U.S. DEPARTMENT OF
ENERGY

2021 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 22, 2021

Dear Colleague:

Welcome to the 2021 Genomic Science and Bioimaging Program Annual PI Meeting. What an extraordinary year it has been since we last met! It was almost exactly a year ago that we were all convening in Washington DC for what was to be the last big face-to-face meeting within BER. Two weeks later we were all on lockdown. BSSD staff have remained on telework status ever since. Despite the challenges that remote work has presented, we have tried to maintain our programmatic activities as much as possible and our portfolio continues to thrive due to your fervent efforts. From all of us in BSSD, we thank you for your endeavors to maintain a sense of normalcy in abnormal times and continuing your best efforts to advance your research as the situation allows. The ongoing COVID-19 pandemic has impacted us all in various ways, and we hope, above all, that you and your families have remained healthy and safe.

From BSSD's perspective there have been some remarkable achievements to note this past year. The 2020 Nobel prize in chemistry was awarded to Dr. Jennifer Doudna (UC-Berkeley and LBNL) for her efforts in developing CRISPR-based gene editing, a discovery whose origins trace back to early DOE support and which now is a standard technique used in much of the research presented at this meeting. Also, Dr. Susannah Tringe (LBNL and JGI) was awarded one of DOE's highest honors as a recipient of the 2020 Ernest Orlando Lawrence Award for her work in environmental metagenomics and fostering entire communities of researchers in this area through her position at JGI. Congratulations to both!

It has also been a year of extraordinary scientific publications in our core topics of bioenergy and bioproduct research from the BRCs, biosystems design, plant genomics, microbial and microbiome science, structural biology, and bioimaging research. Additionally, there have been exciting developments within JGI, KBase, NMDC, at the new cryo-EM centers, and at our resources at the DOE Synchrotron Light and Neutron Sources. As a result, the portfolio continues to be rich with amazing ideas and full of promise. As we transition to a new Administration, we are confident that the Division will have a major role in spurring innovations for renewable energy, biotechnology, environmental research, and enabling an expanding bioeconomy. We therefore remain very optimistic on the portfolio's future and wish to thank you again for making the Genomic Science and Bioimaging programs so successful. It is our pleasure to work with you in this exciting and vibrant program.

This year, we have combined the Genomic Science meeting with our Bioimaging PI meeting. Our Bioimaging program seeks to develop new multimodal methods to image and analyze biological processes of relevance to the Genomic Science program. We encourage cross-connections between our Genomic Science and Bioimaging portfolios, and in this spirit, we are privileged to have Dr. Eva Nogales of UC-Berkeley as our keynote speaker to highlight some of the extraordinary developments in imaging science that will hopefully be of interest to everyone. We also have an excellent lineup of plenary speakers, breakout sessions and presentations from team representing key enabling capabilities being developed within the portfolio.

Finally, I would like to extend a warm welcome to our new Early Career awardees, new Sustainability PIs, new Secure Biosystems Design SFA teams, new Computational Biology projects and new Quantum Science-enabled Bioimaging projects. These last three elements are new efforts in the portfolio for this year. As always, we hope this meeting provides context for how your project fits within the larger portfolio and sparks collaborations and/or contacts to enrich your science.

Sincerely,

Todd Anderson, Ph.D.
Director, Biological Systems Science Division, SC-23.2
Office of Biological and Environmental Research
Office of Science

Table of Contents

(Click on Title to advance to the abstract)

(Click on letter to advance to the corresponding section of the Table of Contents)

A

Paul Adams: High Throughput Isolation and Microbial Community Enrichment Platforms
PRESENTER: Jennifer Kuehl

Paul Adams: Using high-throughput technologies to understand mechanisms of predation
PRESENTER: Vivek Mutalik

Paul Adams: Physical size matters in groundwater bacterial community assembly
PRESENTER: Daliang Ning

Paul Adams: Mechanism across scales: integrating laboratory and field studies for microbial ecology as illustrated by the ENIGMA SFA
PRESENTER: Heidi Smith

Paul Adams: Technologies for High-Throughput characterization of environmental isolates
PRESENTER: Valentine Trotter

