterium, called pTrans, to transiently deliver genome editing reagents as proteins, thus preventing any stable DNA integrations. As preliminary work in tobacco showed promising results, the system has now been improved to target important bioenergy traits in switchgrass. A newer version of pTrans, designed to have higher cas9 expression and more efficient reagent delivery, has also been assembled. This system is currently being tested in switchgrass.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Rapid domestication of poplar using genomic selection and *P. trichocarpa* X *P.deltoides* hybrids

David Kainer*[1] (kainerd@ornl.gov), Jonathon Romero[2], David Macaya-Sanz[3], Roshan Abeyratne[3], Brian Stanton[4], Daniel Jacobson[1], Stephen P. DiFazio[3] and **Gerald A. Tuskan**[1]

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory; [2]Bredesen Center, University of Tennessee, Knoxville, TN; [3]West Virginia University, WV; [4]Xenots Solutions LLC, WA

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

*Populus trichocarpa* is an undomesticated woody species with high potential as a bioenergy feedstock, but a long breeding cycle. To rapidly improve biomass yield per hectare we need to optimize gains per generation. Sustainable and uniform yield is a highly complex conglomerate of traits, so genetically modifying or selecting for a few key loci is unlikely to achieve our goals. Genomic Selection models the genetic (breeding) value of genome-wide variants on target traits in a training population, so is a practical solution. Parents from a breeding population are selected solely on their aggregate genetic value for target traits. The same model can be used to predict progeny phenotypes well before they reach maturity, allowing early culling and rapid cycle times. We have developed a computational pipeline that generates virtual recombinant progeny from a given cross, thus allowing a machine-learning prediction of progeny performance as a result of prospective parental selections and crosses. This "virtual progeny trial" can enable us to predict the relative performance of specific crosses (i.e. families) before they are planted in the field. We tested this pipeline using the 7x7 *P. trichocarpa* proof of concept trial at Clatskanie, OR, which contains 49 full-sib families from 49 controlled crosses. We find a strong ability to predict family rankings for height at year 3.

CBI has access to a great range of genetically diverse genotypes of *P. trichocarpa* (Ptri) for the purposes of genomic selection for Ptri X Ptri crosses. However, crossing *P. trichocarpa* with the closely related *P. deltoides* (Pdelt) may have better outcomes. Firstly, Pdelt is naturally resistant to various pathogens that can devastate pure Ptri plantations, so elite lines from hybrid crosses that inherit Pdelt resistance components will be more sustainable in the long term. Secondly, hybrids (Pdelt X Ptri) often outperform within-species crosses (Ptri x Ptri) for growth due to heterosis

effects. We are now exploring selections of parental materials for Pdelt X Ptri crosses in order to test these advantages. Hybrids do present new challenges in terms of genomic selection models since markers may be inconsistent between the two species, the effects of common markers can vary, genomic regions present in one species may be absent in another, and heterosis effects are difficult to model.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

**Discovery, characterization, and metabolic engineering of Rieske non-heme iron monooxygenases for guaiacol O-demethylation**

Alissa Bleem*[1,2] (alissa.bleem@nrel.gov), Eugene Kuatsjah*[1,2] (eugene.kuatsjah@nrel.gov), Gerald N. Presley[2,4,6], Daniel J. Hinchen[3], David C. Garcia[5], William E. Michener[1], Marco N. Allemann[4], John E. McGeehan[3], Richard J. Giannone[2,4], Gregg T. Beckham[1,2], Joshua K. Michener[2,4], and **Gerald A. Tuskan**[2,4]

[1]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN.; [3]Center for Enzyme Innovation, School of Biological Sciences, Institute of Biological and Biomedical Sciences, University of Portsmouth, UK; [4]Biosciences Division, Oak Ridge National Laboratory, TN; [5]The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN; [6]Department of Wood Science and Engineering, Oregon State University, Corvallis, OR.

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Lignin is the second most abundant renewable biopolymer on earth after cellulose, and aerobic bacteria offer a promising approach to convert mixtures of lignin-related aromatics to a single, value-added product. Aryl-*O*-demethylation represents a critical reaction in the catabolism of lignin-related phenols such as guaiacol, and it is often a bottleneck for both native and engineered bioconversion pathways in these organisms. Here, we utilized randomly barcoded transposon insertion sequencing (RB-TnSeq) to identify a novel Rieske-type guaiacol *O*-demethylase, GdmA, in the aromatic-catabolic bacterium *Novosphingobium aromaticivorans*. Similarity searches identified GdmA homologs in related bacteria, as well as candidate reductase partners, GdmB. Various GdmAB combinations were characterized biochemically for activity on lignin-related substrates, including guaiacol. GdmAB combinations were also evaluated in *Pseudomonas putida* KT2440, a common microbial chassis for bioconversion of aromatic compounds that does not natively utilize guaiacol. The GdmAB pair from *Cupriavidus necator* N-1 demonstrated the highest rate of guaiacol turnover both *in vitro* and in engineered strains of *P. putida*. Additionally, comparison of the GdmAB Rieske-type mechanism with a previously described cytochrome P450 system (GcoAB) demonstrated superior performance of GdmAB for guaiacol catabolism in *P. putida*. The novel GdmAB *O*-demethylases described here represent a new mechanism for

biological assimilation of guaiacol, and they offer a suite of options for microbial conversion of a common lignin-derived substrate.

**The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.**

# Employing rapid, accurate, high-precision phenotyping in poplar, switchgrass, and bacteria

John Lagergren*[1,4] (lagergrenjh@ornl.gov), Jared Streich[1,4], Hari Chhetri[1,4], Delyana Vasileva[1,4], Jaclyn Noshay[1,4], Thomas Pendergast[2,4], Manesh Shah[1,4], Mirko Pavicic[1,4], Joao Felipe Gabriel Machado-Gazolla[1,4], Steve DiFazio[3,4], Jonathon Cumming[3,4], Josh Michener[1,4], Daniel Jacobson[1,4], and **Gerald A. Tuskan**[4]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]University of Georgia, Athens, GA; [3]West Virginia University, Morgantown, WV; [4]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN.

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Traditional plant phenotyping relies on manual measurement of traits, which can be time-consuming, labor-intensive, subject to human error, and often destructive to the samples being scored. Thus, we aim to resolve such bottlenecks by developing field-ready devices (using edge computing and field digitization) and computational tools (deep learning and explainable AI) to facilitate population-scale analyses like GWAS. We apply these methods to data arising from the field and greenhouse in conjunction with ORNL's high performance computing (HPC) resources (e.g., Oak Ridge Leadership Computing Facility).

*Edge computing and phenotyping devices:* Field-ready devices for switchgrass and poplar feedstocks are developed and deployed that utilize (i) edge computing (e.g., GPU-accelerated computing, real-time machine learning, sample tracking, data storage, etc.) for measuring plant traits and (ii) field digitization (e.g., RFID tagging, QR codes, barcodes, plant wearables, soil probes, electrolytic readers, in-situ monitoring devices, etc.) for wireless data capture. This includes the development of handheld phenotyping devices like the Diameter at Breast Height Camera, Rapid Thermal Imager, and μ-tablet, which utilize automated data capture, processing, and storage within the same device. For bacteria, hardware and computational tools are developed for fast and accurate feature extraction from scanned images of bacterial colonies on culture plates.

*Image processing:* We use deep learning and explainable AI to extract and analyze biologically relevant plant features from multimodal image data captured from field devices (cameras, tablets, drones, etc.) and phenotyping facilities (e.g., the Advanced Plant Phenotyping Laboratory at ORNL). Deep learning is used to identify salient image features while explainable AI relates such features to complex plant traits (e.g., environment, biomass yield and composition, etc.).

*Project Impact:* This work enables fast and accurate plant phenotyping in the field and greenhouse, in which a single individual can collect high-precision phenotypes at population scale in a fraction of the time and energy compared to traditional strategies.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Drought Stress Alters Plant-Microbe Interactions but is Contingent on Host Genetic Background

Spencer Roth*[1] (rothsw@ornl.gov), Alyssa Carrell[1], Paul Abraham[1], Dana Carper[1], Mindy Clark[1], Jonathan Cumming[1,2], Dawn Klingeman[1], Melissa A. Cregger[1] and **Gerald A. Tuskan**[1]

[1]Center For Bioenergy Innovation, Oak Ridge National Laboratory; [2]University of Maryland Eastern Shore, Princess Anne, MD

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

To avoid competition with food crops, biofeedstocks should be grown on marginal lands, such as those at risk for drought. Mitigation of drought stress on biofeedstocks therefore plays an important role in increasing the sustainability of biofuels production. Plant-microbial interactions have been shown to alleviate abiotic stress on plants, however the impact of drought stress and host genotype on plant-microbe interactions in the biofeedstock *Populus trichocarpa* remain underexplored. To understand how plant-microbe interactions in *Populus* are impacted by host genotype and drought, we grew drought-tolerant and drought-susceptible genotypes of *Populus trichocarpa* under drought and control conditions in a greenhouse. Following four weeks of growth under varied water conditions, we measured plant growth and characterized the root-associated microbial communities through 16S rRNA and ITS amplicon sequencing. Our results revealed that root biomass was significantly higher in drought-tolerant genotypes, suggesting a potential drought-stress mitigation strategy. Root-associated microbial communities were also significantly impacted by drought condition and host genotype. Bacterial/archaeal and fungal alpha diversity was significantly lower under drought conditions, and drought-tolerant *Populus* genotypes had lower fungal alpha diversity than drought-susceptible hosts. Furthermore, drought condition and host genotype significantly impacted the compositions of the bacterial/archaeal and fungal communities. Drought condition was the most important explanatory variable shaping the bacterial/archaeal community composition, while host genotype accounted for the most variation in the composition of the fungal community. These results suggest that *Populus* genotype exerts control over the selection of root-associated microbial communities and that drought stress alters these communities. The selection of specific microbial communities by *Populus trichocarpa* may confer benefits to the plant. Our results point to exciting future research avenues, such as understanding the mechanisms by which *Populus* selects for beneficial microbial communities.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Integrating Synthetic Biology and Polysaccharide Synthesis for Designer Polymers with Tunable Properties

Peter Smith*[1,3] (peter.smith@nrel.gov), Hsin-Tzu Wang[2,3], Thomas Curry[2,3], Vivek Bharadwaj[1,3], Breeanna Urbanowicz[2,3], Yannick Bomble[1,3] and **Gerald A. Tuskan**[3]

[1]National Renewable Energy Laboratory, Golden, CO; [2]The Complex Carbohydrate Research Center, University of Georgia, Athens, GA; [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN.

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Plant cell wall represent one of the most abundant sources of renewable material available for the production of biobased fuels and products. Plant cell walls are comprised primarily of a complex matrix of cellulose, lignin and a diverse group of polysaccharides known as hemicelluloses. Hemicelluloses such as xylan and xyloglucan are essential to plant growth and fitness and can affect outcomes of biomass deconstruction and fermentation. For these reasons engineering hemicellulose structure is an attractive approach to improving biomass quality for bioprocessing. We are seeking to apply principles of synthetic biology to plant hemicellulose engineering to create novel or well-defined polysaccharides that may serve to improve the properties of biomass for deconstruction and valorization. Using recombinantly expressed plant cell wall glycosyltransferases as biocatalysts, we are building and modifying plant cell wall polymers *in vitro.* This approach allows for the determination of structure-function relationships of polysaccharides and other wall components which are difficult to obtain due to the complexity of structures often exhibited from polysaccharides isolated from plant feedstocks.

As an example of this approach, we are using a Xyloglucan Xylosyl Transferase (XXT1) to modify mixed-linkage glucans (polymers typically devoid of glycosyl branches), with α 1,6 xylosyl or α 1,6 glucosyl residues *in vitro*. By controlling the amount of substitutions on the mixed linkage glucan backbone, we have demonstrated the ability to fine tune the interactions of mixed-linkage glucans with nanocellulose surfaces as determined by a Quartz Crystal Microbalance based assay. In addition, we have developed a platform for the bottom-up synthesis of acetylated xylan polymers using a recombinant Xylan Synthase (XYS1) and Xylan *O*-acetyl Transferase (XOAT1) *in vitro*[1]. Insights gained by computationally modeling enzyme-substrate interactions were used for the rational design of XOAT1 mutants capable of creating xylan polymers with well-defined acetylation patterning. In summary, we are bringing synthetic biology to plant polysaccharide synthesis by using plant biosynthetic enzymes as biocatalysts to build structurally defined

polymers. This work will help to advance our understanding of polysaccharide structure-function relationships and can serve as a basis for the engineering of plant polysaccharides to create biomass better suited for processing to fuels and materials.

**References/Publications**


1. Wang, H. T., Bharadwaj, V. S., Yang, J. Y., Curry, T. M., Moremen, K. W., Bomble, Y. J., & Urbanowicz, B. R. Rational enzyme design for controlled functionalization of acetylated xylan for cell-free polymer biosynthesis. *Carbohydrate polymers*, *2021 Dec 1; 273*, 118564.



***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Systems biology and genomic approaches uncover distinct genes controlling switchgrass biomass and height

Kyle A. Sullivan*[1,4] (sullivanka@ornl.gov), Hari Chhetri[1,4], David Kainer[1,4], Mirko Pavicic[1,4], Yongqin Wang[2,4], Thomas H Pendergast[3,4], Yun Kang[2,4], Katrien Devos[3,4], Daniel A. Jacobson[1,4], and **Gerald A. Tuskan**[4]

[1] Computational and Predictive Biology, Oak Ridge National Laboratory, TN; [2] Division of Plant Biology, Noble Research Institute, Ardmore, OK; [3] Department of Plant Biology, University of Georgia, Athens, GA; [4]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN.

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Switchgrass (*Panicum virgatum*) is a perennial crop and is an important bioenergy feedstock for the future. As such, it is important to understand which genotypes confer increased yield of switchgrass which can lead to increased conversion of lignocellulose into biofuels. To this end, in this study we performed genome wide association studies from multiple common garden sites (Noble Research Institute and University of Georgia) across the United States to determine which genotypes were associated with increased switchgrass biomass. Furthermore, we leveraged a novel systems biology approach (Random Walk with Restart – Filter; RWR-Filter) using multiplex networks from multiple lines of biological experimental data in order to analyze genetic variants below standard thresholds of genome-wide significance. Using this approach, we attempted to glean mechanistic insight into the biological pathways underlying switchgrass biomass and height. RWR-Filter elucidated a number of genetic variants contributing to switchgrass height that were associated with genes known to be involved in abiotic stress signaling, and transcription factors known to modulate jasmonic acid and ethylene biosynthesis. Conversely, variants implicated in above-ground shoot biomass included genes related to cell wall biosynthesis, solute transport, maintenance of osmotic potential, and maintenance of cellular redox state. Our findings indicate that distinct genetic variants and biological pathways underlie two closely related phenotypes (above-ground shoot biomass and height) which are associated with increased switchgrass yield. By understanding which switchgrass genotypes confer optimal growth and biomass yield, we can better inform genomic selection efforts to increase feedstock productivity.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

## Understanding the genetic basis of drought tolerance in *Populus trichocarpa*

Marie C Klein[1], Jack Bailey-Bale[1], Troy Magney[1], Tom Buckley[1], Wellington Muchero[2], Jay Chen[2], Stephen DiFazio[3], Timothy J. Tschaplinski[2], Gail Taylor*[1] (gtaylor@ucdavis.edu) and **Gerald A. Tuskan**[2]

[1]Department of Plant Sciences, University of California- Davis, CA [2]Center for Bioenergy Innovation and Biosciences Division, Oak Ridge National Laboratory, TN, [3]Department of Biology, West Virginia University, WV

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

*Populus trichocarpa* is a fast-growing tree and a potential biofuel feedstock crop, but future large-scale crop production is likely to be required. In 2020 we initiated a new multi-environment multi-year field trial with over 1000 genotypes of *P. trichocarpa* on the UC Davis Farm, California, with the aim of discovering the genomic basis of phenotypic diversity and drought tolerance, enabling the development of tree breeding for water limited environments. We are focused on finding adaptive traits that enable poplar trees to maintain yield in water-limited conditions. We are identifying genetic loci and candidate genes linked to these traits and these will be tested and deployed in breeding pipelines, enabling the development of energy tree crops suited to marginal, low input sites. We established the site in 2020, planting multiple replicates of over 1,000 unique genotypes across 15 acres. In 2021, we implemented a controlled drought treatment across 5 of the 15 acres, throughout the growing season. We are deploying the latest molecular, leaf, proximal and remote sensing technologies for high through-put phenotyping. For example, 6 phenocams mounted at 80 feet high on three separate towers take images every 30 min of several hundred trees, every day of the year. This is producing valuable data on canopy phenology of relevance to yield and drought tolerance. We have identified a set of tree genotypes that are able to tolerate this drought and we are already identifying gene loci explaining traits of interest. We are currently characterizing the relationship between climate of origin and alleles contributing to the phenological and yield traits to address hypotheses about their role in local adaptation to climate. We have further identified loci contributing to performance traits (e.g., biomass estimates) under drought and well-watered conditions and tested for evidence of this variation reflecting historical selection across climate gradients of origin in *P. trichocarpa*. The loci identified are beginning to provide insights into the genetic and phenotypic basis of climate adaptation in *P. trichocarpa* with potential to facilitate genomic assisted breeding.

## References/Publications

1. G Taylor IS Donnison, D Murphy-Bokern, M Morgante et al. (2019). Sustainable bioenergy for climate mitigation: developing drought-tolerant trees and grasses, Annals of Botany **124**, 513-520

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

**Quantitative-trait loci (QTL) mapping: a novel method for dissecting the genetic basis of complex phenotypes in bacteria**

Delyana P. Vasileva*[1,2] (vasilevadp@ornl.gov), Jared C. Streich[1,2], Hari B. Chhetri[1,2], Leah H. Burdick[1], Dawn M. Klingeman[1], J. Christopher Ellis[1,2], Daniel A. Jacobson[1,2], Joshua K. Michener[1,2], and **Gerald A. Tuskan**[1,2]

[1]Biosciences Division and [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Most phenotypes in bacteria are dictated by multiple loci spread across the entire chromosome which have varying magnitudes of effects. However, despite rapid increase in the number of sequenced bacterial genomes, our ability to decipher the genetic architecture of such complex traits has lagged far behind the ability to gather genomic data. Assigning genotypes to phenotypes is of great importance in order to achieve a highly complete picture of metabolic, regulatory and signalling networks that would accelerate rational engineering of cell functions in nonmodel microbes.

QTL mapping is a powerful technique in eukaryotic genetics for identifying genetic loci that affect a phenotype of interest. However, the potential of this method for gene discovery in bacteria has remained unexplored because it traditionally relies on sexual recombination to break linkages between genetic variants. In this study, our approach uses genome shuffling by protoplast fusion to mimic the effects of sexual recombination in bacteria. Using various *Bacillus subtilis* strains as parents, we showed that genome shuffling produces multiple random recombination events throughout the chromosome creating mosaic genomes with unique recombination landscapes suitable for QTL mapping. We developed an integrated computational workflow for bacterial QTL mapping and through recursive protoplast fusion we constructed a QTL population of highly recombinant progeny that can be phenotypically screened in a high-throughput manner to identify causal genetic variants. Using several phenotypic assays, we demonstrated the power of bacterial QTL mapping to link phenotypes to genotypes at the level of biosynthetic pathways and even subgene regions. Furthermore, we developed experimental methods for protoplast fusion in *Clostridium thermocellum* and showed that genome shuffling between *C. thermocellum* strains yields genome-wide recombination patterns. Our work sets a

platform for studying the genetic architecture of complex phenotypes in model and non-model bacteria providing knowledge that will facilitate rapid design for biotechnology applications.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Genome-wide association studies of drought tolerance and water use efficiency related traits in switchgrass

Yongqin Wang*[1,2,3] (yqwang@uga.edu), Hari B. Chhetri[3], Kyle Sullivan[3], David Kainer[3], Larry M. York[2,3], Malay C. Saha[2,3], Michael K. Udvardi[2,3], Daniel Jacobson[3], Yun Kang[2,3], Ali Missaoui[1,3] and **Gerald A. Tuskan**[3]

[1]University of Georgia, Athens, Georgia; [2]Noble Research Institute, LLC, Ardmore, OK; [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN.

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Switchgrass (*Panicum virgatum* L.) is a promising feedstock for biofuels in the United States. As for most crops, periodic drought often limits its productivity, especially in marginal lands where it is likely to be planted. Improving drought tolerance and water use efficiency (WUE) in crops through breeding and biotechnology strategies is essential to meet global food and fuel demands and to achieve sustainable agriculture. In this CBI project, we aim to evaluate drought tolerance and WUE related physiological and biochemical traits of a switchgrass genome-wide association study (GWAS) population (Juenger et al, 2016), and to understand mechanisms and identify key genes underlying drought tolerance and WUE in switchgrass through GWAS.

Phenotyping experiments were conducted under greenhouse conditions using a switchgrass GWAS panel of 299 genotypes with five biological replicates (two tillers per replicate) each for drought-stressed and well-watered treatments per genotype. Three weeks after tiller transplanting, drought stress was applied by withholding watering, and drought stress levels were monitored with a soil moisture sensor. When the soil volumetric water content dropped to 5% (or at the wilting point), which usually occurred in two to four weeks after withholding water, phenotypic data on related traits were collected. For well-watered plants, phenotypic data were collected eight weeks after tiller transplanting. Traits characterized included leaf area, specific leaf weight, stomatal density, leaf cuticular wax, leaf osmotic pressure, shoot and root biomass/ratio, water use efficiency, and root traits. Large genotypic variations were observed in all the traits characterized. GWAS analyses have been performed using the standard GCTA MLMA model, and multiple significant marker associations were identified. Top candidate genes have been evaluated by networks analysis using Random Walk with Restart/Lines of Evidence on Multiplex Graphs (Wen

et al, 2021), and will be further evaluated by PCR, RT-PCR, and functional genomics tools such as CRISPR, stable transformation or virus-mediated gene silencing or overexpression.

**References/Publications**


1. Juenger TE, Schmutz J, Wolke T, Fritschi F, Zare A, Bartley L, et al. Climate adaptation and sustainability in switchgrass: Exploring plant-microbe-soil interactions across continental scale environmental gradients. IN: Genomic Science Contractors–Grantees Meeting XIV and USDA-DOE Plant Feedstock Genomics for Bioenergy Meeting. 2016.
2. Wen Y, Song X, Yan B, Yang X, Wu L, Leng D, et al. Multi-dimensional data integration algorithm based on random walk with restart. BMC Bioinformatics. 2021, 22: 97.



***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# Continuous hydrodeoxygenation of lignin to jet-range aromatic hydrocarbons

Michael L. Stone[1,3], Matthew S. Webber*[1,3] (mswebber@mit.edu)[1,3], William P. Mounfield III[1,3], Earl Christensen[2,3], Ana R. C. Morais[2,3], Yanding Li[1,3], Eric Anderson[1,3], Gregg T. Beckham[2,3], Yuriy Román-Leshkov[1,3], and **Gerald A. Tuskan**[3]

[1]Massachusetts Institute of Technology, Department of Chemical Engineering, Cambridge MA; [2]National Bioenergy Center, National Renewable Energy Laboratory, Golden CO; [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

The need for sustainable aviation fuels (SAFs) is highlighted by the increasing global demand for air travel, which is expected to more than double by 2050. Meanwhile, the IATA is currently aiming to decrease carbon emissions from aviation sources by half relative to 2005 levels within the same timeframe. Current SAFs are limited in that they fail to provide the requisite aromatic content to obtain the necessary physical properties to function as a drop-in fuel substitute, necessitating a 50% blend limit with conventional jet fuels. This underscores an interesting opportunity to convert lignin, the largest source of renewable aromatics available in nature, to jet-range aromatic blendstocks, thereby allowing for the production of a 100% renewable SAF. Deoxygenation technologies are essential for the upgrading of lignin for applications in aviation fuels, as lignin possesses an extremely high oxygen content (27-34 wt%) relative to the trace levels allowable in aviation fuels. Here, we demonstrate a flowthrough process combining reductive catalytic fractionation (RCF) with hydrodeoxygenation to selectively produce a stream of deoxygenated aromatic hydrocarbons from raw lignocellulosic biomass. Using a dual-pass system and a three-phase trickle-bed reactor, we both evaluate catalyst stability while achieving complete deoxygenation of a neat poplar lignin oil derived from RCF with an overall 73 C-mol% recovery (86% of theoretical when accounting for methoxy groups). Through a suite of characterization techniques, the product stream's properties were subsequently evaluated for use in aviation fuels. This process demonstrates a simple system for generating SAF blendstocks from lignin, the properties and product distributions of which can potentially be tuned via using different biomass types and their corresponding *in planta* lignin composition.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

# UAV-based Sustainability Traits Modeling of Field-grown Switchgrass GWAS Population

Yaping Xu*[1,2] (yxu86@utk.edu), Vivek Shrestha[1,2], Cristiano Piasecki[1,2], Benjamin Wolfe[1,2], Lance Hamilton[1,2], Reginald J. Millwood[1,2], Mitra Mazarei[1,2], C. Neal Stewart Jr. [1,2], and **Gerald A. Tuskan**[2]

[1]Department of Plant Sciences, University of Tennessee, Knoxville; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN.

cbi.ornl.gov

**Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition, and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to biofuels using CBP with cotreatment at high rates, titers and yield in combination with catalytic upgrading into drop-in hydrocarbon fuel blendstocks.**

Unmanned aerial vehicles (UAVs) provide an intermediate scale of spatial and spectral data collection that yields increased accuracy and consistency in data collection for morphological and physiological traits than satellites and expanded flexibility and high-throughput compared to ground-based data collection. In this study, we used UAV-based remote sensing for automated phenotyping of field-grown switchgrass (*Panicum virgatum*), a leading bioenergy feedstock. Using vegetation indices calculated from a UAV-based multispectral camera, statistical models were developed for rust disease caused by *Puccinia novopanici*, leaf chlorophyll, nitrogen, and lignin contents. For the first time, UAV remote sensing technology was used to explore the potentials for multiple traits associated with sustainable production of switchgrass, and one statistical model was developed for each individual trait based on the statistical correlation between vegetation indices and the corresponding trait. Also, for the first time, lignin content was estimated in switchgrass shoots via UAV-based multispectral image analysis and statistical analysis. The UAV-based models were verified by ground-truthing via correlation analysis between the traits measured manually on the ground-based with UAV-based data. Th the normalized difference vegetation index (NDVI) for rust and nitrogen, while NDVI performs better than NDRE for chlorophyll and lignin. Overall, linear models work well for rust disease, chlorophyll and lignin, but for nitrogen, non-linear models outperform linear models. As the first comprehensive study to model switchgrass sustainability traits from UAV-based remote sensing, these results suggest that this methodology can be utilized for switchgrass high-throughput phenotyping in the field.

***The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.***

**High-throughput Screening for Carboligase Activity in ThDP-dependent Enzymes**

Tracey Dinh[1]*(tracey.dinh@u.northwestern.edu), Rohin Devanathan[1], Bradley W. Biggs[1], Lindsay Caesar[2], Neil L. Kelleher[2,3,4], Paul M. Thomas[4], Linda J. Broadbelt[1], Keith E.J. Tyo[1]

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL; [2]Department of Molecular Biosciences, Northwestern University, Evanston, IL; [3]Department of Chemistry, Northwestern University, Evanston, IL; and [4]Proteomics Center of Excellence, Northwestern University, Evanston, IL

https://pamspublic.science.energy.gov/CCBond

**Project Goals:** The goal of this project is to prospect novel carbon-carbon bond ligation reactions in thiamine diphosphate-dependent enzymes using predictive activity models. Carboligase enzymes will be screened for α-keto acid condensation and machine learning models will be developed to elucidate their catalytic landscape. Activity models will be also be used to predict potential effects of selected enzymes on the *E. coli* metabolome.

**Abstract:** Promiscuous enzymatic reactions create 'biological noise' that can be either advantageous or disruptive to global metabolism. Redundancy among promiscuous reactions can act as resistance to deleterious genetic mutations while others can inadvertently siphon essential metabolites or create toxic products. The potential metabolomic burden resulting from promiscuous, heterologous enzyme expression remain challenging to characterize and mitigate. Toward this end, our team seeks to develop a platform to systematically predict enzyme substrate promiscuity and the resulting metabolomic consequences. Merging both experimental and computational approaches, this work aims to comprehensively characterize the catalytic landscape of thiamine diphosphate (ThDP)-dependent enzymes and uncover novel promiscuous transformations for carbon-carbon bond formation. Predictive cheminformatics-based tools will be developed to gain insight into the chemical properties contributing to altered substrate selectivity and the resulting impact of promiscuous activity on cell metabolism.
A high-throughput activity assay was developed for precise functional analysis of ThDP-dependent enzymes capable of catalyzing carbon-carbon ligation (carboligases). Diverse sets of α-keto acid substrates were screened in multiplexed reactions, and the resulting products were detected by liquid chromatography mass spectrometry (LC-MS) to generate high-quality enzyme activity data. The developed carboligase activity assay will be combined with machine learning classification to rapidly characterize the enzyme-specific activity landscapes. Establishing a method for rapid enzyme screening and characterization will also facilitate subsequent rational mutagenesis of select enzymes and assessing the engineered enzymes for metabolic burden. Predictive activity models will be applied to E. coli metabolites to identify potential cross-reactivity and toxic byproducts. This work demonstrates a platform for rapid biocatalyst development based on substrate promiscuity. We present novel reactions discovered for multiple carboligase enzymes; however the analytical methods and cheminformatics tools developed for reaction screening can be widely applied to other chemistries.

This work is supported by DOE grant DE-SC0019339.

**Title: Curation and Characterization of Conserved Green Lineage Proteins**

**Authors: James Umen**[1]* (jumen@danforthcenter.org), Chen Chen[2], Jianlin Cheng[2], Eric Knoshaug[3], Jian Liu[2], Vladimir Lunin[3], Ambarish Nag[3], Huong Nguyen[1], Peter St. John[3], Peipei Sun[1], Ru Zhang[1].

*PI, senior author, presenter

**Institutions:** [1]Donald Danforth Plant Science Center, St. Louis, MO; [2]University of Missouri, Columbia, MO; [3]National Renewable Energy Laboratory, Golden, CO.

**Website:** http://tulip.rnet.missouri.edu/deepgreen/deepgreen/index.html

**Project Goals:**

Around half of the predicted proteins in most sequenced green-lineage genomes remain as unknowns, with no information on structure or function. Through this project, we will characterize plant proteins of unknown function (Deep Green proteins), including around 500 unknown proteins from the model dicot *Arabidopsis thaliana* (Arabidopsis) and/or the model $C_4$ monocot *Setaria viridis* (Setaria) with homologs in the model green alga *Chlamydomonas reinhardtii* (Chlamydomonas), where we will perform high-throughput functional genomics screening. Our objectives include: 1. Assembly and ongoing curation of the Deep Green candidate protein set; 2. *in silico* structural predictions and network analyses to assign structures and predict function; 3. Assembly and curation of reverse genetic resources in Chlamydomonas; 4. Functional genomics characterization and prioritization in Chlamydomonas; and 5. structural validation of selected candidates and functional validation in Arabidopsis and Setaria.

**Abstract text:**

Sequence-homology and experimental approaches have enabled functional annotation of many plant and algal genes, but around half of the predicted proteins in most sequenced green-lineage genomes remain as unknowns, with no information on structure or function. While some of these unknown proteins are lineage-specific or even species-specific, a sizable number are conserved within the Viridiplantae (green algae + land plants) or within large sub-groups of plants (e.g. monocots and dicots). This project will help fill a major gap in the annotation for large sets of plant proteins whose structures and functions have not yet been characterized, and which represent a relatively untapped resource for bioenergy and synthetic biology applications that underlie the DOE mission. Expertise in structural genomics and high-performance bioinformatics computing from team members at the National Renewable Energy Laboratory (NREL), omics-based computational predictions from team members at University of Missouri (MU), and algal and plant functional genomics expertise from team members at Donald Danforth Plant Science Center will be leveraged to provide this functional annotation. Ongoing work on Deep Green proteins has produced three curated lists of unknown protein families from the three

focal species Arabidopsis, Setaria and Chlamydomonas as well as overlaps between these sets established based on sequence homology criteria. 412 Chlamydomonas Deep Green genes have been identified with homologs in either Setaria (134), Arabidopsis (97) or both (181). Expression profiling revealed non-random distributions of Deep Green gene expression across the diurnal cycle with enrichment during the light phase when photosynthesis-related proteins are upregulated, and they were also predicted to be enriched in chloroplast localization. Secondary structure analysis indicated Deep Green proteins are more structured (i.e. less disordered) in general than the total set of unknown proteins in each species. A manuscript describing the curation process and preliminary characterization of Chlamydomonas Deep Green proteins is in preparation. Under Objective 3 (assembly of reverse genetic resources for Chlamydomonas Deep Green Proteins) we have identified one or more Chlamydomonas CLiP library (1) mutants for 296 Deep Green genes, and for the remainder we have adapted an efficient genome editing procedure (2) that uses CRISPR-Cas9 and a barcoded selectable marker cassette to generate around 180 tagged CRIPSR mutants (2 alleles per gene). Under Objective 2 we applied our MULTICOM tool ranked among top predictors in the 14 Critical Assessment of Protein Structure Prediction (CASP14) to predict the tertiary structures and structural features (i.e., secondary structure, solvent accessibility, disorder, domain boundaries, inter-residue contacts) for 825 out of 1658 Setaria and Arabidopsis Deep Green proteins. The prediction results are available at a user-friendly, browsable website (http://tulip.rnet.missouri.edu/deepgreen/deepgreen/index.html ). These results are being compared and integrated with structure predictions obtained using Alphafold2, I-TASSER and the Rosetta AbinitioRelax module.

The rich new data resources produced under the Deep Green project will be curated in one or more public databases, including DOE-supported KBase. These data will help guide researchers in investigating the contribution of conserved unknown proteins to diverse aspects of plant biology that impact photosynthesis, biomass accumulation, and stress responses. This work will also help fill a major gap in the annotation for large sets of plant proteins whose structures and functions have not yet been characterized, and which represent a relatively untapped resource for bioenergy and synthetic biology applications that underlie the DOE mission.

**References/Publications**

1. X. Li, *et al.*, An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in Chlamydomonas reinhardtii. *The Plant Cell* **28**, 367–387 (2016).

2. T. Picariello, *et al.*, TIM, a targeted insertional mutagenesis method utilizing CRISPR/Cas9 in Chlamydomonas reinhardtii. *Plos One* **15**, e0232594 (2020).