Paul Adams: The Development, Progress, and Cross-Campaign Investigation of the Abiotic Influences on Denitrification Processes Partitioned Between Synthetic Communities
PRESENTER: Jacob Valenzuela

Paul Adams: Strain dynamics and functional diversity of 22 high-quality single cell genomes from ENIGMA ground water
PRESENTER: Anni Zhang

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

Andrew Allen: Developing episome-based gene expression platforms in the model diatom *Phaeodactylum tricornutum*
PRESENTER: Tessema Kassaw

Andrew Allen: Cas9-mediated mutagenesis of GS-GOGAT genes in the diatom Phaeodactylum tricornutum
PRESENTER: Mark Moosburner

Andrew Allen: Mapping transcription factor-mediated remodeling of diatom metabolism in response to shifting environmental conditions
PRESENTER: Sarah Smith

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: Diel Activity of Microbial Communities in Surface Soil Litter
PRESENTER: Eric Morrison

Steven Allison: Differential Response of Microdiversity to Simulated Global Change Within a Bacterial Genus
PRESENTER: Nick Scales

Ana Alonso: Unravelling the Role of Pennycress (*Thlaspi arvense* L.) Proteins in the Modulation of Neutral Lipid Droplet Abundance
PRESENTER: Julius Ver Sagun

Ana Alonso: Elucidating the Temporal and Spatial Organization of Storage Lipids using 13C-labeling in Developing Embryos of pennycress, a Promising Source for Aviation Fuel
PRESENTER: Umesh Prasad Yadav

Daniel Amador-Noguez: *In vivo* thermodynamic analysis of glycolysis in *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum* using ^{13}C and ^{2}H tracers
PRESENTER: Tyler Jacobson

James Anderson: Genetic improvement of seed yield and oil content in field pennycress
PRESENTER: Zenith Tandukar

Jean-Michel Ané: Creating representative microbial communities to promote biological nitrogen fixation on sweet sorghum
PRESENTER: April MacIntyre

Jean-Michel Ané: A Systems Understanding of Nitrogen-Fixation on the Aerial Roots of Sorghum
PRESENTER: Wilfred Vermerris

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

Dionysios Antonopoulos: Grammar and Language of CRISPR/Cas-Targetable Sites in *Escherichia coli*, *Shigella*, *Pseudomonas*, and *Salmonella*: A Comprehensive Survey
PRESENTER: Carla Mann

Dionysios Antonopoulos: CRISPR-Act: AI-guided Prediction of a CRISPR Kill Switch Under Diverse Physiological Conditions
PRESENTER: Rebecca Weinberg

Adam Arkin: ModelSEED release 2: High Throughput Genome-Scale Metabolic Model Reconstruction and Analysis in KBase
PRESENTER: José P. Faria

Adam Arkin: KBase Partners with SFA Developers to Increase Functionality of Tools and Analysis
PRESENTER: Elisha Wood-Charlson

Jose Avalos: Systems Biology of Isobutanol Production in Saccharomyces Cerevisiae
PRESENTER: Jose Montano Lopez

B

Ivan Baxter: Phenotyping for water use efficiency and related traits in C_4 grasses *Setaria* and *Sorghum*
PRESENTER: Ivan Baxter

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

Ivan Baxter: The Development of Transgenic Lines and Improved Technologies for the Analysis of Photosynthetic and Water Use Efficiencies in Sorghum
PRESENTER: Albert Kausch

Ivan Baxter: Targeted Mutagenesis and Programmed Transcriptional Regulation in Setaria and Sorghum
PRESENTER: Dan Voytas

Ivan Baxter: Mathematical modeling strategies to study the impact of drought on plant growth and metabolism in Setaria and sorghum.
PRESENTER: Cheng Zhao

Jeff Bennetzen: Systems analysis of the beneficial associations of sorghum with arbuscular mycrorrhizal fungi studied with genetics, genomics, imaging and microbiomics
PRESENTER: Jeff Bennetzen

Michael Betenbaugh: Revealing Metabolic Exchange and Optimizing Carbon Transformation in Co-Culture for Applications to Sustainable Biosynthesis
PRESENTER: Jackson Jenkins