**Funding statement:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0020400 and by JGI Community Sequencing Proposal 506032

**Infernet: Gene Function Inference By Leveraging Large, Organ-Specific Expression Datasets And Validation Of Non-Redundant Regulators**

Rajeev Ranjan*(ranjan9@purdue.edu, rajeevmbbtac@gamil.com )[1], Ying Li[1], Karen Hudson[2] and **Kranthi Varala**[1],

[1]Purdue University, West Lafayette, IN; [2] USDA-ARS, West Lafayette, IN

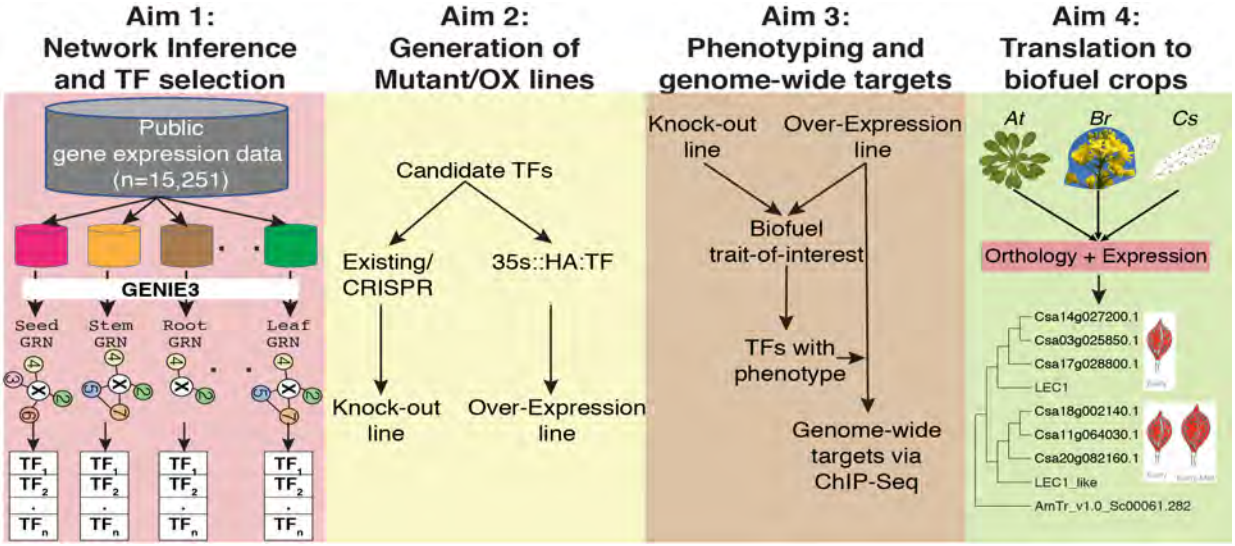
https://www.purdue.edu/hla/sites/varalalab/infernet

## Project Goals

This project combines computational approaches, e.g., machine learning, network inference and phylogenomics, with molecular approaches, e.g., metabolite profiling and ChIP-Seq, to find novel transcription factors (TF) that regulate traits of agronomic or biofuel interest. This project focuses on the biofuel trait of seed oil synthesis as a proof of concept that is extensible to any agronomic/biofuel trait of interest. This project focuses on regulation of a biological process of interest (e.g., lipid biosynthesis) in an organ specific manner (e.g., in seeds) and by estimating the likelihood of a given TF being redundant in its function (Aim 1). We then validate our functional predictions, using transgenic lines (Aim 2), via phenotypic assays (Aim 3a) and by identifying the specific targets these TFs regulate (Aim 3b). Finally, we translate the validated TF regulation knowledge gained in a model species (*Arabidopsis*) to biofuel crops (e.g., *Camelina sativa*) (Aim 4).

## Abstract

Gene regulatory network inference [1] from public RNA-Seq data (Aim 1) predicted TF regulators of seed lipid biosynthesis. The list of top 10 candidate TFs included four known regulators of this process and many novel TFs that are predicted to have a strong influence on seed lipid biosynthesis. We have identified and collected mutant lines and completed phenotyping for traits like seed size and seed weight. Phenotyping for alteration in seed lipid profile



is under progress (Table 1). In addition, we generated over-expression lines for each of these candidate TFs using a set of 3 promoters: i. a generic 35S promoter for robust plant-wide over-expression ii. a seed-specific NapinA promoter for robust seed-specific over-expression and iii. native promoter driven expression. Screening of 35S and Napin promoter driven $T_1$ over-expression transgenic lines is completed and isolation of homozygous T2 lines is under progress. While screening of native promoter driven T1 lines is under progress. T2 overexpression lines showing higher level of TF expression will be used for phenotyping of seed lipid profile along with seed size and weight. Preliminary results show a significant change in total FA content for one novel TF regulator, and significant changes in FA composition for another novel TF regulator.

***Table 1*** *Status of mutant identification and phenotyping, along with generation of over-expression lines for each candidate TF. All TFs were conjugated with a HA tag to enable ChIP-Seq assays to identify global targets of TF binding.*

| TF Name | Mutant | | | | 35Spro:TF-HA | | | NapApro: TF-HA | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mutant Name | ABRC ID | Seed phenotype | Lipid profile | Transfo ration | Transgene Screening (T1) | Transgene Screening (T2) | Transfor mation | Transgene Screening (T1) | Transgene Screening (T2) |
| bHLH93 | *bhlh093_3* | SALK_121082 | Done | Ongoing | Done | Done | Ongoing | Done | Done | Ongoing |
| HB25 | *hb25_2/hb25_4* | SALK_014023/SAIL_517_E03 | Done | Ongoing | Done | Done | Ongoing | Done | Done | |
| DIV2 | *div2_2* | SALK_208938 | | | | | | | | |
| SRM1 | *srm1_2* | SALK_206518 | Done | Ongoing | Done | Done | Ongoing | Done | Done | Ongoing |
| MYB30 | *myb30_1* | SALK_122884 | Done | Ongoing | Done | Done | | Done | Done | |
| DAG2 | *dag2_1* | SALK_201125 | Done | Ongoing | | | | Done | Done | Ongoing |
| CESTA | *cesta_1/cesta_3* | SALK_124840/SAIL_674_A01 | | | Done | Done | Ongoing | | | |
| TGA4 | *tga4_1* | SALK_127923 | Done | Ongoing | Done | Done | Ongoing | Done | Done | Ongoing |
| SPL12 | *spl12_1* | SALK_142295 | Done | Ongoing | | | | | | |
| AGL18 | *agl18_2/agl18_5* | SALK_004483//SALKseq_69591 | Done | Ongoing | | | | | | |
| CAMTA6 | *camta6_1* | SALK_078900 | Done | Ongoing | | | | | | |
| MYB118 | *myb118_1* | SALK_111812 | Done | Ongoing | | | | | | |
| MYB115 | *myb115_1/myb115_2* | SALK_202795/SALK_044168 | Done | Ongoing | Done | Done | Ongoing | Done | Done | Ongoing |
| WRI 1 | *wri1_1* | CS69538 | | | Done | Done | Ongoing | | | |

The GRN inference used in Aim 1 only considered TFs as regulators of gene expression. Subsequently, we repeated the inference pipeline using TFs + epigenetic (i.e., DNA and histone) modifiers as potential regulators of gene expression. Using the shoot and root apical meristems as use cases we identified many known and novel non-TF regulators of plant development. A manuscript describing this study was recently published (McCoy *et. al.,* 2021).

## Publications

McCoy, R.M., Julian, R., Kumar, S.R.V., Ranjan, R., Varala, K., Li, Y. A systems biology approach to identify essential epigenetic regulators for specific biological processes in plants. Plants, 2021, 10 (2), 364.

## References

1. Huynh-Thu, V.A., A. Irrthum, L. Wehenkel and P. Geurts, Inferring regulatory networks from expression data using tree-based methods. PLoS One, 2010. 5(9).

*The InferNet grant is supported by the Genomics Science program within the Office of Biological and Environment Research in the DOE Office of Science.*

**Title:** Elucidating the Evolution of Interspecies Metabolic Interactions within a Methanotroph-Cyanobacteria Coculture using Dynamic Genome-scale Metabolic Modeling

**Authors:** Kiumars Badr[1]*(kzb0054@auburn.edu), Q. Peter He[1], and **Jin Wang**[1]

**Institutions:** [1]Auburn University, Auburn, AL;

**Project Goals:** In nature, microbial communities have developed a highly efficient way to recover energy and capture carbon from both $CH_4$ and $CO_2$ through interspecies coupling of methane oxidation to oxygenic photosynthesis. However, in order to successfully utilize mixed culture for biotechnology applications, both fundamental knowledge and technological gaps have to be addressed. The knowledge gap refers to the lack of systematic study for identifying and quantifying the interactions between community members and how the interactions affect system dynamics. The technological gap refers to the lack of effective methodology, and fast and low-cost analytical tools to characterize mixed culture systems frequently or in real-time. The overall objective of this research is to help address those gaps through developing experimental and computational tools to characterize a synthetic methanotroph-photoautotroph (M-P) binary consortium, to identify and validate interspecies interactions at both systems and cellular levels, and to engineer a model methanotroph-photoautotroph coculture pair for enhanced production of chemicals.

**Abstract Text:** Microbial communities are ubiquitous in nature and play a significant role in various biogeochemical cycles on earth. Microbial communities are dynamic in nature and exhibit different metabolic interactions under different environmental conditions. However, due the inherent complexity of microbial communities, their dynamic responses are not well understood. Recently, genome-scale metabolic modeling (GEM) emerged as an effective tool to study microbial communities. However, the representative dynamic GEM approaches cannot predict the interspecies metabolic interactions within a community, while the steady-state GEM approaches cannot capture the dynamic evolvement of the metabolic interactions. In this work, using a methanotroph-cyanobacteria (*Methylomicrobium buryatense* 5GB1- *Arthrospira platensis*) coculture as the model system, we present a dynamic GEM approach, termed DynamiCom that can predict the evolution of interspecies metabolic interactions within the coculture over time (Fig. 1). The proposed approach builds upon a semi-structured kinetic model (Fig. 2) we developed recently that can accurately predict the growth dynamics of the model coculture under a wide range of conditions. The cross-membrane fluxes, i.e., the individual substrate pickup rates and product excretion rates predicted by the kinetic model, are used as additional constraints for a steady-state coculture GEM (SteadyCom is used in this work). Besides consistently predicting the same set of top metabolic exchanges within the coculture under different *in silico* setups, the proposed approach

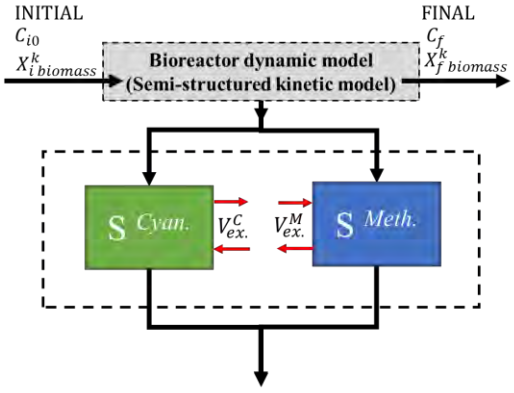



Fig. 1. Schematics of the proposed dynamic GEM model

predicted the establishment of the mutualistic relationship between the methanotroph and cyanobacteria (Fig. 3). Specifically, our results suggest the establishment of an emergent N-exchange within the coculture, where the cyanobacteria consume nitrate, while the methanotroph exclusively consumes ammonium excreted by the cyanobacteria after the mutualistic relationship is established.

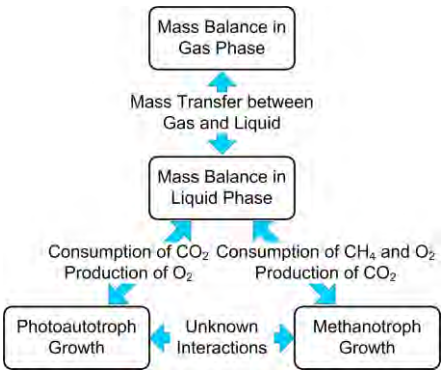
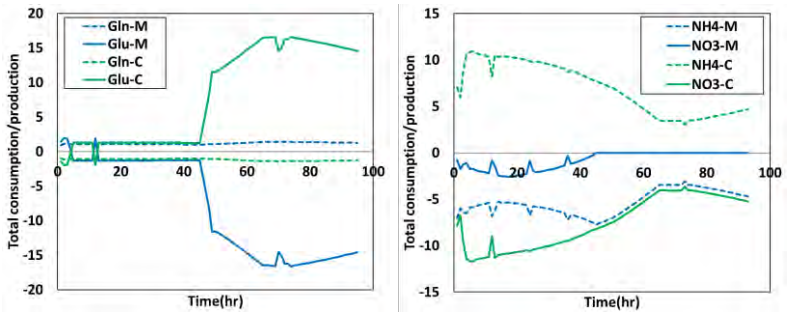



Fig. 2. Overview of the semi-structured kinetic modeling framework




(a) (b)

Figure 3. Dynamic evolution of (a) Glutamine (Gln) and Glutamate (Glu) as an example of top interspecies metabolic interactions; (b) ammonium (NH4) and Nitrate (NO3). M is produced(+)/consumed(-) by methanotroph and C produced(+)/consumed(-) by cyanobacteria. The irregularity in the prediction is the consequence of the change of $CH_4/CO_2$ uptake by the species during and after refeeding the system

**References/Publications**


1. Badr K., He Q.P. & **Wang J.** (2022), A novel semi-structured kinetic model of methanotroph-photoautotroph cocultures for biogas conversion, ***Chemical Engineering Journal***, Vol. 431, 133461; https://doi.org/10.1016/j.cej.2021.133461
2. Stone K., Hilliard M., Badr K., Bradford A., He Q.P. & **Wang J.** (2020), Comparative study of oxygen-limited and methane-limited growth phenotypes of Methylomicrobium buryatense 5GB, ***Biochemical Engineering Journal***, 161, 107707; https://doi.org/10.1016/j.bej.2020.107707
3. He Q.P., **Wang J.**, Hilliard M.V., Culture systems and methods of using same, **USPTO Patent Application** # 16,934,766, filed on July 21, 2020;
4. Badr K., Whalen W., He Q.P. & **Wang J.** (2020), Fast and Easy Quantitative Characterization of Methanotroph-Photoautotroph Cocultures, ***Biotechnology and Bioengineering***, 2020; 1-12; https://doi.org/10.1002/bit.27603;
5. Roberts N., Hilliard M., He Q.P. & **Wang J.** (2020), A microalgae-methanotroph coculture platform to convert wastewater into microbial biomass for fuels and chemical production, ***Frontiers in Energy Research,*** Vol. 8, Article 563352. https://doi.org/10.3389/fenrg.2020.563352;
6. Stone K., He Q.P., & **Wang J.** (2019), Two Experimental Protocols for Accurate Measurement of Gas Component Uptake and Production Rates in Bioconversion Processes, ***Nature - Scientific Report***, 9 (1), 5899.



**Funding Statement:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019181; and by the U.S. Department of Education, Graduate Assistance in Areas of National Need Program under Award Number P200A180002.

**Microbiome transfer and synthetic community approaches for determining the genetic and environmental factors underlying mutualism within a *Sphagnum* peatmoss system**

Alyssa A. Carrell[1], Jun Lee[1], Sara S. Jawdy[1], Dale A. Pelletier[1], Jon Shaw[2], Wellington Muchero[1], Kai Feng[1], Adam Healey[4], John Lovell[3], Jeremy Schmutz[3], David J. Weston[1*] (westondj@ornl.gov)

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]Duke University, Durham NC, Uppsala University, Sweden; [4]Joint Genome Institute, Walnut Creek, CA

**Project Goals: To characterize the *Sphagnum*-diazotroph symbiosis by incorporating plant host *Sphagnum* and microbial genetic variation, variable climatic drivers, and complex communities that scale across biological organizations to regulate peatland carbon and nitrogen cycling.**

The importance of plant-microbiome systems on terrestrial carbon and nitrogen processes is perhaps most pronounced in *Sphagnum* dominated ecosystems, which occupy 3% of the Earth's land surface yet store approximately 30% of terrestrial carbon as recalcitrant organic matter (i.e., peat). The foundation plant *Sphagnum* is responsible for much of the primary production in peatland ecosystems and produces recalcitrant dead organic matter. *Sphagnum* together with associated $N_2$-fixing microorganisms, contributes substantial nitrogen inputs to peatlands. *Sphagnum* growth and production (carbon gain) depends, in part, on a symbiotic association with $N_2$-fixing, diazotrophic microbes. Under changing environmental conditions, a central question about these ecosystems is whether the *Sphagnum*-diazotroph symbiosis will maintain its beneficial interaction, or will it shift to neutral or even antagonistic interactions that ultimately influence peatland carbon gain and storage. To begin to address this question, we initiated a project using synthetic communities, microbiome transfers, genotype-to-phenotype associations, and metabolic characterization to address the overarching hypothesis that genetic variation in *Sphagnum* host and associated microbiome play a key role in determining the environmental tipping point of beneficial symbiosis (i.e., environmental disruption).

To address this hypothesis and more specifically investigate the interaction of host genetics and microbiome thermal origin on host thermotolerance, we mechanically separated the microbiome from *Sphagnum* plants residing in a whole-ecosystem warming study, transferred the component microbes to a 180 member pedigree of germ-free plants, and exposed the new hosts to temperature stress. In some *Sphagnum* individuals the addition of a microbiome from a thermal origin that matched the experimental temperature completely restored plants to their pre-warming growth rates while other *Sphagnum* individuals did not receive a benefit from the addition of a conditioned microbiome. To confirm the results we repeated the experiment with a subset of *Sphagnum* genotypes that corresponded with the highest and lowest benefit from the addition of thermal conditioned microbiomes and found similar results. To identify candidate plant genes responsible for receiving thermal benefits from the microbiome, we performed genome scans using a single normal QTL (quantitative trait loci) model, Haley-Knott regression, and default parameters. A seven gene QTL region was identified and is being tested in heterologous expression systems. Together, our findings show that the microbiome can transmit thermotolerant phenotypes to host plants, providing a valuable strategy for rapidly responding to

environmental change. Future research efforts from this award are progressing in two areas: 1) candidate gene functional analysis for microbial provided thermal benefits for bioenergy crop improvement, and 2) field-based biogeochemical studies exploring the role of genetically predisposed *Sphagnum* genotypes to warming.

**Funding:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Early Career Research Program; and the DOE JGI Community Science Program 504399; and FICUS 504306.

**Development of emerging model microorganisms: Megasphaera elsdenii for biomass and organic acid upgrading to fuels and chemical**s.