Michael Betenbaugh: Sustainable production of biofuels by consortia of *Synechococcus elongatus* and *Aspergillus* species.
PRESENTER: Cristal Zuniga

Jennifer Bhatnagar: Plant- Mycorrhizal - Decomposer interactions and their impacts on terrestrial biogeochemistry
PRESENTER: Nahuel Policelli

Crysten Blaby: Delivering the goods: leveraging functional genomics to understand cofactor trafficking
PRESENTER: Crysten Blaby

Crysten Blaby: A New Structural Paradigm In Heme Binding – A Novel Family Of Plant Heme Oxidases
PRESENTER: Nicolas Grosjean

Crysten Blaby: Optimizing hydroponic growth system for metal stress studies of bioenergy crops
PRESENTER: Meng Xie

Eduardo Blumwald: Contribution of Serine Biosynthesis and Degradation to Carbon and Nitrogen Metabolism During Salinity stress in Poplar.
PRESENTER: Maria del Mar Rubio-Wilhelmi

Eduardo Blumwald: Rational design and testing of abiotic stress inducible synthetic promoters from poplar cisregulatory
elements
PRESENTER: Yongil Yang

Nicholas Bouskill: Microbial metabolic feedbacks following hydrological perturbation
PRESENTER: Stephany Chacon

Nanette Boyle: Advanced Metabolic Modeling of Diurnal Growth in *Chlamydomonas reinhardtii*
PRESENTER: Alex Metcalf

Tom Bruns: Functional and comparative genomics of pyrophilous fungi
PRESENTER: Andrei Steindorff

Daniel Buckley: Using Metagenomic Stable Isotope Probing to Identify Genomic Signatures of Bacterial Life History Strategies
PRESENTER: Samuel Barnett

Kristin Burnum-Johnson: Spatiotemporal mapping of the origin of linolenic acid in signaling transduction cascades during leaf deconstruction
PRESENTER: Kristin Burnum-Johnson

C

Jeffrey Cameron: Spatiotemporal Dynamics of Photosynthetic Metabolism in Single-Cells at Sub-Cellular Resolution
PRESENTER: Colin Gates

William Cannon: A Multiscale Model of Fungal Growth & Metabolism
PRESENTER: Jolene Britton

William Cannon: A Physics-Based Model of Fungal Metabolism and its Regulation
PRESENTER: William Cannon

John Carlson: Breeding Resilient, Disease-Resistant Switchgrass Cultivars for Marginal Lands
PRESENTER: John Carlson

Patrick Chain: Adaptive hydrophobic and hydrophilic surface response of fungi to changing growth conditions
PRESENTER: Saskia Bindschedler

Patrick Chain: Endo-hyphal microbiome: a core of bacteria associated with the nitrogen cycle
PRESENTER: Simone Lupini

Patrick Chain: Microscopic methods to identify, locate, and study endofungal bacteria.
PRESENTER: Demosthenes Morales

Patrick Chain: Exploring Mechanisms of Bacterial-Fungal Interactions Using *Ralstonia pickettii* Genomes Obtained from Diverse *Monosporascus* Isolates
PRESENTER: Aaron Robinson

Clint Chapple: Identification of the tyrosine- and phenylalanine-derived soluble metabolomes of sorghum
PRESENTER: Jacob Olson

Clint Chapple: Metabolic Source Isotopic Pair Labeling and Genome-Wide Association Are Complementary Tools for the Identification of Metabolite-Gene Associations in Plants
PRESENTER: Jeffrey Simpson

George Church: Precise genome editing in new microbial species using SSAP libraries and broad-host recombineering methods
PRESENTER: Gabriel Filsinger

George Church: Programming high-order combinatorial genetics with Cas9-mediated gene drive approach for cellular engineering
PRESENTER: Xiaoge Guo

George Church: Towards genetic incorporation of an Orthogonal Ribosome-tRNA pair and D-amino-acids in E.coli.
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: *In situ* sequencing for deciphering spatial taxonomic structures of plant-associated microbial communities.
PRESENTER: Andrew Pawlowski

George Church: High-throughput pooled functional screens via in-vivo production of Single-stranded DNA, toward applications in photosynthetic microbial hosts
PRESENTER: Max Schubert

John Coates: Prevalence of autotrophy and the characterization of carbon reduction in dissimilatory phosphite oxidizing microbes
PRESENTER: Sophia Ewens

John Coates: Environmental and Biological Constraints on Dissimilatory Phosphite Oxidizing Microorganisms
PRESENTER: Kyle Metcalfe

John Coates: Determination of Phosphite ($HPO_3^{-2)}$ by a New IC/MS/MS Method Using an ^{18}O-labeled HPO_3^{-2} Internal Standard
PRESENTER: Sepideh Sadeghi

Hugo Cuevas: Genomic dissection of anthracnose resistance response in the sorghum nested association mapping populations
PRESENTER: Hugo Cuevas

D

Gautam Dantas: Elucidating Aromatic Utilization Mechanisms in Engineered *Rhodococcus opacus* Strains for Lignin Valorization
PRESENTER: Winston Anthony

Brian Davison: Chemical and Morphological Structure of Solubilized Lignin Extracted via Ethanol, Tetrahydrofuran, and γ-Valerolactone Pretreatments from Wild Type and Transgenic Switchgrass
PRESENTER: Luna Liang

Brian Davison: Developing next-generation modeling techniques to analyze biomembrane neutron scattering data
PRESENTER: Jonathan Nickels

Brian Davison: Visualization of plant cell wall polymers using neutron scattering and deuterium labelling in *planta*
PRESENTER: Sai Venkatesh Pingali

José Dinneny: Divergence in stress tolerance mechanisms across the Brassicaceae family highlight strategies for maintaining growth and physiological balance under extreme environments
PRESENTER: Jose Dinneny

José Dinneny: Identification of Cell-type Marker Genes from Plant Single-cell RNA-seq Data Using Machine Learning
PRESENTER: Haidong Yan

Mitch Doktycz: Plant-Microbe Interfaces: Metaproteomics reveals shifts in microbial activity and regulation in assembling communities as they respond to environmental variation
PRESENTER: Paul Abraham

Mitch Doktycz: Plant-Microbe Interfaces: Simplified community approach to investigate multiple levels of selection in a host-microbiome relationship
PRESENTER: Dana Carper

Mitch Doktycz: Plant-Microbe Interfaces: Beneficial mycorrhization mediated by a lectin receptor-like kinase
PRESENTER: Jay Chen

Mitch Doktycz: Plant-Microbe Interfaces: Determining the rate and consequences of horizontal gene transfer in the rhizosphere by simulating lateral spread of salicylate catabolism genes
PRESENTER: Stephan Christel

Mitch Doktycz: Plant-Microbe Interfaces: Temporal dynamics of the Populus microbiome across scales
PRESENTER: Nicholas Dove

Mitch Doktycz: Plant-MIcrobe Interfaces: Comparative genome evolution of Populus root endophytes
PRESENTER: Yi Hong Ke

Mitch Doktycz: Plant-MIcrobe Interfaces: *In situ* chemical monitoring and imaging of Populus and its root microbiome using engineered devices with a porous membrane
PRESENTER: Muneeba Khalid

Mitch Doktycz: Plant-Microbe Interfaces: Functional characterization of Populus proteins involved in plant-microbe interactions
PRESENTER: Haiwei Lu

Mitch Doktycz: Plant-Microbe Interfaces: Exploring plant-microbe associations uisng Random Walk with Restart on multiplex networks
PRESENTER: Izaak Miller

Mitch Doktycz: Plant-Microbe Interfaces: Characterizing the perception of lipochitooligosaccharides signaling in fungi
PRESENTER: Tomas Rush

Mitch Doktycz: Bio-Scales Pilot Project: Defining gene function and its connection to ecosystem processes
PRESENTER: Chris Schadt

Mitch Doktycz: Plant-Microbe Interfaces: Structural basis for a bacterial Pip system plant effector recognition protein
PRESENTER: Amy Schaefer