Janet Westpheling, University of Georgia and Adam Guss ORNL

The native ability to condense acetyl-CoA groups to efficiently generate C4 to C8 compounds makes *M. elsdenii* a compelling platform for the production of fuels and chemicals from lactate and plant carbohydrates. Our overall objective is to develop *M. elsdenii* into a platform for the conversion of lignocellulosic biomass sugars and organic acids to hexanol and other valuable chemicals. *M. elsdenii* produces organic acids as fermentation products when growing on lactate and glucose, including butyric (four carbon), valeric (five carbon), hexanoic (six carbon), and in some cases octanoic (eight carbon) acids as major fermentation products, likely via a chain elongation pathway using acetyl-CoA. As the carbon chain length increases for the corresponding alcohols, fuel properties improve, making hexanol an appealing target as a next-generation gasoline blend stock beyond ethanol. We are engineering *M. elsdenii* to efficiently produce next-generation, drop-in lignocellulosic fuels such as hexanol at high yield and titer as a general bioengineering platform. While previous efforts to make products like alcohols in *E. coli* have been moderately successful, production of C6 and larger products remains low, suggesting that extending the chain elongation pathway beyond a single cycle remains a significant challenge in model organisms. Engineering *M. elsdenii* is an alternative approach with the potential for exceptionally high yields and titers of C6 products such as hexanol. We developed an initial method for transformation of DNA into two strains of *M. elsdenii* via methylome analysis, heterologous expression of DNA methyltransferases in *E. coli*, and electroporation. We further improved efficiency by creating an *E. coli* methylating strain that uses arabinose-inducible expression, resulting in an additional 10-fold increase in *M. elsdenii* transformation efficiency, to approximately 10,000 cfu/ug DNA. This has facilitated efficient and rapid strain construction. Expression of a bifunctional aldehyde/alcohol dehydrogenase (*adhE2)* gene using the native *M. elsdenii* ribosomal S4 protein promoter resulted in 5.3 mM butanol in *M. elsdenii* ATCC 25940 from lactic acid as the growth substrate. The deletion of a propionyl-CoA transferase in the *M. elsdenii* chromosome resulted in increased acetate co-consumption, loss of propionate production, reduced valerate production, and increased hexanoate and butyrate production from lactic acid as the growth substrate. *This is the first demonstration of metabolic engineering in Megasphaera and proof of concept that this approach may lead to the accomplishment of our longer-term goals.*

**Tools for Faster and More Sensitive Sequence Annotation, and Visualization of Those Annotations**

Genevieve Krause[1], Dave Rich[1], Jack Roddy[1], **Travis Wheeler**[1*](travis.wheeler@umontana.edu)

**Institutions:** [1]University of Montana, Missoula

**Website URLs:**
https://github.com/TravisWheelerLab/hmmer/tree/frameshift
https://github.com/TravisWheelerLab/MMOREseqs
https://sodaviz.org/

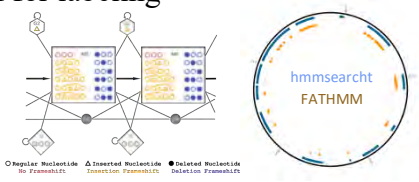
## Project Goals

Microbial communities are ingrained in essentially every conceivable niche. We are particularly motivated by the need to understand soil communities that play a key role in the plant-soil dynamic, with impact on food and fuel crop production. To understand the roles of these microbial communities, it is vital that we maximally annotate their genomic and functional capacity. In this work, we are primarily concerned with (i) the problem of accurately annotating protein-coding DNA that contains insertions or deletions that induce frameshifts in the coding sequence, (ii) performing this annotation at high speed, and (iii) generating complex and interactive visualizations of annotations.

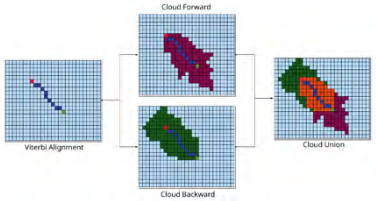
## Abstract Text

We describe advances in methods for annotation with profile hidden Markov models (pHMMs):

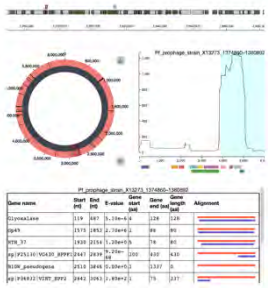
- We have developed an implementation of an extended pHMM for labeling protein-coding DNA with frameshifts. This tool, FATHMM, is substantially more sensitive than competing software methods in the fact of frameshifts, and is only slightly slower than frameshift-oblivious translated search with pHMMs.

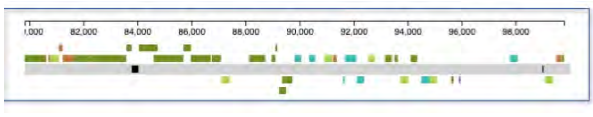


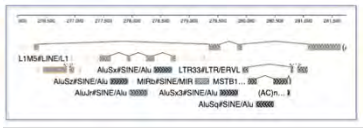

- We have developed a sparse dynamic programming algorithm to produce highly accurate Forward/Backward profile HMM alignments with 20-100x reduction in memory and runtime requirements (MMOREseqs);




- We have created an open-source TypeScript library that supports efficient development of custom, dynamic, and interactive visualizations of annotations of linear and circular genomic sequence (SODA).







**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021216, and NIH NIGMS R01GM132600.

**Title:** A mechanistic approach to parsing pyrophilous lifestyles of fungal and bacterial isolates

**Authors:** Monika Fischer[1]* (monikaf@berkeley.edu), Timothy Berry[2]* (tdberry@wisc.edu), Andrei Steindorff[3]* (assteindorff@lbl.gov), Frances Stark,[1] Nayela Zeba,[2] Igor Grigoriev,[3] **Thea Whitman**,[2] Matthew Traxler[1]

**Institutions:** [1]University of California-Berkeley, Berkeley, CA; [2]University of Wisconsin-Madison, Madison, WI; and [3]Lawrence Berkeley National Laboratory, Berkeley, CA

**Website URL:** https://whitmanlab.soils.wisc.edu/research/ https://traxlerlab.berkeley.edu/

**Project Goals:** In this work, we aim to dissect the effects of microbes (fungal and bacterial) on carbon (C) and nitrogen (N) dynamics in post-forest fire soils. Our conceptual framework is rooted in systems biology and ecology, while our experimental approach combines genomics, transcriptomics, metabolomics, microbial community profiling, stable isotope techniques, small scale fire systems (pyrocosms), tightly controlled methods for producing labeled pyrogenic organic matter (PyOM), and high-throughput monitoring of C mineralization rates. We have three major research objectives: (1) To determine how dominant post-fire soil microbes affect the fate of PyOM; (2) To assess the interaction between N availability and PyOM mineralization by post-fire microbial communities and individual pyrophilous microbes; (3) To define the network of microbial interactions that facilitate PyOM breakdown over time and the key genes involved in this process.

**Abstract Text:** Wildfires are a fundamental ecological process with dramatic effects on ecosystem function and nutrient cycling (including carbon). Soil fungi and bacteria play a key role in post-fire cycling of pyrogenic organic matter, in addition to their roles in other biogeochemical cycles. We are taking a multifaceted approach to investigating the response of specific organisms to wildfire and pyrogenic organic matter additions, using comparative genomics and transcriptomics on isolates that are known pyrophiles or that can grow with PyOM as their only C source. This research has included sequencing the genomes of pyrophilous fungal taxa, including *Lyophyllum atratum*, *Coprinellus angulatus*, *Pholiota molesta*, *Crassisporium funariophilum*, *Pyronema domesticum, Pyronema omphalodes*, *Tricharina praecox*, *Geopyxis carbonaria*, *Morchella snyderi*, and *Peziza echinospora*. Comparative genomics have indicated that the genomes of pyrophilous fungi may be enriched in or have expansion of carbohydrate-active enzymes (CAZymes), families associated with stress response and the initiation of fruiting bodies, both of which may be relevant in the post-fire environment.

Fungi in the genus *Pyronema* dramatically dominated the soil fungal community within weeks to months after the 2013 Rim Fire around Yosemite National Park (CA, USA). *Pyronema domesticum* isolated from this fire was used to identify how it survives and interacts with the cycling of carbon or pyrolyzed carbon. Leveraging the recently sequenced genome, we did an RNAseq experiment with *P. domesticum* grown on four different carbon sources; 750°C PyOM, wildfire burned soil, sucrose minimal medium, and water (no carbon). The PyOM induced

expression of a comprehensive set of genes required for the breakdown and metabolism of aromatic and polyaromatic compounds, which are characteristic of PyOM. We further confirmed that *P. domesticum* could mineralize $^{13}$C-labelled 750°C PyOM to $CO_2$. Taken together, our results show that the pioneering post-fire fungus *P. domesticum* can liberate carbon from complex PyOM and return it to the carbon cycle.

In addition to the fungal isolates described above, over 70 bacterial isolates have been collected from fire-affected soils and assayed for their ability to mineralize PyOM. The 12 isolates with the most robust growth on PyOM agar were cultured on a medium containing $^{13}$C-enriched maple wood PyOM. The isotopic composition of the $CO_2$ produced by each of these isolates conclusively demonstrates the mineralization of PyOM. A similar approach has been taken to assay the potential of cultured isolates to degrade specific molecular components found in PyOM extracts (*i.e.* polyaromatic hydrocarbons and methylphenols). After sequencing the genomes of these 12 isolates, we aim to use isotope-enabled respiration studies to determine the primary pathways necessary for bacterial PyOM mineralization.

Together, these findings illustrate decisive steps toward identifying the microbial taxa involved in post-fire PyOM degradation and the molecular mechanisms they use to do so.

**References/Publications**


1. Fischer, M.S., F.G. Stark, T.D. Berry, N. Zeba, T. Whitman, and M.F. Traxler. (2021). Pyrolyzed substrates induce aromatic compound metabolism in the post-fire fungus, *Pyronema domesticum*. *Frontiers in Microbiology*, 12, 729289.

2. Berry, T.D., C. Creelman, N. Nickerson, A. Enders and T. Whitman. (2021). An open-source, automated, gas sampling peripheral for laboratory incubation experiments. *HardwareX,* 10, e00208.

3. Steindorff, A.S., A. Carver, S. Calhoun, K. Stillman, H. Liu, A. Lipzen, G. He, M. Yan, J. Pangilinan, K. LaButti, V. Ng, T.D. Bruns, and I.V. Grigoriev. (2021). Comparative genomics of pyrophilous fungi reveals a link between fire events and developmental genes. *Environmental Microbiology*, 23, 99-109.



**Funding Statement:** *This work was funded by the Department of Energy, Systems Biology Enabled Research on the Roles of Microbiomes in Nutrient Cycling Processes program, grant DE-SC0020351 to Thea Whitman, Thomas D. Bruns, Matthew Traxler, and Igor Grigoriev. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

**Title:** A community-based approach to understanding fungal and bacterial responses to wildfire

**Authors:** Nayela Zeba[1]* (nayelazeba@wisc.edu), Neem Patel[2]* (neempatel@berkeley.edu), Monika Fischer,[2] Timothy Berry,[1] Matthew Traxler,[2] **Thea Whitman[1]**

**Institutions:** [1]University of Wisconsin-Madison, Madison, WI; and [2]University of California-Berkeley, Berkeley, CA

**Website URL:** https://whitmanlab.soils.wisc.edu/research/; https://traxlerlab.berkeley.edu/

**Project Goals:** In this work, we aim to dissect the effects of microbes (fungal and bacterial) on carbon (C) and nitrogen (N) dynamics in post-fire forest soils. Our conceptual framework is rooted in systems biology and ecology, while our experimental approach combines genomics, transcriptomics, metabolomics, microbial community profiling, stable isotope techniques, small scale fire systems (pyrocosms), tightly controlled methods for producing labelled pyrogenic organic matter (PyOM), and high-throughput monitoring of C mineralization rates. We have three major research objectives: (1) To determine how dominant post-fire soil microbes affect the fate of PyOM; (2) To assess the interaction between N availability and PyOM mineralization by post-fire microbial communities and individual pyrophilous microbes; (3) To define the network of microbial interactions that facilitate PyOM breakdown over time and the key genes involved in this process.

**Abstract Text:** Forest wildfires in the western U.S. have been increasing in both frequency and size with each decade since the 1970s (Westerling, 2016). Understanding the effects of wildfires on microbial communities is required to predict their effects on processes such as biogeochemical cycling and plant recovery post-fire. We took a complementary field-laboratory approach to understanding how microbial communities respond to fire, and to probe the mechanisms controlling the effects of PyOM additions on soil organic carbon (SOC) cycling (often referred to as "priming") in a Californian mixed conifer forest.

For the field study, we undertook extensive field sampling of two controlled-burn plots (one high and one low-temperature burn) in parallel with two unburned control plots. Analysis of both the fungal and bacterial communities indicates non-neutral processes (*e.g.*, selection) play a key role in structuring the post-fire microbial community. For the laboratory study, we worked with soil from the same site, that was burned during the 2014 King Fire, adding PyOM to isolate the effects of PyOM addition on microbial communities and SOC cycling. To better elucidate the mechanisms of PyOM-SOC interaction, we exchanged water-extractable fractions of PyOM from $^{13}$C-labelled and unlabelled PyOM, produced from pine biomass grown in our custom labelling chamber and charred at 350°C and 550°C in our "charcoalator". This exchange of isotopically-labelled fractions allowed us to trace the water-extractable PyOM separately from the non-water-extractable fraction. We found that the water extractable fraction was the most mineralizable in both 350°C and 550°C PyOM compared to the SOC and non-water-extractable fraction. The mineralizability of the 350°C water-extractable PyOM fraction was higher than the

mineralizability of the 550°C water-extractable PyOM fraction. We observed short-term positive priming upon addition of 350°C PyOM to soil, most likely due to co-metabolism of easily mineralizable PyOM-C and the SOC. On the other hand, addition of 550°C PyOM to soil induced negative priming which could be attributed to the physical protection of SOC through sorption of SOC on the PyOM surface. We also examine the effects of PyOM additions on bacterial and fungal communities, and compare PyOM-responsive taxa with those identified in the field study.

**References/Publications**

Westerling ALR. 2016. Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring. *Phil. Trans. R. Soc.* B 371: 20150178.

**Funding Statement:** *This work was funded by the Department of Energy, Systems Biology Enabled Research on the Roles of Microbiomes in Nutrient Cycling Processes program, grant DE-SC0020351 to Thea Whitman, Thomas D. Bruns, Matthew Traxler, and Igor Grigoriev. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

# $N_2O$ Formation and Organic Nitrogen Utilization in Wetland Microbial Communities

Juliet Johnston[1*](johnston50@llnl.gov), James Britt Abrahamson[2*], Zach Flinkstrom[2*], Stephany Wei[2*], Xiangpeng Li[3], Hanyan Li[3], Michael Wells[3], Long Ngo[3], Eunkyung Choi[4], Xianhui Wan[5], Xin Sun[5], David Stahl[2], Peter K. Weber[1], Xavier Mayali[1], Chongle Pan[3], Wei Qin[3], Willm Martens-Habbena[4], Bess B. Ward[5], **Mari Winkler[2]**
*Denote co-first authors

1. Lawrence Livermore National Laboratory, Livermore, California 2. University of Washington, Seattle, Washington 3. University of Oklahoma, Norman, Oklahoma; 4 University of Florida, Davie, FL; 5 Princeton University, Princeton, New Jersey.

**Project Goals:** Wetlands capture and release large amounts of greenhouse gases ($CO_2$, $CH_4$ and $N_2O$) and it is of prime importance to predict their response to climate change induced stressors such as drought and sea level rise associated saltwater intrusion. This project aims to link wetland microbial activities to ecosystem-scale processes by developing a reproducible experimental model for lacustrine and estuarine wetland ecosystems to quantify responses to controlled manipulations representing climate impacts. Hydrogel beads, controllable in size, with entrapped wetland microbes and soil and plant-like carbon, act as models for sediment aggregates. Bioreactors with real-time gas and liquid metabolite flux monitoring, integrated multi-omics analyses, and stable isotope tracing will be conceptually incorporated into mathematical models to predict how climate change stressors impact C and N fluxes across different wetland spatial and temporal scales.

**Overview:** Nitrous oxide ($N_2O$), a potent GHG (265 times stronger than $CO_2$), is released in wetlands by heterotrophic denitrification as well as ammonia-oxidizing organisms (AOO). The three types of AOO are 1) ammonia oxidizing bacteria (AOB) which are prevalent in ammonium laden soil and emit the largest flux of $N_2O$ 2) ammonia oxidizing archaea (AOA), which are ubiquitous in oligotrophic environments and 3) comammox bacteria, which were recently discovered and have redefined our understanding of nitrification by performing both ammonia oxidation and nitrite oxidation. Due to their novelty, the contribution of $N_2O$ emissions from comammox bacteria remains under explored. Previous AOO investigations have focused on ammonia as the primary N substrate, despite organic nitrogen compounds being more prominent in wetland ecosystems. To better understand $N_2O$ emissions in wetlands from AOOs as well as their organic nitrogen utilization patterns we **a)** explored the role of aerobic and anaerobic comammox metabolisms on $N_2O$ emissions and **b)** studied the physiology and regulations controlling the use of alternative nitrogen sources (inorganic and organic) by different AOOs. In addition, we also studied the effect of drought events on $N_2O$ emissions originating from nitrification and denitrification events (**c**).

### (a) $N_2O$ emissions from comammox bacteria under aerobic and anaerobic conditions

We confirmed aerobic $N_2O$ production by comammox and could reproduce the recently published stoichiometry of $N_2O$ emitted per ammonium oxidized (*1*). In addition, we studied the use of alternative electron donors and acceptors by comammox. We measured $N_2O$ production during anaerobic comammox incubations and are currently investigating if it is the result of an abiotic conversion. These observations suggest that nitrifier induced $N_2O$ production could be induced by comammox under physiologically relevant conditions. The results also indicate the role of comammox within environmental systems might be more complicated than previously thought.

### (b) Physiological Studies of the use of alternative nitrogen sources in AOO

We characterized the anabolic incorporation and metabolic respiration of different inorganic and organic nitrogen sources in different AOO isolates, to determine N resource utilization and regulation among phylogenetically distinct AOOs. Isolates were fed organic and/or inorganic nitrogen substrates in batch

incubations. The organic nitrogen, ammonia, and nitrite concentrations were measured over time to determine substrate utilization. Stable isotope ($^{15}N$ and $^{13}C$) labeled organic nitrogen, ammonia, and bicarbonate were used to determine cellular respiration products and intermediates via Isotope-Ratio Mass Spectrometry (IRMS) and biomass incorporation using Nanoscale Secondary Ion Mass Spectrometry (NanoSIMS). Significant differences in N drawdown patterns were observed among different AOO isolates. For example, organic N was consumed rapidly in $\beta$-AOB, being drawdown at higher rates than inorganic N. We confirmed the selective incorporation of organic-N over inorganic-N by $\beta$-AOB with NanoSIMS. In contrast to $\beta$-AOB, soil AOA preferentially used inorganic-N over organic-N. These observations suggest that phylogenetically distinct AOO have different mechanisms for nitrogen transport and utilization and challenge the assumption that ammonia N is always the primary substrate for ammonia oxidizers growth. Stable isotope tracking of $^{15}NO_2^-$ production is in progress to determine the dynamics in different N-substrates utilization for each AOO. Results from microrespirometry showed the differences in affinities amongst different AOO and nitrogen sources. RNA-seq and proteomics analyses are underway to infer regulatory mechanisms, which will lead to a better understanding of nitrogen partitioning and ecological niches across different nitrifiers.