Mitch Doktycz: Plant-Microbe Interfaces: Protein post-translational modifications in a 10-member microbial community from the Populus rhizosphere offer critical insights into how microbes adapt to changing environments
PRESENTER: Ivan Villalobos Solis

Mitch Doktycz: Plant-MIcrobe Interfaces: Formation and characterization of emergent microbial communities
PRESENTER: Jia Wang

Mitch Doktycz: Plant-Microbe Interfaces: Microbial responses to the chemical environment of the rhizosphere
PRESENTER: Amber Webb

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

Tim Donohue: A High-Efficacy CRISPR Interference System for Gene Function Discovery in Zymomonas mobilis
PRESENTER: Amy Enright

Tim Donohue: Disruption of Brachypodium distachyon Lichenase Alters Metabolism of Mixed-linkage Glucan and Starch
PRESENTER: Mingzhu Fan

Tim Donohue: Multiomics analysis of mitochondrial versus cytosolic compartmentalization of the isobutanol pathway in *Saccharomyces cerevisiae*
PRESENTER: Francesca Gambacorta

Tim Donohue: Analysis of Carbon Capture in Lignocellulosic Biorefineries
PRESENTER: Caleb Geissler

Tim Donohue: Evolution of Cellulose Structure throughout gamma-Valerolactone-Assisted and Enzymatic Biomass Deconstruction
PRESENTER: Elise Gilcher

Tim Donohue: Engineering DXS for Improved Flux into the MEP Pathway
PRESENTER: Jonathan Greenhalgh

Tim Donohue: Reinforcement Learning to Optimize Medium Chain Fatty Acid Production from Lignocellulosic Stillage using Anaerobic Microbial Communities
PRESENTER: Madeline Hayes

Tim Donohue: Plant-soil-microbial interactions in detritusphere and its impact on N_2O emission
PRESENTER: Alyssa Kim

Tim Donohue: Enzymatic Deconstruction of Cellulosic Biomass
PRESENTER: Nathaniel Kuch

Tim Donohue: Engineering naringenin into the lignins of transgenic poplar
PRESENTER: Elizabeth Mahon

Tim Donohue: Metabolomic and Proteomic Analysis of *Zymomonas mobilis* During Nitrogen Fixation Reveals Metabolic Remodeling of Biofuel Producing Pathways
PRESENTER: Julia Martien

Tim Donohue: Investigating the Interplay between the Phyllosphere Microbiome, Epicuticular Wax, and Root Mucilage on Sorghum Resilience to Water and Nitrogen Limitation.
PRESENTER: Marco Mechan Llontop

Tim Donohue: Deletion of the *ntrYX* two Component System in *Rhodobacter sphaeroides* Causes the Generation of Diverse Extracellular Membrane Structures
PRESENTER: Daniel Parrell

Tim Donohue: Redundancy in Aromaric O-demethylation and Ring Opening Reactions in *Novosphingobium aromaticivorans* and their Impact in Biological Conversion of Lignin
PRESENTER: Jose Perez

Tim Donohue: Investigating the Effects of Environmental Stress on Feedstock Quality and Biofuel Production by Field-to-Fuel Optimization Research Pipeline
PRESENTER: Trey Sato

Tim Donohue: Lignin Valorization by Integrating Chemical Depolymerization and Microbial Funneling
PRESENTER: Canan Sener

Tim Donohue: Drought Caused a 33% Reduction in Switchgrass (*Panicum Virgatum L.*) Biomass with Minimum Effects on Net CO_2 Assimilation
PRESENTER: Mauricio Tejera Nieves

Tim Donohue: Chemical genomic profiling of hydrolysates and toxins: implications for yeast strain engineering
PRESENTER: Elena Vanacloig

Tim Donohue: Impact of genome assemblies, genotyping methods, variant types, and ploidy levels on genomic prediction in switchgrass
PRESENTER: Peipei Wang

Tim Donohue: Perenniality Drives Soil Microarthropod Community Differences Across Three Potential Bioenergy Cropping Systems
PRESENTER: Allison Zahorec

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

John Dunbar: Microbial Control of Mineral-Bound Carbon
PRESENTER: Joany Babilonia

John Dunbar: Cycles of Wetting and Drying Reduce Carbon Efflux and Litter Decomposition in Soil Microcosms
PRESENTER: Tayte Campbell

John Dunbar: Microbial driven variation in carbon flow & stabilization during root litter decomposition
PRESENTER: Rae DeVan

John Dunbar: Influence of Microbial Surface Litter Decomposer Communities on CO_2 Emissions from Natural Soils
PRESENTER: Sanna Sevanto

John Dunbar: Comparing In-Situ, Individual Bacterial Growth Rates in Cropped and Successional Soils Using a 16S rRNA Internal Standard
PRESENTER: Cassandra Wattenburger

Mary Dunlop: Live-cell imaging of E. coli biofuel synthesis by spectroscopic stimulated Raman scattering microscopy
PRESENTER: Haonan Lin

Mary Dunlop: Single Cell Chemical Imaging with Stimulated Raman Scattering for Biofuel Production Screening
PRESENTER: Nathan Tague

E

Robert Egbert: Identifying growth harness actuator genes for Pseudomonas putida
PRESENTER: Shara Balakrishnan

Robert Egbert: Serine integrase-assisted genome engineering (SAGE) enables efficient, iterative site-specific genome engineering in bacteria
PRESENTER: Joshua Elmore

Robert Egbert: 3D Cartography of the Sorghum Rhizosphere
PRESENTER: Pubudu Handakumbura

Robert Egbert: Rapid Functional Annotation Using Degenerate Kmers
PRESENTER: Jason McDermott

Robert Egbert: Identifying persistence control strategies in plant-growth promoting soil bacteria with RB-TnSeq, CRISPRi, and proteomics
PRESENTER: Aaron Ogden

Edward Eisenstein: Transgenic Poplar Lines Reveal Host Genes Involved in Defense Against Rust
PRESENTER: Edward Eisenstein

Emiley Eloe-Fadrosh: Enhanced metadata standards supported by the National Microbiome Data Collaborative
PRESENTER: William Duncan

Emiley Eloe-Fadrosh: Multi-omics workflows to support data integration for the National Microbiome Data Collaborative
PRESENTER: Bin Hu

Emiley Eloe-Fadrosh: The National Microbiome Data Collaborative community survey on microbiome research data
PRESENTER: Pajau Vangay

Joanne Emerson: Infective Viruses and Inert Virions: Illuminating Abundant Unknowns in Terrestrial Biogeochemical Cycles
PRESENTER: Joanne Emerson

James Evans: Opportunities Linking Omics and Structural Biology at PNNL: Excelling at Cryo-EM
PRESENTER: James Evans

Sarah Evans: Nitrogen Addition Impacts on Structure and Function of the Switchgrass Root-associated Diazotrophic Community
PRESENTER: Darian Smercina

Andrea Eveland: PhytoOracle: Leveraging Open-Source Tools for Phenomic Data Processing at Scale
PRESENTER: Michele Cosi

Andrea Eveland: Elucidating the Molecular Mechanisms Underlying Drought Resilience in Sorghum
PRESENTER: Andrea Eveland

F

Kjiersten Fagnan: Genesearch: A Sequence Similarity Search Service for Genomics Workflows
PRESENTER: Kjiersten Fagnan

Mary Firestone: Cross-Kingdom Characterization of Community Dynamics and C flow in Grassland Soils
PRESENTER: Javier Ceja-Navarro

Mary Firestone: Belowground Allocation and Dynamics of Recently-Fixed Plant Carbon in a California Annual Grassland Soil
PRESENTER: Christina Fossum

Mary Firestone: Carbon Transported by Arbuscular Mycorrhizal Fungi to Soil Alters the Characteristics of Soil Organic Carbon as well as the Soil Microbial Community
PRESENTER: Anne Kakouridis

Mary Firestone: Soil Viral Community Composition Differs Spatially and in Response to Wet-up in Mediterranean Grasslands
PRESENTER: Christian Santos-Medellin