**(c) $N_2O$ emissions at varying water table submergence**

To understand the competition among nitrifiers and their interplay with denitrifiers in terrestrial wetlands (Lake Washington, WA) we conducted laboratory soil column experiments. This experimental model system was exposed to water table changes to understand the effect of climate change induced drought or flooding/inundation on globally significant microbially-mediated lacustrine wetland nitrogen cycling. GHG flux measurements indicate that the wetland soil columns were an $N_2O$-neutral system when submerged, but shifts to become an $N_2O$ source (0.5 nmol $s^{-1}$ $m^{-2}$) when the water table decreases. As a next step, we will encapsulate the wetland community in hydogels to generate a simplified microbial system. These results will be used to develop predictive models that link microbial physiology of terrestrial systems to overall ecosystem processes (e.g., nitrate leaching and $N_2O$ emissions). The model will be used to simulate how climate change induced environmental changes such as different redox gradients, e.g. the change in oxygen concentration on a micrometer scale within individual hydrogels, impact GHG emissions, microbial abundance, and activities.

**Funding Statement:** *This study is supported by the Genome Sciences Program of the Office of Biological and Environmental Research under project FWP SCW1677. Part of this work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344. Part of this work was funded through Florida Agricultural Experiment Station (Hatch project FLA-FTL-005680) and UF IFAS Early Career award to W.M.H.*

## References

1. K. D. Kits *et al.*, Low yield and abiotic origin of N2O formed by the complete nitrifier Nitrospira inopinata. *Nature Communications* **10**, 1836 (2019).

**Experimental models bridging single cell-to-ecosystem scales to evaluate climate-wetland feedback mechanisms**

Dongyu Wang[2*], Zach Flinkstrom[1*], Pieter Candry[1*], Bruce Godfrey[1], James Britt Abrahamson[2], Juliet Johnston[3], Sam Bryson[1], Ludmila Chistoserdova[1], David Stahl[1], Chongle Pan[2], Xavier Mayali[3], and Mari Winkler[1*] (mwinkler@uw.edu)
*Denote co-first authors
[1] University of Washington, Seattle, Washington, USA
[2] University of Oklahoma, Norman, Oklahoma, USA
[3] Lawrence Livermore National Laboratory, Livermore, California, USA

**Project Goals:** Wetlands capture and release large amounts of greenhouse gases ($CO_2$, $CH_4$ and $N_2O$) and it is of prime importance to predict their response to climate change induced stressors such as drought and sea level rise associated saltwater intrusion. This project aims to link wetland microbial activities to ecosystem-scale processes by developing a reproducible experimental model for lacustrine and estuarine wetland ecosystems to quantify responses to controlled manipulations representing climate impacts. Hydrogel beads, controllable in size, with entrapped wetland microbes and soil and plant-like carbon, act as models for sediment aggregates. Bioreactors with real-time gas and liquid metabolite flux monitoring, integrated multi-omics analyses, and stable isotope tracing will be conceptually incorporated into mathematical models to predict how climate change stressors impact C and N fluxes across different wetland spatial and temporal scales.

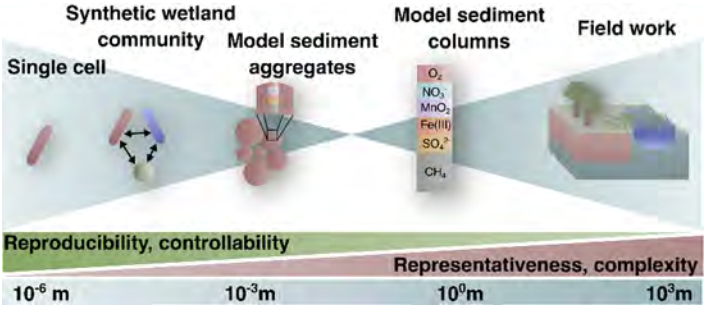



**Figure 1** Bottom-up and top-down approaches to begin bridging the microbe-climate scale divide.

**Abstract Text:** This work aims to construct reproducible experimental model systems to test the relationships between macro-scale climate stressors and processes at the microbial scale to understand future trajectories of wetlands as carbon cycle nodes (Figure 1).

Estuarine wetlands are expected to experience significant alterations of greenhouse gas fluxes due to the introduction of sulfate ($SO_4^-$) ions that may promote sulfate reduction over methanogenesis for the decomposition of organic matter. We have developed a four-member synthetic community (Figure 2) converting cellulose to $CH_4$ and $CO_2$ representative of estuaries at the edge of transition towards saline systems. Using *Desulfovibrio vulgaris* as key node in the community enables investigating the role of sulfate on community dynamics in a simplified system. Proteomics of the four-member community grown with and without sulfate revealed the impact of sulfate on the population size and functional behaviors. The increased population size of *Desulfovibrio vulgaris* and the enhanced abundance of key functional proteins in the sulfate reduction pathway indicated that the addition of sulfate promoted sulfate reduction. Reduced population size of *Clostridium cellulolytiucm* and the abundance of identified cellulolases suggested a reduced cellulose degradation after being treated by sulfate. There were no significant changes on the population sizes of two methanogens between two conditions, but the sulfate addition significantly reduced the protein abundance of enzymes in the acetoclastic methanogenesis pathway in *Methanosaeta concilii* and the hydrogenotrophic methanogenesis pathway in *Methanospirilum hungatei.* This was consistent with the observed reduction in the methane production in the sulfate addition condition. Moving towards higher-scale processes, diffusion limitations become more important. We have developed an *in silico* biofilm model of *D. vulgaris* and *Methanospirillum hungatei* to

investigate the role of electron donor (lactate) and electron acceptor (sulfate) concentration gradients on the spatial organization of methanogenic syntrophy by these partners. In parallel, we have entrapped a cellulose degrader (*Clostridium cellulolyticum*) and the four-member synthetic community in hydrogel-based synthetic sediment aggregates to investigate how organisms and communities access solid carbon substrates in diffusion limited environments. Connecting these experiments with nanoSIMS and SIP-proteomics provides a bottom-up experimental model connecting single cells to millimeter-scale processes.

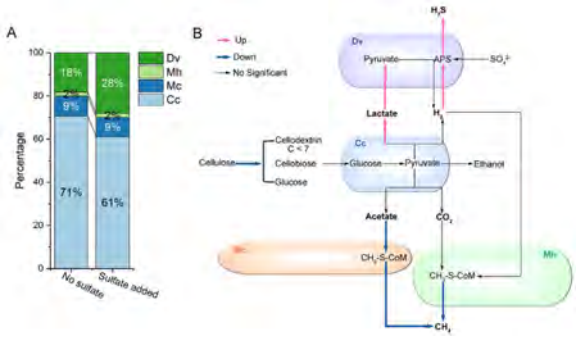



**Figure 2** Metaproteomics analysis of the synthetic community. Dv, *Desulfovibrio vulgaris,* Mh, *Methanospirilum hungatei* Mc, *Methanosaeta concilii* Cc, *Clostridium cellulolytiucm*

Freshwater wetlands are expected to experience longer droughts and more extreme floods in the future. To investigate the consequences of these impacts on GHG fluxes from freshwater wetlands, we are using a top-down approach from field sites to spatially explicit lab-scale model systems. In a wetland field site on Lake Washington (a managed, freshwater lake), we measured *in situ* surface GHG fluxes as well as depth-resolved metabolite (e.g., ammonia, sulfate, sulfide, organic acids, etc.) and microbial community profiles. Field campaigns across seasons and associated water table changes will show the extent to which the GHG fluxes are controlled by temperature and varying level of submergence. Our results suggest that disturbance of anaerobic conditions, can increase decomposition of organic matter in soils and continue to promote the release of GHGs even if the wetland's condition is restored. In parallel, we are investigating the effect of water table changes on GHG emissions in a more reproducible, on-line monitored bioreactor systems, enabling long-term non-destructive measurements of metabolite, gas, and microbial community depth profiles throughout a 20-centimeter soil column. Preliminary results show that lowering water tables triggers a massive transient $CH_4$ efflux (100-fold higher than before disturbance), observations that may be due to both physical and biological responses to water table drops. Conversely, increasing water tables drastically lowered $CH_4$ fluxes in the system. Such lab-scale experimental systems provide powerful templates to design simple experiments answering questions on centimeter-scale community organization, carbon conversion and meter-scale process outcomes. These columns will be sampled at different column depth for transcriptomic and/or proteomic analyses to better understand the key players contributing to carbon release.

Towards the future, these bottom-up and top-down experimental models will converge in a hydrogel-based sediment column bioreactor. This will enable leveraging the single-cell to millimeter-scale strengths of the hydrogel model system, while providing the centimeter-to-meter-scale process heterogeneity needed to appropriately investigate the responses of wetland ecosystems to climate disturbances.

**Funding Statement:** *This study is supported by the Genome Sciences Program of the Office of Biological and Environmental Research under project FWP SCW1677. Part of this work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344.*

**Title:** Genomes to Ecosystem Function: Targeting Critical Knowledge Gaps in Soil Methanogenesis and Translation to Updated Global Biogeochemical Models

**Authors:** Angela M. Oliverio[1,2*] (Angela.Oliverio@colostate.edu), Adrienne B. Narrowe[1+], Mikayla A Borton[1,3], Jorge A. Villa[4], Rebecca A. Daly[1], Pengfei Liu[1], Bridget B. McGivern[1], Michael Shaffer[1], Rory Flynn[1], Christopher S. Miller[4], Trent R. Northen[5], David W. Hoyt[3], William R. Riley[5], Timothy H. Morin[7], Gil Bohrer[8], **Kelly C. Wrighton**[1]

**Institutions:** 1 Colorado State University, Fort Collins; 2 Yale University, New Haven; 3 Pacific Northwest National Lab, Richland, WA; 4 University of Colorado, Denver; 5 Lawrence Berkeley National Lab, Berkeley, CA; 6 University of Louisiana, Lafayette; 7 State University of New York ESF, Syracuse; 8 The Ohio State University, Columbus

**Project Goals:**

Despite their relatively small land coverage, wetlands represent the largest source of atmospheric methane (20-40%). However, variations in these wetland emission budgets are high, with over 25% uncertainty. Accurately predicting net methane fluxes from these systems depends on multiple interacting geochemical, ecological, and metabolic constraints that are poorly understood, oversimplified, or missing in global biogeochemical models. The overarching objective of this early career proposal is to identify the biogeochemical and genomic determinants impacting methane production, including the scale and physical distribution over which they operate, along freshwater wetland gradients.

The project goals can be summarized into three main objectives:

1) Characterize wetland variables affecting methanogen distribution and activity.

2) Quantify fine scale oxygen dynamics in soils and the impact on methane emissions.

3) Improve model capability to include a representation for small-scale processes.

**Abstract Text:**

Freshwater wetlands are important contributors to ecosystem services and mediate global biogeochemical cycles with important contributions to soil carbon sequestration and greenhouse gas (GHG) flux. Toward our goal of understanding controllers on microbial carbon decomposition in soils, we constructed a greenhouse gas (GHG) observatory at Old Woman Creek, a freshwater wetland in the Ameriflux network with some of the highest annual methane fluxes [1]. To track the end products of microbial carbon decomposition we measured methane and carbon dioxide production at cm scale depths in soils and GHG fluxes from soils at meter and wetland scales. We defined the methane and carbon dioxide production hotspots occurring within the top 50 cm of the soil profile and extended these data across multiple months and years. Paired to these data we collected over 600 microbial samples across the wetland including genome-resolved metagenomics with metatranscriptomes, 16S rRNA gene sequencing, and metabolomics (NMR, LC-MS, and FT-ICR MS). During our 5-year sampling period, climatically-driven water level rise in the adjacent Lake Erie flooded the wetland by ~3-6 ft, with this hydrological inundation serving as a natural laboratory for interrogating how climatic

induced shifts in hydrology impact microbial metabolism to regulate soil carbon decomposition and flux.

Essential to our microbial analyses, we constructed a first of its kind genomic catalog of freshwater soil microbial communities, resulting in over 13,000 genomes spanning 64 bacterial and 13 archaeal phyla. To first understand the microbial processes directly regulating soil methane, we genomically inventoried 300 genomes from methanogen and methanotroph lineages, defining a new family and 12 new genera. We coupled this genomic catalog to metatranscriptome data from spatially and temporally distributed samples along with geochemical, metabolomic, and gas flux measurements. Here we provide evidence that contrasts with widely held paradigms about the soil carbon cycle: (1) methanogenesis using methylated organic nitrogen and sulfur substrates, not acetate or hydrogen, best explained soil methane production, (2) soil oxygen concentrations or depth are not effective approximates for the activity of methanogens or methanotrophs in soil cores [2,3], and (3) redox changes caused by sustained flooding, did not alter the methanogenic substrate pools or the metabolic networks governing methane production. Collectively, these results highlight the soil microbial metabolisms influencing the terrestrial methane cycle, thereby offering direction for increased realism in predictive process-oriented models of methane flux in wetland soils.

To better understand the nutrient and carbon metabolic networks that control soil carbon stability and its decomposition into GHGs, we used our extensive metatranscriptome data set to determine coordinated gene expression networks across season and depth and correlated these networks to key ecosystem chemical features. Our analysis revealed cohorts of metabolically active microorganisms and their associated metabolites that persisted across months and even years, impervious to different soil geochemical conditions. For example, we found the same microbial network predicted soil carbon dioxide and methane production, as well as sulfate concentrations. This active soil organic matter decomposing cohort was comprised of soil carbon respiring members that produce substrates for simple carbon utilizing sulfate reducing bacteria, which generate the methanogenic substrates to support active methane cycling microorganisms. The strong co-occurrence between sulfate-reducers and methanogens is of particular interest, as it is often thought that sulfate reduction is a thermodynamic constraint on methane production in wetland soils. Alternatively, we suggest these groups act syntrophically, as the high dissolved organic carbon is not a competitive constraint, and thus sulfate reducers oxidize lactate and ethanol to hydrogen and acetate, which cross-feeds methanogens. We are now using field-designed reactor studies amended with labeled substrate to validate and probe the resiliency of these carbon decomposing metabolic networks.

While we have demonstrated that microbial genome resolution does increase predictions of emergent ecosystem properties like soil methane production, omics data are not easily ingested into terrestrial ecosystem biogeochemistry models (e.g., *ecosys*). These models incorporate data across ecosystem layers including climate, macrofauna, chemistry, and microbes including microbial functional guilds. Therefore, we developed a software application to summarize individual genome-resolve transcriptomic data into functional guilds using DRAM (Distilled Refined Annotation of Metabolism) [4,5], which was developed as part of this early career award. We demonstrated that DRAM identifies trait-level information by leveraging annotation

and metabolism content from microbial genomes and expression patterns and translates this into "functional guilds".

To test if we could predict the observed variability in soil carbon dioxide and methane using only our functional guild data, we built random forest models and evaluated the importance of predictors (e.g., functional guilds) by determining the decrease in prediction accuracy between observations and predictions. For soil methane production, the strongest guild predictors were methanogens, homoacetogens, and fermenters, while aerobic heterotrophs and nitrifiers predicted carbon dioxide production. Lastly, these explanatory models and coordinated gene expression networks informed a conceptual, community-based metabolic model identifying possible abiotic and biotic control points on the interconnected microbial metabolisms that control carbon dioxide and methane production. Our analyses suggest that translating complex microbial data into a 'guild' framework may represent a path to readily incorporate highly dimensional multi-omics into terrestrial ecosystem biogeochemistry models while still maintaining biological integrity.

**References**

1. Rey-Sanchez AC, TH Morin, KC Stefanik, KC Wrighton, G Bohrer 2018. Determining total emissions and environmental drivers of methane flux in a Lake Erie estuarine marsh. *Ecological Engineering*.114, 7-15

2. Angle JC, TH Morin, LM Solden, GJ Smith, AB Narrowe, MA Borton, RA Daly, DW Hoyt, WR Riley, CS Miller, G Bohrer, KC Wrighton 2017. Methanogenesis in oxygenated soils is an unrecognized driver of wetland methane emissions. *Nature Communications*, (8)1, 1567

3. Smith GJ, JC Angle, LM Solden, MA Borton, TH Morin, RA Daly, MD Johnston, KC Stefanik, R Wolfe, G Bohrer, KC Wrighton 2018. Members of the Methylobacter are inferred to account for the majority of aerobic methane oxidation in oxic soils from a freshwater wetland. *mBio* 9 (6) e00815-18.

4. Shaffer M, MA Borton, BB McGivern, AA Zayed, SL La Rosa, LM Solden, P Liu, AB Narrowe, J Rodriguez-Ramos, B Bolduc, C Gazitua, RA Daly, GJ Smith, DR Vik, PB Pope, MB Sullivan, S Roux, KC Wrighton 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*.