Brian Fox: Identification and Characterization of a New Set of Monocot BAHD Monolignol Transferases
PRESENTER: Rebecca Smith

G

Ryan Gill: Extension of Genetic Circuit Design Optimization to Industrially-relevant Organisms
PRESENTER: Hamid Doosthosseini

Ryan Gill: Engineering of Regulatory Networks for Complex Phenotypes in *E. coli* and *S. cerevisiae*
PRESENTER: Emily Freed

Ryan Gill: Integration of Random Barcode Transposon Sequencing Applications into KBase
PRESENTER: Omree Gal-Oz

Ryan Gill: Functional genetic screening in the thermotolerant yeast *Kluyveromyces marxianus*
PRESENTER: Maria Priscila Lacerda

Michael Guarnieri: IMAGINE BioSecurity: Integrative Modeling and Genome-Scale Engineering for Biosystems Security
PRESENTER: Michael Guarnieri

H

William Hancock: Nanoscale Dynamics of Cellulase TrCel7A Digesting Cellulose
PRESENTER: Zachary Haviland

William Hancock: Characterizing the degradation of cellulose by combinations of cellulolytic enzymes
PRESENTER: Thomas Starr

Christopher Henry: Integrating BER Facilities and Resources for Multi-Omics to Reactive Transport Modeling Workflows
PRESENTER: Timothy Scheibe

Kirsten Hofmockel: Quantitative Analysis of the Fate of Microbial Residues in Biofuel Soils
PRESENTER: Kirsten Hofmockel

Bruce Hungate: The temperature sensitivity of soil: microbial biodiversity, growth, and carbon mineralization
PRESENTER: Ember Morrissey

Bruce Hungate: The Phylogenetic Organization of Tundra Bacterial Growth in Response to Short-Term and Long-Term Warming
PRESENTER: Jeffrey Propster

Bruce Hungate: Scaling of ^{18}O-informed Microbial Growth Rates Links Microbial Activity to Soil Carbon Flux
PRESENTER: Bram Stone

J

Tiffany Jamann: Differential Regulation of Maize and Sorghum Orthologs in Response to the Fungal Pathogen *Setosphaeria turcica*
PRESENTER: Pragya Adhikari

Tiffany Jamann: Genomic regions associated with pathogenicity in *Exserohilum turcicum* identified by linkage mapping in a biparental population
PRESENTER: Pummi Singh

Janet Jansson: Fungal Nutrient Acquisition and Transport in Soil Micromodels is Regulated by Organic Acid Chelation and Specific Membrane Transporters
PRESENTER: Arunima Bhattacharjee

Janet Jansson: Unveiling the Molecular Mechanisms Underlying the Microbiome Response to Soil Rewetting
PRESENTER: Mary Lipton

Janet Jansson: Interspecies Interactions During Chitin Decomposition in a Naturally Enriched Model Soil Microbial Consortium
PRESENTER: Ryan McClure

Janet Jansson: Prediction of Spatial Assemblies in Soil Microbiomes Governed by Interspecies Interactions and Environmental Gradients
PRESENTER: Hyun-Seob Song

Janet Jansson: Soil Moisture Impacts Composition, Activity and Ecosystem Functions of Soil DNA and RNA Viruses
PRESENTER: Ruonan Wu

Kolby Jardine: Volatile Emissions and Cell Wall Ester Signatures of Abiotic Stress in Poplar
PRESENTER: Rebecca Dewhirst

Michael Jewett: Closing the Carbon Cycle: Design, Optimization and Scaling-Up Production of Carbon-Negative Platform Chemicals
PRESENTER: Steve Brown

Michael Jewett: Multi-omics analyses reveal temperature-induced metabolic changes in the industrially relevant microbe *Clostridium autoethanogenum* affecting its product profile
PRESENTER: Payal Chirania

Michael Jewett: Rapid Prototyping of Novel Bioproduct Pathways in an Acetogen through Integrated Computational Modeling and High-throughput Candidate Screening
PRESENTER: Zach Cowden

Michael Jewett: Modeling Tools to Predict Metabolic Network Behavior in Non-Model Systems
PRESENTER: Jacob Martin