5. Rubinstein RL, MA Borton, H Zhou, M Shaffer, DW Hoyt, J Stegen, CS Miller, C Henry, KC Wrighton, R Versteeg. ORT: A workflow linking genome-scale metabolic models with reactive transport codes. BioRxiv https://doi.org/10.1101/2021.03.02.433463

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018022. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

**Enabling Structure Determination of Challenging Samples with new cryo Electron Microscopy Methods**

Jose A. Rodriguez [1], David Eisenberg,[1] Logan Richards,[1] Ambarneil Saha,[1] Matthew Agdanowski,[1] Roger Castells-Graells (rcastellsg@ucla.edu),[1] and **Todd O. Yeates**[1]

[1]UCLA-DOE Institute for Genomics and Proteomics; University of California, Los Angeles. Los Angeles, CA 90095

https://www.doe-mbi.ucla.edu/ucla-doe-institute-missions/

**Project Goals:**

Research in the UCLA-DOE Institute for Genomics and Proteomics (IGP) includes major efforts in the area of imaging science, with key applications in microbial biosystems, from genomics to function. Our team is pioneering new enabling capabilities that facilitate the discovery of molecular structural features affecting protein function and specificity. These capabilities span the broad areas of X-ray diffraction, electron microscopy, and micro-electron diffraction, along with protein engineering and selection methods designed to advance those techniques. Emerging MicroED techniques present challenges in processing, refinement and phasing of diffraction data that our team is tackling in part by enabling rapid access to robust public-facing tools (webservers) that facilitate ED structure determination. Our team is also making critical advances to resolve protein structures smaller than about 50 kDa in size by cryo-EM. Success in that goal will be transformative for applying cryo-EM to broad problems in bioenergy science.

**Abstract:**

*Breakthroughs in cryo-EM –*
Cryo-electron microscopy has become a major tool for atomic structure determination. While its power has been widely demonstrated for analyzing large macromolecular complexes, signal-to-noise challenges make it difficult or impossible to resolve the structure of smaller proteins by cryo-EM. With a practical lower limit around 50 kDa, a large fraction of cellular proteins of interest remain outside the scope of this leading structural method. In recent work, our DOE-UCLA IGP team has broken through this barrier by engineering novel scaffolds with sufficient rigidity and modularity to achieve resolution useful for interpreting atomic structure. These unique scaffolds are based on designed protein cages fused to Darpins as modular protein adaptors. We will discuss soon-to-be published work, where we have created a second generation of scaffolds that are rigid enough to reach the critical resolution of 3 Å (Fig. 1). Our scaffolding approaches should open up EM-based structural studies on wide-ranging proteins and enzymes with bioenergy importance. Applications to select enzymes involved in cellulosic breakdown are planned, along with future applications of cage-based scaffolds as markers in cryo-EM tomography.

*Enabling micro-ED methods -*
Electron diffraction is reemerging as a frontier method for atomic structure determination from three-dimensional microcrystalline proteins and peptides. With micro-electron diffraction

(MicroED), the ability to work with very small crystals circumvents the most common obstacle to successful x-ray crystallographic work: obtaining large well-ordered crystals. However, other distinctions of the MicroED method bring up new challenges that need to be overcome in order for the method to reach its potential. Our UCLA-DOE IGP team is making important inroads on multiple key challenges in both crystal imaging and diffraction.

(i) We are developing new atomic scattering parameterizations, which are distinctly different from X-ray scattering profile for charged atoms, in order to clarify the important issues of charge states using MicroED. We have developed a public web server to make those parameters fully accessible at: https://srv.mbi.ucla.edu/faes/. This resource enables Gaussian parameterization of elastic electron scattering factors in a form amenable to refinement in the program Phenix. Current applications will be discussed.

(ii) For biomolecules that are within the size range of total synthesis (which includes many small proteins), we are applying methods of racemic crystallography, where phasing challenges are greatly mitigated, and for larger molecules with potentially novel folds, we are taking advantage of fragment-based phasing approaches (Fig. 2) and, in collaboration with PNNL, direct crystal imaging.

We are also developing and updating tools for validating and assessing the quality of three-dimensional protein structure models. That important topic is reemerging as a major challenge and opportunity in wake of the coming flood of predicted models from recent advances in machine-learning-based structure prediction.

Collectively, the enabling capabilities we are developing will broadly facilitate the determination and refinement of unknown macromolecular structures with importance for bioenergy.

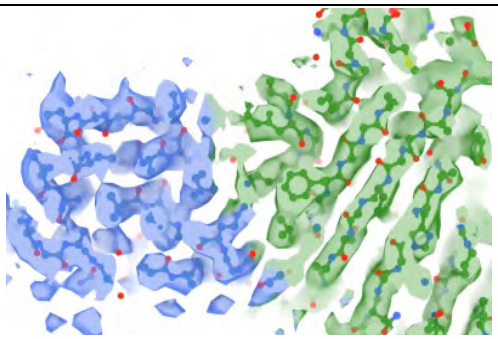


**Figure 1.** A cryo-EM map of a small protein (26 kDa GFP) bound to its modular adaptor (Darpin). The small cargo protein (green) has a resolution of 3.0 Å. The large designed cage that serves as the scaffold is not shown.

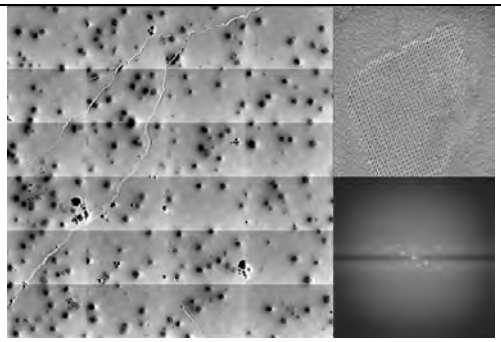


**Figure 2.** A field of cell-grown frozen-hydrated nanocrystals (left) is surveyed for tomographic reconstruction of a single crystal (right, top), whose Fourier Transform shows a lattice (right, bottom).

## References

1. M.C. Thompson, T.O. Yeates, J.A. Rodriguez (2020). Advances in methods for atomic resolution macromolecular structure determination. *F1000Res* **9**: doi: 10.12688/f1000research.25097.1.

*This work is supported by DOE Grant DE-FC02-02ER63421.*

**Metabolism in Microbial Communities and the Associated Biochemistry of Polymer Deconstruction**

**Authors:** Janine Fu[1], Brendan Mahoney[1], John Muroski[1], Allen Takayesu[1], Orlando E. Martinez[1], Marco Morselli[1], Matteo Pellegrini[1], Rachel R. Ogorzalek Loo[1], Katherine Chou[2], Robert Clubb[1], Joseph A.. Loo[1] and Robert Gunsalus[1]* (robg@microbio.ucla.edu ) and **Todd Yeates[1]**

**Institutions:** [1]UCLA DOE Institute, University of California, Los Angeles, Los Angeles, CA; [2] National Renewable Energy Laboratory, Bio Science Center, Golden, CO.

https://www.doe-mbi.ucla.edu/

**Project Goals:** Our microbiology team project within the UCLA DOE Institute employs a coordinated set of molecular and *in silico* approaches to examine model microbial communities and their component parts to better understand the processes that drive anaerobic carbon recycling in nature. These processes impact multiple areas of BER interest including bio-conversions of model substrates in natural and manmade environments, the associated biochemistry of key degradative enzymes and in the design of plant-based biomass deconstruction strategies for biofuel production. We are elucidating the undocumented metabolic pathways in syntrophic communities for model substrates with focus on key enzymes and associated oxidation reactions. We are developing next-gen omics methods to interrogate environmentally relevant pathways and interactions among organisms within microbial communities, and testing these proposed functions where possible. Using the model cellulolytic microorganism, *Clostridium thermocellum*, we are examining how anaerobic microbes synthesize and assemble their extracellular cellulosome structures that degrade lignocellulose.

**Abstract text:** Major activities within the UCLA-DOE Institute in the past year deal with three core areas of investigation.

*Elucidation of syntrophic microbial pathways for metabolism of model substrates.* Genomic, proteomic and informatic studies were performed on defined microbial communities to elucidate how representative fatty acid substrates are metabolized syntrophically. Key pathway enzymes were identified and characterized in two strains, *Syntrophomonas wolfei* and *S. wolfei* sub sp methylbutyratica for short and branched chain fatty acids along with the supporting electron transport reactions. Recombinant and structural studies of ten additional enzymes of the carbon oxidation pathway in *S. wolfei* were performed to further explore the biochemical basis for the thermodynamic limiting steps occurring during syntrophic cell growth. The resulting structures were also employed for subsequent modeling of acyl-lysine modifications of syntrophic pathway enzymes.

In companion studies, PacBio long read sequencing approaches are being used to sequence, assemble and annotate genomes of five previously unstudied syntrophic bacterial strains that utilize other model substrates when grown in co-culture with suitable methanogen partners. We are extending the gene annotation methods beyond the standard homology-based interferences to those based on co-evolution such as phylogenetic profiling, phenotypic profiling and operon conservation with the goal of supporting microbial pathway prediction and modeling.

*Acyl-lysine modification of syntrophic pathway proteins.* Proteomic and mass spectrometry studies were performed to further characterize protein post-translational modifications of carbon and electron transfer pathway enzymes in our model syntrophic strains. As protein modification can affect enzyme activity, these data will decipher their relationship with the metabolism of syntrophic microbial communities. Acyl-lysine modifications, which can arise from reactive metabolites, were strikingly found in high abundance in the proteome of model syntrophic bacteria. In *S. aciditrophicus*, *S. wolfei*, and *S. wolfei* sub sp *methylbutyratica*, multiple types of acyl-modifications were identified, including three types not reported before in any other organism. We have also identified the proteins modified and their sites of modification. Our recent work has shown that the type and relative abundance of these modifications do significantly change in response to different carbon sources, correlating with metabolic bottleneck points in the microbes' degradation pathway. These findings give us a glimpse at how thermodynamically-challenged organisms employ reversible catalysis to survive.

*Cellulosome assembly and display in cellulolytic anaerobic bacteria.* In companion microbial studies we are investigating how highly cellulolytic anaerobic bacteria synthesize, assemble and display cellulosomes. *Clostridium thermocellum,* a model bacterium capable of directly converting cellulosic substrates into ethanol and other biofuels is being used to investigate how the cell fine-tunes the enzyme composition of its cellulosome using anti-σ factors to control gene expression in response to sensing extracellular polymers. Our recent studies have shown that the RsgI9 anti-σ factor interacts with cellulose via a C-terminal bi-domain unit. A 2.0 Å crystal structure reveals that the unit is constructed from S1C peptidase and NTF2-like protein domains that contain a potential binding site for cellulose. Small angle X-ray scattering experiments of the intact ectodomain indicate that it adopts a bi-lobed, elongated conformation. In the structure a Conserved RsgI Extracellular (CRE) domain is connected to the bi-domain via a proline-rich linker, which is expected to project the carbohydrate binding unit ~160 Å from the cell surface. The CRE and proline-rich elements are conserved in several other *C. thermocellum* anti-σ factors, suggesting that they will also form extended structures that sense carbohydrates. We hypothesize that cellulolytic anaerobic bacteria assemble and display cellulosomes using a conserved molecular pathway. We are employing *in silico* comparative genomics approaches to identify conserved pathway components whose functional importance is being assessed in *C. thermocellum*. The results of these studies will provide new insight into anaerobic carbon recycling by naturally cellulolytic bacteria and could guide rational engineering efforts to create microbes that are capable of converting of plant biomass into biofuels, materials and chemicals.

**References/Publications:**

Muroski JM, Fu JY, Nguyen HH, Ogorzalek Loo RR, and Loo JA. Leveraging immonium ions for targeting acyl-lysine modifications in proteomic datasets. Proteomics. 2021 Feb;21(3-4).

Muroski JM, Fu JY, Nguyen HH, Wofford NQ, Mouttaki H, James KL, McInerney MJ, Gunsalus RP, Loo JA, Ogorzalek Loo RR. The acyl-proteome of *Syntrophus aciditrophicus* reveals metabolic relationship with benzoate degradation. (in revision).

Mahoney BJ, Takayesu A, Zhou A, Cascio D, Clubb RT. The structure of the *Clostridium thermocellum* RsgI9 ectodomain provides insight into the mechanism of biomass sensing. (submitted).

***Funding statement:*** *This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-FC-02-02ER63421.*

**Title:** Automated Knowledge Harvesting from Literature Text, Tables, and Figures

**Authors: Shinjae Yoo,[1,*] (sjyoo@bnl.gov),** Ian Blaby[2], Sean McCorkle[1], Gilchan Park[1], and Carlos Soto[1]

**Institutions:** [1] Computational Science Initiative, Brookhaven National Laboratory, Upton, NY; [2]Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA

**Project Goals:** The overall goal of this project is to develop tools and techniques enabling the efficient extraction of protein-associated information from large volumes of literature. This high-level goal is divided into three core objectives: i) develop machine learning (ML) methods to visually process documents which are not in machine-readable formats; ii) use natural language processing (NLP) to identify protein relations described in text; iii) leverage machine learning techniques to demonstrate automatic extraction of structure and information from tables and figure embedded in documents.

**Abstract:** A significant knowledge gap currently exists between sequenced genomes and the cellular function of the encoded proteins. This gap is growing as sequencing techniques accelerate while gene function-validating experiments continue at a slower pace. There is substantial cost (financial and time) in investigations seeking to capitalize on genome-enabled organisms by biological redesign to meet BER goals. Therefore, the automated -- and up-to-date with the current literature -- annotation of target genes is essential. Current techniques for managing this resource are inadequate: keyword-based search is largely limited to hand-picked terms or at best the contents of the abstract; reference crawling helps to expand a query, but not to refine it. Consequently, at present the most reliable functional annotations in databases are manually curated, which clearly cannot keep pace with the ever-growing body of literature. Moreover, much of the scientific contents of a publication are found within tables and figures, which are all but ignored by current literature search techniques. In this work, we use machine learning (ML) and natural language processing (NLP) techniques to move past these limitations and develop new tools to harvest knowledge from the literature at scale.

We identified several challenges to the goal of scalable scientific literature mining for functional genomics: full-text document processing; non-machine-readable formats; inconsistent gene and protein identifiers; semantic ambiguity and complex relationship ontologies; scale and diversity of table and figure structures and contents; and extensible knowledge representations. For this year, **we focus on two major subproblems**, their associated challenges, and our approaches, methods, and results in addressing them in this work: 1) NLP for **identifying protein entities and relationships (i.e., regulation) between them in the text**, and 2) ML for **automated information extraction from tables and figures**.

Due to broad inconsistencies in the in-text gene/protein identifiers found within the literature, a simple dictionary approach would not suffice for seeking textual evidence of relationships between these entities. We therefore used NLP techniques to train a named entity recognition (NER) model

specialized in identifying mentioned of genes and proteins in the main-body text of biology articles. We then built upon this NER model to develop and train an entity-relationship model that identifies a refined set of relationships from the semantics of the textual evidence surrounding identified gene/protein entities. Based on our last success of PPI (protein-protein interaction) inference from literature, we leverage multisource gene databases and refine their labels to train natural language processing (NLP) models for predicting ***gene regulatory relations***. Our best model achieves 90.04 F1 performance and we have used it to identify over 200 promising regulation candidates for our demonstration organism of Pseudomonas putida. We are currently working to combine other evidence sources (i.e., expression profiles) to further improve regulatory relation detection.

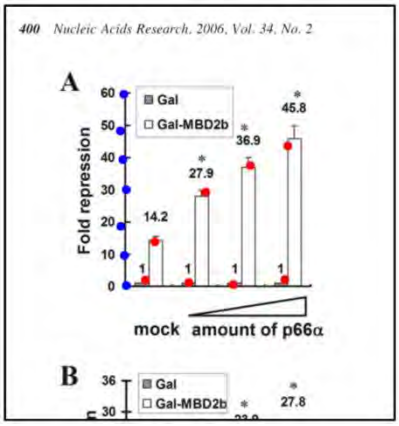



*Figure 1 the predictions are bar peaks (red), and data axis values (blue). The results show the robust Figure detection results on over-cropped images.*

Although tables and figures often contain much of the scientific contents in research publication, the information contained in these has largely remained opaque to automated information extraction techniques. To address this opening, we are adapting existing ML techniques as well as developing new ones to identify and isolate tables and figures of relevance, as well as to extract their structure and contents. We are building upon semantic segmentation ML methods to accurately capture the structure and contents of document-embedded tables, after which we may apply NLP techniques to process the text contents. We identified bar charts as a case study for **demonstrating the novel ability to identify data plots of interest and automatically extract the data values they contain**. For this purpose, we developed a novel point proposal network (PPN) for automatically identifying data points in charts and graphs. Our PPN model efficiently leverages synthetic data during training and achieves 87.05 F1 on real scientific bar charts from unseen literature; we are working on extending our PPN algorithm to other chart types such as pie chart. We also significantly improved Table structure detection accuracies, compared to the last year by leveraging direct cell structure inference instead of regular cell structure prediction, in order to better handle irregular table structures.

Table 2. Folate profiles of *E. coli* strains

| Strain | Folates, pmol mg$^{-1}$ protein* | | | | |
|---|---|---|---|---|---|
| | THF | $CH_2$-THF | CH=THF + 10-CHO-DHF† | 5-CHO-THF | Total |
| Wild type | 48.1 ± 10.7 | 10.6 ± 1.9 | 738 ± 93 | 68.9 ± 10.9 | 866 ± 114 |
| Δ*folB* | <0.05 | <0.05 | <0.05 | <0.05 | <0.2 |
| Δ*folP* | <0.05 | <0.05 | <0.05 | <0.05 | <0.2 |
| Δ*gcvP*-*aghyA* | 845 ± 171 | <0.05 | <0.05 | <0.05 | 845 ± 171 |
| Δ*folE*Δ*thyA* + 5-CHO-THF‡ | 152 ± 100 | 7.1 ± 0.7 | 14.4 ± 3.5 | 5.8 ± 1.5 | 180 ± 98 |

This project aims to provide biologists with new tools to accelerate their work and to discover promising new directions of research informed by the wealth of knowledge buried in the published literature.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012704.

**Accelerating Strain Phenotyping with Desorption Electrospray Ionization-Mass Spectrometry Imaging and Untargeted Molecular Analysis of Intact Microbial Colonies**

**Authors:** Berkley M. Ellis,[1] Hawkins S. Shepard[1]* (hawkins.s.shepard@vanderbilt.edu), Jody C. May,[1] Piyoosh Babele,[1] Carl H. Johnson,[1] Brian F. Pfleger,[2] **Jamey D. Young**,[1] and John A. McLean[1]

**Institutions:** [1]Vanderbilt University, Nashville; [2]University of Wisconsin-Madison, Madison

**Project Goals: In an effort to accelerate the timeframe associated with the screening of genetic edits, an analytical workflow using desorption electrospray ionization mass spectrometry imaging was developed that allows for the rapid analysis of genetically engineered biological systems in-situ. The method has demonstrated the ability to differentiate individual phenotypes in co-cultures and can provide comparisons of the relative levels of metabolites produced.**

Developments within the field of synthetic biology have allowed for the production of genetic variants to outpace the screening capabilities of traditional analytical workflows. The acceleration of strain phenotyping is a critical need for effective determination of optimal target sequences for genetic engineering. In order to increase the throughput associated with this process, we have developed a workflow implementing desorption electrospray ionization-mass spectrometry imaging (DESI-MSI) that reduces sample preparation and allows for the rapid characterization of biosynthetic systems. DESI-MSI is unique in its ability to perform detailed spatially-resolved chemical mapping *in situ* that enables broadscale molecular phenotyping, including the analysis of labile and volatile small molecules. Through untargeted acquisitions and unsupervised segmentation, this multiplexed method is able to efficiently determine the metabolic phenotypes of bacterial colonies by analyzing their molecular profiles. The effectiveness of the workflow was demonstrated through the analysis of *Escherichia coli* strains engineered to overproduce free fatty acids (FFAs), wherein a single acquisition is able to distinguish individual phenotypes present within a co-culture and identify strains with the highest FFA production. Because the untargeted nature of the DESI-MSI approach, global metabolic information related to off-target production, product sinks, and membrane lipid compositions is also garnered, providing a more holistic screening of engineering efficacy. Further developments are currently being undertaken in order to optimize the workflow for the characterization of genetic edits associated with the production of high-value chemicals by cyanobacteria, with products such as FFAs having the potential to provide the basis of a renewable feedstock for biodiesel.

*This work was supported by Department of Energy awards DE-SC0019404 and DE-SC0022207.*

**Title:** Systems biology to enable modular metabolic engineering of fatty acid production in cyanobacteria

**Authors: Jamey Young**,[1,2,*] (j.d.young@vanderbilt.edu), Carl Johnson,[2,3] John McLean,[4] Doug Allen,[5,6] Bradley Evans,[6] Brian Pfleger[7]

**Institutions:** [1]Chemical and Biomolecular Engineering, [2]Molecular Physiology and Biophysics, [3]Biological Sciences, [4]Chemistry, Vanderbilt University, Nashville, TN, USA. [5]USDA-ARS [6]Donald Danforth Plant Science Center, St. Louis, MO, USA. [7]Chemical and Biological Engineering, University of Wisconsin, Madison, WI, USA.

**Project Goals:** This project will apply systems biology to identify metabolic control points and bottlenecks that regulate flux to lipid products in cyanobacteria. Our long-term goal is to develop strains of cyanobacteria that are optimized for producing renewable chemicals at commercially feasible rates and titers.

**Abstract Text:** Cyanobacteria are attractive hosts for biomanufacturing because of their ability to rapidly fix CO2, grow in nutrient-poor environments, and produce renewable chemicals directly from photosynthesis. Unlike triacylglycerol production in green algae, production of free fatty acids (FFAs) using genetically engineered cyanobacteria results in secretion of the product into the culture medium where it can be efficiently recovered. However, there is a major gap in our understanding of how lipid metabolism is regulated in cyanobacteria that limits our ability to rationally engineer high-titer FFA production in cyanobacterial hosts. The overall objective of this project is to use systems biology to identify metabolic control points and bottlenecks that regulate flux to FFAs in cyanobacteria. Our central hypothesis is that cyanobacterial lipid metabolism can be modularized into pathways that are "upstream" and "downstream" of the nodal metabolite acetyl-CoA, which can be separately studied and optimized to enhance overall FFA production. We will apply a suite of systems biology approaches including $^{13}C$ flux analysis, metabolomics, lipidomics, proteomics, and CRISPRi screens to rigorously define the regulation of flux within each module. As proof of principle, this modular approach will be applied to optimize cyanobacterial production of FFAs in the fast-growing, halotolerant strain *Synechococcus* sp. strain PCC 7002. The rationale for the proposed research is that a deeper understanding of how fatty acid flux is regulated upstream and downstream of acetyl-CoA will enable integrated "push-pull" metabolic engineering strategies to produce lipid products directly from photosynthetic $CO_2$ fixation in cyanobacteria. This research will directly contribute to DOE's mission by advancing toward biological production of renewable fuels that do not compete with agriculture.

**Funding Statement:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022207.

**Designing Novel Enzymes for Complete Degradation of Recalcitrant Polyamides**

Jacob Bale,[1] Amandeep Sangha[1*] (amandeep.sangha@arzeda.com), Karen Eaton [1*] (karen.eaton@arzeda.com), Chris Voigt,[2] **Alexandre Zanghellini**[1*] (alexandre.zanghellini@arzeda.com)

[1]Arzeda Corporation, Seattle, WA; [2]Massachusetts Institute of Technology, Cambridge, MA

**Project Goals:**

**The goal of this project is to design enzymes capable of complete depolymerization of nylon 6 and nylon 66 and engineer bacterial strains able to metabolize the degradation products to higher-value sustainable materials.**

As of 2015, a total of 6.3 billion tons of plastic waste had been generated globally. It is estimated only 9% of this total had been recycled, while 12% had been incinerated to recover energy values, and the remainder entered landfills or the natural environment. New technologies are needed to address this ever-growing problem. While closed-loop recycling methods offer new potential routes for dealing with plastic waste, competition with cheap, fossil fuel-derived precursors is likely to inhibit progress on this front for the foreseeable future. An alternative approach, harnessing the power of biology to not just depolymerize plastics back to their monomer precursors, but convert them into higher-value products offers stronger economic incentives and in turn would be expected to drive more rapid and widespread adoption. Toward that end, our work focuses on combining cutting-edge computational protein design and synthetic biology to address the challenge of complete biodegradation and upcycling of the recalcitrant polymers nylon 6 and nylon 66. Although natural enzymes have been shown to be able to degrade amorphous portions of polyamides such as nylon 6 and nylon 66, complete enzymatic degradation has not been demonstrated. We hypothesize this to be due in large part to a lack of natural enzymes able to efficiently catalyze degradation of the crystalline portion of the polymer. We are computationally designing enzymes to alleviate this limitation by introducing and optimizing polyamide hydrolysis activity in scaffolds with open active sites. In conjunction, we are screening and engineering bacterial strains able to metabolize nylon 6 and nylon 66 degradation byproducts directly into central metabolism. Once achieved, such platform strains can be used to produce a wide variety of fermentation products from central metabolites. Integration of our designer nylon 6 and nylon 66 depolymerizing enzymes into these engineered hosts will provide a novel, elegant, and cost-effective consolidated fermentation process for nylon upcycling to higher-value sustainable materials.

**References**

1. Geyer, R., Jambeck, J. R. & Law, K. L. Production, use and fate of all plastics ever made. *Sci. Adv.* 3, e1700782 (2017).
2. Kakudo, S., Negoro, S., Urabe, I. & Okada, H. Nylon oligomer degradation gene, nylC, on plasmid pOAD2 from a Flavobacterium strain encodes endo-type 6-aminohexanoate oligomer hydrolase: purification and characterization of the nylC gene product. *Appl. Environ. Microbiol.* **59**, 3978-3980 (1993).

3. Kinoshita, S. *et al.* 6-Amniohexanoic acid cyclic dimer hydrolase. A new cyclic amide hydrolase produced by Acromobacter guttatus KI72. *Eur. J. Biochem.* **80**, 489-495 (1977).
4. Kato, K. *et al.* Amino acid alterations essential for increasing the catalytic activity of the nylon-oligomer-degradation enzyme of Flavobacterium sp. *Eur. J. Biochem.* **200**, 165-169 (1991).
5. Yasuhira, K. *et al.* 6-Aminohexanoate oligomer hydrolases from the alkalophilic bacteria Agromyces sp. Starin KY5R and Kocuria sp. Strain KY2. *Appl. Environ. Microbiol.* **73**, 7099-7102 (2007).

**Funding Statement:** This work is supported by the award number DE-SC0022077 by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program.

# Experimental Approaches to Understand Rhizosphere Processes for Improved Bioenergy Crop Productivity

Yuntao Hu[1*](yuntaohu@lbl.gov), Vlastimil Novak[1], Yi Zhai[1], John P. Vogel[1,2,3], Trent R. Northen[1,2], **Karsten Zengler**[4]
[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Joint Genome Institute, Berkeley, CA; [3]University of California, Berkeley, CA; [3]University of California, San Diego, CA

**Project Goals: This project couples novel lab and field studies to develop the first predictive model of grass-microbiomes based on new mechanistic insights into dynamic plant-microbe interactions in the grasses *Sorghum bicolor* and *Brachypodium distachyon* that improve plant nitrogen use efficiency (NUE). The results will be used to predict plant mutants and microbial amendments which improve low-input biomass production for validation in lab and field studies. To achieve this goal, we will determine the mechanistic basis of dynamic exudate exchange in the grass rhizosphere, with a specific focus on the identification of plant transporters and proteins that regulate root exudate composition and how specific exudates select for beneficial microbes that increase plant biomass and NUE. Further, we will develop a predictive plant-microbe model for advancing sustainable bioenergy crops and will predictively shift plant-microbe interactions to enhance plant biomass production and N acquisition of N from varied forms.**

Up to 25% of all photosynthetically fixed carbon is released into the rhizosphere through root exudates. Exudates, consisting of >100 distinct metabolites, provide a critical source of growth substrates for the root microbiota, consisting of diverse microorganisms and fauna actively recruited by plants from bulk soil. Yet, we lack the foundational understanding of plant exudate biology including if and how exudation is controlled by active transporters and regulatory mechanisms. Unraveling the nature of plant and microbial drivers behind these dynamic interactions is crucial for understanding the dynamics of N cycling and uptake in the rhizosphere, and will guide the development of predictive models to inform the design of sustainable bioenergy systems.

Plants use multiple forms of N from soil including nitrate ($NO_3^-$), ammonium ($NH_4^+$), and various forms of organic N. To investigate how inorganic N affects plant root exudation, we grew *B. distachyon* in sterile fabricated ecosystem devices (EcoFABs) supplied with $NO_3^-$, $NH_4^+$, or N-depleted media and then characterized the resulting plant phenotypes, Additionally, we applied LC-MS/MS-based metabolomics to analyze root exudates. Our untargeted metabolomics results demonstrate strong modulation of root exudation by N supply. N source significantly changed root and shoot biomass and N uptake kinetics. To examine the circadian rhythms of root exudation, we quantified the diurnal root exudation patterns of *B. distachyon* cultivated in EcoFABs supplied with inorganic N. This analysis revealed more sugars and amino acids, such as hexose, isoleucine, and glutamic acids, are exudated during the daytime. Finally, to elucidate how plant transporters regulate the root exudation, we analyzed metabolites secreted by *B. distachyon* ABC and N transporter mutants grown in EcoFABs. Together these studies are providing new foundational insights into the regulation of plant-microbe interaction by evaluating the dynamics and transporter of plant exudation.

*This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021234.*

# Metabolic modeling of a rhizosphere microbial community

Manish Kumar[1*] (m3kumar@health.ucsd.edu), Maxwell Neal[1], Deepan Thiruppathy[1], Gabriela Canto Encalada[1], Anurag Passi[1], Rodrigo Santibañez[1], William Kang[1], Diego Tec-Campos[1], Oriane Moyne[1], Cristal Zuniga[1], **Karsten Zengler**[1,2,3]

[1]Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0760, USA
[2]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA
[3]Center for Microbiome Innovation, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0436, USA

## Project Goals:

The goal of this project is to integrate novel systems biology-based tools and genome-scale metabolic modeling with laboratory ecosystems (EcoFABs), plant genetics, and integrated field trials to gain insights into the dynamic plant-microbe interactions governing nitrogen exchange to improve plant productivity in marginal soils. The vast complexity of plant-microbe interactions and their dynamic nature renders elucidation of the underlying mechanisms of these interactions. Thus, we are developing a predictive metabolic model of plant root and rhizosphere community to guide experimental designs and generate hypotheses that are then tested in laboratory and field studies.

## Abstract:

Genome-scale metabolic models for microorganisms derived from the rhizosphere can be employed as a framework to unravel microbe-microbe and host-microbe interactions. Here, we present manually-curated genome-scale metabolic models for 17 bacteria of the rhizosphere. These bacteria were isolated from switchgrass rhizospheres, and represent dominant members commonly found in the rhizosphere of grasses belonging to the genera *Arthrobacter, Bacillus, Bosea, Bradyrhizobium, Brevibacillus, Burkholderia, Chitinophaga, Lysobacter, Methylobacterium, Mucilaginibacter, Mycobacterium, Niastella, Paenibacillus, Rhizobium, Rhodococcus, Sphingomonas,* and *Variovorax.* Genome-scale models were generated using standard pipelines developed by our lab. Collectively, the models incorporate 3,877 reactions (including exchange reactions that represent uptake and secretion of metabolites) and 2,663 metabolites. Moreover, the individual metabolic models contain 790 to 1,788 genes, covering 15% to 30% of all functionally annotated genes in these microbial genomes. All models were manually curated using information based on literature and public databases, such as KEGG, modelSEED, MetaCyc, UniProt, BRENDA, IntEnz, TCDB, and TransportDB. The curation resulted in the removal of 271 genes and 21 reactions from the models on average, while 131 new genes and 3 new reactions were introduced into the models during curation. We simulated these models on 95 different carbon sources and compared predicted phenotypes with experimental measurements. This practice helped to validate the prediction capabilities of these models and provided information on pathways absent in the models. Subsequent gap-filling added 288 new reactions to the models. We are currently validating growth predictions under hundreds of different conditions experimentally. Collectively, the study will lay the foundation to

advance our understanding of dynamic plant-microbe interactions in the rhizosphere and how these interactions contribute to nitrogen use efficiency.

**Funding:** This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under Award DESC0021234.

# Reconstruction of a genome-scale community metabolic model of a microbial co-culture to enable next generation biochemical production

Gustavo Lastiri-Pancardo[1*] (glastiripancardo@health.ucsd.edu), Deepan Thiruppathy[1], Neal N. Hengge[4], Sean Woodworth[5], Violeta Sánchez i Nogué[5], Jeffrey G. Linger[5], Michael T. Guarnieri[4], **Karsten Zengler**[1,2,3]

[1]Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0760, USA

[2]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0412, USA

[3]Center for Microbiome Innovation, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, 92093-0436, USA

[4]Biosciences Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO, 80401, USA

[5]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO, 80401, USA

## Project Goals:

The goal of this project is to integrate innovative systems biology tools and genome-scale metabolic modeling to acquire a deeper understanding of the mutualistic interactions in a clostridial co-culture for the effective conversion of lignocellulosic biomass to butyric acid. Moreover, we aim to design a pipeline that facilitates the creation of engineering strategies to increase the complexity, stability and productivity of designer microbial communities for the production of next-generation biofuels and bioproducts from diverse biomass feedstocks.

## Abstract:

Multi-species microbial consortia are widespread in nature. The majority of these communities are very diverse and highly dynamic. These complex communities depend on intricate interactions among their members, thereby accomplishing metabolic functions that are unfeasible for individual members. Understanding the relationship inside a community and being able to predict perturbation outcomes is critical for the design of robust and productive consortia for biochemical

production. This project studies a lignocellulosic biomass-degrading and butyrate-producing co-culture of *Clostridium thermocellum* and *Clostridium thermobutyricum*, to unravel inter-microbial interactions with the goal to design more effective consortia for production of bioproducts from biomass.

To untangle metabolic traits and interactions of the clostridial co-culture, we first constructed and manually curated two individual genome-scale metabolic models (GEMs) for these strict anaerobic organisms, based on protein homology to other model microbial species and annotated data from public databases and literature; covering around 27% of all genes found in their annotated genome sequences. The two GEMs were then integrated into a compartmentalized community metabolic (CM) model. The compartmentalized model iGL1101, accounts for 1,888 reactions, 1,636 metabolites and 1,101 genes. We will experimentally validate and refine the CM model though an iterative procedure that involves integration of multi-omics data and substrate utilization information obtained from Phenotype MicroArrays (Biolog) plate experiments. Model simulations will subsequently identify the biosynthetic requirements and trade-offs for butyrate production. The CM model offers a computational framework to contextualize butyrate production from lignocellulosic biomass and to determine metabolic bottlenecks of this co-culture. Identification of these bottlenecks will serve as a foundation for targeted strain engineering and the design of consortia of increased complexity for the effective conversion of biomass into bioproducts.

**Funding:** This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under Award DE-SC0022137.

**Title: Improved Biofuel Production through Discovery and Engineering of Terpene Metabolism in Switchgrass.**

**Authors:** Kira Tiedge[1,2]*(kjtiedge@ucdavis.edu) and **Philipp Zerbe[1]**

**Institutions:** [1]Department of Plant Biology, University of California, Davis, USA; [2]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, Netherlands

**Website URL:**

**Project Goals:** Of the myriad specialized metabolites that plants form to adapt to environmental challenges, terpenoids form the largest group. In many major crops, unique terpenoid blends serve as key stress defenses that directly impact plant fitness and yield. In addition, select terpenes are used for biofuel manufacture. Thus, engineering of terpenoid metabolism can provide a versatile resource for advancing biofuel feedstock production but requires a system-wide knowledge of the diverse biosynthetic machinery and defensive potential of often species-specific terpenoid blends. This project merges genome-wide enzyme discovery with comparative -omics and protein structural studies to define the biosynthesis and stress-defensive functions of switchgrass (Panicum virgatum) terpenoid metabolism. These insights would be combined with the development of genome editing tools to design plants with desirable terpene blends for improved biofuel production on marginal lands.

**Abstract Text:** Diterpenoids constitute a diverse class of metabolites with critical functions in plant development, defense, and ecological adaptation. Major monocot crops, such as maize (*Zea mays*) and rice (*Oryza sativa*), deploy diverse blends of specialized diterpenoids as core components of biotic and abiotic stress resilience. This project reports the genome-wide discovery and functional characterization of the stress-related diterpenoid-metabolic network in the bioenergy crop switchgrass (*Panicum virgatum*). Mining of the allotetraploid switchgrass genome identified expansive diterpene synthase (diTPS) and cytochrome P450 monooxygenase (P450) enzyme families critical for the chemical diversity of bioactive diterpenoids. Tissue-specific transcriptome and metabolite analyses of drought-resistant (Alamo) and drought-susceptible (Cave-in-Rock) genotypes showed an earlier onset of transcriptomic changes and significantly more differentially expressed genes in response to drought in Cave-in-Rock. Diterpenoid-biosynthetic genes showed drought-inducible expression in Alamo roots, contrasting largely unaltered triterpenoid and phenylpropanoid pathways. In addition, metabolomic analyses identified common and genotype-specific terpenoids. Consistent with transcriptomic alterations, several root diterpenoids showed significant drought-induced accumulation. Structural analysis of drought-responsive root diterpenoids verified these metabolites as oxygenated furanoditerpenoids

that are perhaps unique to switchgrass. Together, these findings support a role of diterpenoids in switchgrass drought stress tolerance and provide resources for understanding the molecular mechanisms underlying switchgrass environmental resilience.

**References/Publications**


1. Muchlinski A, Jia M, Tiedge K, Fell JS, Pelot KA, Chew L, Davisson D, Chen Y, Siegel J, Lovell JT, Zerbe P (2021) Cytochrome P450-catalyzed biosynthesis of furanoditerpenoids in the bioenergy crop switchgrass (*Panicum virgatum* L.). Plant J. 108:1053-1068.
2. Tiedge K, Muchlinski A, Zerbe P (2020) Genomics-enabled analysis of specialized metabolism in bioenergy crops: current progress and challenges. Synth Biol (Oxf). 5:ysaa005.
3. Pelot KA, Chen R, Hagelthorn DM, Young CA, Addison JB, Muchlinski A, Tholl D, Zerbe P (2018) Functional Diversity of Diterpene Synthases in the Biofuel Crop Switchgrass. Plant Physiol. 178:54-71.



**Funding Statement:** Financial support for this work was provided by the U.S. Department of Energy (DOE) Early Career Research Program (DE-SC0019178, to PZ), the German Research Foundation (DFG) Research Fellowship (TI 1075/1-1, to KT), and the DOE Joint Genome Institute (JGI) DNA Synthesis Science Program (grant #2568, to PZ). The gene synthesis work conducted by the U.S. Department of Energy Joint Genome Institute (JGI), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**Title:** A gene-editing system for large-scale fungal phenotyping in a model wood decomposer

**Jiwei Zhang**[1, *] (zhan3437@umn.edu), Weiran Li[1], Jonathan S Schilling[1], Igor Grigoriev[2], Daniel Cullen[3], and Hugh D Mitchell[4]

[1]University of Minnesota, Saint Paul, MN; [2]DOE Joint Genome Institute - Lawrence Berkeley National Laboratory, Walnut Creek, CA; [3]USDA, FPL - University of Wisconsin, Madison, WI; and [4]PNNL - Environmental Molecular Sciences Laboratory, Richland, WA

**Website URL:** http://url.SpecificToProject.gov (optional)

**Project Goals:** Our goal is to combine systems biology approaches and gene-editing to develop a high-throughput genetic platform for large-scale phenotyping in a model wood decomposer fungus relevant to the DOE mission area. We will develop an efficient, CRISPR/Cas9-based gene-editing platform with model brown rot fungi, discovering and validating at a large scale the genetic functions of key wood-decaying gene pathways identified by network analysis. By this project, we expect to provide stand-alone tools and resources for discovering novel fungal genetic features that can also be used in combination to advance relevant plant biomass conversion research in the post-genomic era.

**Abstract:**

Wood decay fungi offer industrially-relevant pathways to extract carbohydrates from lignocellulose, and their mechanisms have broad relevance to global carbon cycling. Among these organisms, white rot-type fungi use ligninolytic enzymes (e.g., lignin peroxidases) to break down the lignin barrier, while brown rot fungi use non-enzymatic reactive oxygen species (ROS) mechanisms to modify lignin and selectively extract sugars.[1, 2] Brown rot fungi evolved from white rot ancestral lines multiple times and, from a process efficiency standpoint, they represent a pathway 'upgrade' to approaches that mimic white rot. Brown rot is generally faster than white rot, the mechanism selectively releases soluble sugars by leaving lignin relatively intact as a by-product (value addition).[1, 2] These fungi achieved this, evolutionarily, by contracting rather than enhancing their repertoire of carbohydrate-active enzymes (CAZys), shedding ~65% of their ancestral CAZy genes, on average.[3, 4, 5] Although DOE mission-relevance is clear, and we have made major genomically-informed advances, progress is limited by an inability to manipulate genes in any brown rot fungal strain.

Using multi-omics tools, our team has been leading the research of dissecting this brown rot mechanism 'upgrade'. 1) We reported the first brown rot genome in *Postia placenta* (now *Rhodonia placenta*) in PNAS in 2009.[3] 2) We then used phylogenomics to discover the genetic inventories unique to brown rot fungi in Science in 2012[4] and in PNAS in 2014[5]. 3) More recently, we used functional genomics to further elucidate a staggered 'two-step' (i.e., oxidation-then-hydrolysis) gene regulation model for brown rot in PNAS in 2016 and in mBio in 2019.[6, 7] Although these genomic studies have greatly advanced our understanding of brown rot, its

genetic basis remains uncharacterized and unharnessed. This is, to a large extent, due to the lack of a robust genome editing tool – these fungi have one of the most mission-relevant mechanisms for deconstruction, and we cannot currently engineer their genomes to learn *how* they work.

Targeted gaps are remaining for understanding the brown rot genetic mechanism. *First*, multi-omics approaches have advanced our knowledge of the 'two-step' mechanism,[6, 7] but the functions of genes involved remain unverified and ambiguous. *Second*, brown rot fungi have adapted distinct gene regulatory mechanisms to precisely control and consolidate the two steps during wood decay,[8, 9] but little is known about this process. This limits our ability to refer to brown rot gene functions from other fungi with known decay pathways. *Third*, the majority of genes identified by multi-omics are of hypothetical/unknown function, [5, 6, 7] leaving major gaps for discovering novel functional genes.

This project will integrate systems biology, genome-editing, and network modeling to address these key gaps. Our **objective 1** is to optimize a CRISPR/Cas9-mediated gene-editing system based on the established genetic manipulation platform in brown rot fungi, and use it for targeting genes that have been isolated by multi-omics studies. This will generate a single-gene mutant library for phenotypic studies. Our **objective 2** is to build an extensive carbon utilization network using transcriptomic analyses and network modelling, discovering the distinctive gene regulation features adapted by brown rot. Leveraging this objective, we will also complement the gene pool for large-scale phenotypic screening. Finally, the **objective 3** aims to develop a pipeline to use the multiplexing sgRNA library for genome-editing, building the mutant library for large-scale phenotypic screens. This will allow us to rapidly link genotype to the phenotypes that enable brown rot efficacy.

With this project, we anticipate to provide stand-alone tools and resources to elucidate fundamental microbial processes relevant to DOE mission area, advancing new engineering designs for lignocellulose bioconversion.

**References:** [1]Hibbett & Donoghue 2001 Syst Biol 50:215-; [2]Eastwood et al. 2011 Science 333:762-; [3]Martinez et al. 2009 PNAS 106:1954-; [4]Floudas et al. 2012 Science 336:1715-; [5]Riley et al. 2014 PNAS 111:9923-; [6]Zhang et al. 2016 PNAS 113:109682-; [7]Zhang et al. 2019 mBio 10:e02176-19; [8]Zhang et al. 2017 Fungal Genet Biol 106:1-; [9]Kubicek et al. 2014 Annu Rev Phytopathol 52:427-.

**Funding Statement:** This research is supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022151.

**Multi-omic investigation of lipid accumulation mechanisms in *R. toruloides***

Anshu Deewan[1], Shekhar Mishra[1], Christopher Rao[1], Huimin Zhao[1]
[1]University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois
https://www.igb.illinois.edu/research-areas/gsecenter

**Project Goals: Our project aims to develop new metabolic engineering, omics analysis, and computational modeling tools at the genome scale for strain development, which may be implemented in an automated manner at the Illinois Biological Foundry for Advanced Biomanufacturing. Two non-model yeasts, *Rhodosporidium toruloides* for the production of oleaginous compounds and *Issatchenkia orientalis* for the production of organic acids, are selected as the platform organisms. Milestones achieved so far include reconstruction of comprehensive genome-scale metabolic models, development of carbon mapping models, and $^{13}$C-metabolic flux analysis at the genome-scale. The final goal is to develop kinetic models to guide metabolic engineering accounting for reaction kinetics and allosteric regulations.**

Oleaginous yeasts are, in recent years, being studied in greater detail for their higher lipid accumulating capacity. In the context of metabolic engineering, they offer the promise of platform hosts that can be used to produce lipid-related biochemicals. Their higher accumulation potential over model yeasts suggests the possibility of achieving production titers that can compete with chemical synthesis routes, thus ensuring economic viability. One such microorganism is the non-model yeast *Rhodosporidium toruloides*, which has been documented to accumulate high titers of intracellular lipids. However, barriers such as lack of precision genetic editing or an incomplete understanding of its metabolic and regulatory topology can hinder the creation of successful overproduction phenotypes. Here we adopt a multi-omic approach to investigate the molecular mechanisms governing lipid accumulation. The *R. toruloides* strain IFO0880 was cultivated in nitrogen-rich and nitrogen-deficient conditions and analyzed via RNA-Seq and lipidomic analysis experiments.

The results from transcriptomic analysis indicate global metabolic shifts were governed by nitrogen limitation. From hierarchical clustering of gene expression data, a pattern emerged where samples grown in high carbon-to-nitrogen ratio clustered together despite different values of the ratio. This trend was observed in both fatty acid biosynthesis and fatty acid utilization pathways implying that culture conditions exerted a greater amount of control on gene expression profiles than the exact function of different pathways. Clustering analysis across different timepoints showed that in the logarithmic stage of growth, all cultures clustered together independent of the carbon-to-nitrogen ratio, while the distinct clusters were only seen in the stationary phase of culture, indicating the role that logarithmic and stationary phase play in switching to a lipid accumulation mode.

In addition, we identified the genes and pathways involved in transcriptional reprogramming during lipid production induced by nitrogen limitation and mapped them to metabolic pathways of *R. toruloides*. We further used bioinformatics analysis to identify networks of co-regulated gene clusters. The multi-omics analysis of regulatory mechanisms of lipid accumulation in such oleaginous yeast thus paves a useful path towards identifying control levers of its metabolism.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018260.

**Title:** Population phenomics of *Issathenkia orientalis* for creation of a safer and more robust chassis for next-generation industrial biotechnology

**Authors:** Yusuke Sasaki[1](yusukesasaki@lbl.gov), Ping-Hung Hsieh[1], Zia Fatma[2], Huimin Zhao[2], and Yasuo Yoshikuni[1]

**Institutions:** [1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of Illinois at Urbana-Champaign, Urbana, IL

**Project Goals:** To alleviate reliance on chemical synthesis and shift into the bioeconomy regime, it is imperative to develop a biological chassis for producing organic acids from renewable biological resources. *Issatchenkia orientalis*, a nonconventional acid-tolerant yeast, has been characterized, engineered, and employed as a production chassis for several organic acids, including succinate acid[1] and D-lactic acid[2]. In addition to tolerating acid, *I. orientalis* has extraordinary tolerance to other stressors, including heat, lignocellulosic inhibitors, and antifungal drugs, although the biological mechanisms for these tolerances are still unknown. These innate tolerances and the limited information on their underlying mechanisms arouse safety concerns, making use of *I. orientalis* as a production chassis a double-edged sword. To help alleviate these concerns, we elucidated genotype-phenotype relationships by analyzing populations obtained by collecting *I. orientalis* strains from around the globe. This effort will provide a knowledge database to help design and build a safer and more robust *I. orientalis* chassis for industrial biotechnology.

**Abstract Text:** The bioeconomy revolves around efforts to use bio-based resources and products to help tackle grand societal challenges such as climate change, nature conservation, and food security. Microbial chassis mediate the bioeconomy scheme, serving as molecular factories that convert various biological resources into a broad spectrum of chemicals, fuels, and materials. Among these products, organic acids are value-added platform molecules that have gained attention in the food, pharmaceutical, and chemical industries. Producing them microbially using bioenergy crops such as miscanthus and switchgrass as feedstocks is a viable way to reduce reliance on chemical synthesis and avoid its environmental burdens. Tolerance of final organic acids, low pH, and inhibitory compounds present in lignocellulosic hydrolysates are prerequisites in order for any organic acid production chassis to be cost-effective and environmentally friendly. A nonconventional yeast, *Issatchenkia orientalis*, is well aligned with the requirements because of its renowned tolerance to multiple stresses, including various organic acids[1], low pH, phenolic compounds, furan aldehydes, and salts. However, it also resists antifungal drugs, potentially raising concerns (most pathogenic fungi are opportunistic, even *Saccharomyces cerevisiae*, or baker's yeast)[3]. These potential concerns can be alleviated by following laboratory practices appropriate for *I. orientalis*'s assessment at biosafety level 1, the lowest security level. Nevertheless, considering that *I. orientalis* would typically be genome engineered to optimize it as a production chassis, it is important to identify genotype-phenotype relationships that might bring to light opportunities to engineer robust but safer properties. In this study, we focused on

expanding knowledge of *I. orientalis* through population phenomics with associated statistical analyses. We developed a high-throughput phenotyping assay method that allowed us to measure growth profiles of respective strains on an agar plate in various chemical compounds and under various environmental conditions. We used the assay to analyze 160 *I. orientalis* strains across 58 representative conditions. Analysis of the correlation between conditions indicated a trend in population-level behavior, suggesting that significantly (anti-)correlated cellular responses likely occur through a mechanism that affects the phenotypes and no significantly correlated pairs may not arise from a common mechanism. Therefore, we highlighted specific paired conditions in the context of generating a safer and more robust chassis. We also compared the geographical origins of each isolate with the phenotypic behaviors to investigate potential evolutionary effects of surrounding environments. In addition, a genome-wide association study (GWAS) identified potential genetic loci, and deletion of these loci allowed us to select for a genetic locus that affects phenotypic changes. Our study showcases the utility of this population growth assessment strategy for evaluating phenotypic changes for further statistical analyses. The knowledge database we developed will offer potential opportunities to engineer safer and more robust chassis for biotechnological applications.

**References**


1. Xiao, H., Shao, Z., Jiang, Y., Dole, S. & Zhao, H. Exploiting *Issatchenkia orientalis* SD108 for succinic acid production. *Microbial Cell Factories* **13**, 1–11 (2014).

2. Park, H. J. *et al.* Low-pH production of d-lactic acid using newly isolated acid tolerant yeast *Pichia kudriavzevii* NG7. *Biotechnology and Bioengineering* **115**, 2232–2242 (2018).

3. Pérez-Torrado, R. & Querol, A. Opportunistic strains of *Saccharomyces cerevisiae*: A potential risk sold in food products. *Frontiers in Microbiology* **6**, 1522 (2016).



**Funding Statement:** This work was supported by the Biosystems Design program (DOE, Office of Science contract DE-SC0018260), US Department of Energy (DOE) Joint Genome Institute (DOE Office of Science User Facility contract DE-AC02-05CH1123), and the Genomic Science Program (DOE, Office of Science contract DE-SC0020390).

**Examining organic acid production and model-driven strategies in *Issatchenkia orientalis***

Patrick F. Suthers[1*] (pfs13@psu.edu), Teresa A. Martin[2], Zia Fatma[2], Vinh G. Tran[2], Hoang V. Dinh[1], Costas D. Maranas[1], and **Huimin Zhao**[2]
[1]The Pennsylvania State University, University Park, Pennsylvania; [2]University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois

https://www.igb.illinois.edu/research-areas/gsecenter

**Project Goals: Our project aims to develop new metabolic engineering, omics analysis, and computational modeling tools at the genome scale for strain development, which may be implemented in an automated manner at the Illinois Biological Foundry for Advanced Biomanufacturing. Two non-model yeasts, *Rhodosporidium toruloides* for the production of oleaginous compounds and *Issatchenkia orientalis* for the production of organic acids, are selected as the platform organisms. Milestones achieved so far include reconstruction of comprehensive genome-scale metabolic models, development of carbon mapping models, and $^{13}$C-metabolic flux analysis at the genome-scale. The final goal is to develop kinetic models to guide metabolic engineering accounting for reaction kinetics and allosteric regulations.**

Many platform chemicals can be produced from renewable biomass by microorganisms, with organic acids making up a large number of these chemicals. Intolerance to the resulting low pH growth conditions and inhibitors in biomass hydrolysates, however, remains a challenge for the industrial production of organic acids by microorganisms. The unique metabolic capabilities and resilience to inhibitory stressors enable some non-model yeasts to be attractive microbial cell factories. The non-model yeast *Issatchenkia orientalis* is a promising host for industrial production because it is tolerant of low pH conditions.

Here, we explore engineering synthetic pathways in *I. orientalis* to produce 21 different organic acids from glucose under aerobic or microaerobic conditions. These organic acids include seven of the eight organic acids identified by the US Department of Energy as part of the top twelve bio-based building-blocks. We use a genome-scale metabolic (GSM) model for *I. orientalis* SD108 to blueprint pathways and to examine yields and dependences of product formation on oxygen uptake levels. Based on the model, we determined that the production of 3-hydroxypropionic acid in *I. orientalis* can be enhanced by microaerobic fermentation conditions. Our experimental results have recently supported this finding. We also use constraint-based methods to assess the potential of computationally designing growth-coupled strains and to propose genetic modifications that bolster product formation. We identified growth-coupled strategies for 15 of the substrate-product pairs. We detail our work to introduce pathways *in vivo* and the use of a recently developed CRISPR/cas system for *I. orientalis*. We also highlight recent updates to the GSM model.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018260.