Michael Jewett: Spacer2PAM: an R Package for CRISPR Protospacer Adjacent Motif Prediction from Spacer Sequences
PRESENTER: Grant Rybnicky

Yongqin Jiao: From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms
PRESENTER: Dan Park

Yongqin Jiao: An Electrochemical Model of Rhizosphere ROS Generation to Analyze Commensal Engineering Strategies
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: Environmental and Genetic Effects on Switchgrass (*Panicum virgatum*) Biomass Composition across Diverse Environments
PRESENTER: Laura Bartley

Thomas Juenger: Genetic signatures influencing host-microbiome interactions in switchgrass (*Panicum virgatum*)
PRESENTER: Joseph Edwards

Thomas Juenger: The switchgrass genome: polyploidy and introgressions facilitate climate adaptation and biomass yield
PRESENTER: John Lovell

Thomas Juenger: Community Resources to Study Switchgrass Adaptation Using Genome-Wide Association
PRESENTER: Alice MacQueen

K

Jay Keasling: Integrated *Pseudomonas putida* Strain Design For Maximizing Biomass Conversion To Biofuels And Bioproducts.
PRESENTER: Deepanwita Banerjee

Jay Keasling: Production Cost and Carbon Footprint of High Performance Biomass-Derived Dimethylcyclooctane as a Jet fuel Blendstock
PRESENTER: Nawa Baral

Jay Keasling: Title: An semi-automated workflow for high-throughput and quantitative proteomic analysis of metabolic engineered microorganisms
PRESENTER: Yan Chen

Jay Keasling: Multiple Ions in Ionic Liquid for Lignocellulosic Biomass Processing – A Boon or a Bane
PRESENTER: Hemant Choudhary

Jay Keasling: The development of an automated platform for the use of Polyketide Synthases for Biofuels and Sustainable Chemical Production
PRESENTER: Cameron Coates

Jay Keasling: A mass spectrometry based high-throughput platform to study lignin modifying enzymes
PRESENTER: Kai Deng

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

Jay Keasling: ART: A machine learning Automated Recommendation Tool for synthetic biology
PRESENTER: Hector Garcia Martin

Jay Keasling: Development of Automated Workflows and Data Processing to Enable the Conversion of Bioenergy Crops into Biofuels and Bioproducts
PRESENTER: Nurgul Kaplan

Jay Keasling: Lignin manipulation via plant synthetic biology to produce a biodegradable polyester precursor with concurrent improvement in biomass quality for biofuel production
PRESENTER: Chien-Yuan Lin

Jay Keasling: Engineering *Rhodosporidium toruloides* for Bioproduction of Polyketide Synthase and p-Coumarate Derived Compounds
PRESENTER: Peter Otoupal

Jay Keasling: *Pseudomonas putida* soil isolates that metabolize C5 sugars and lignin-derived aromatics
PRESENTER: Mee-Rye Park

Jay Keasling: Expression of fungal laccases in *Pichia pastoris* and characterization using lignin depolymerization
PRESENTER: Mai Pham

Jay Keasling: Microscale thermophoresis as a powerful tool for screening glycosyltransferases involved in cell wall biosynthesis
PRESENTER: Wanchen Shao

Jay Keasling: Engineering Plants with Novel Metabolic Pathways as a Production Platform for Bioproducts
PRESENTER: Khanh Vuu

Jay Keasling: Modifying terpenes for the development of biofuels and bioproducts
PRESENTER: Xi Wang

Jay Keasling: High Throughput Bioengineering Using a Microfluidic Platform
PRESENTER: Lauren Washburn / Jess Sustarich

Matias Kirst: Evolution of root nodule symbiosis & engineering of symbiotic nitrogen fixation in *Populus* sp.
PRESENTER: Thomas Irving

Matias Kirst: Identification of the Minimal Genetic Toolkit Required for Nitrogen Fixation Using Comparative Genomics and Single-Cell Transcriptomics
PRESENTER: Matias Kirst

Matias Kirst: Leveraging temporal change in chromatin accessibility to predict regulators of N-fixing symbiosis in *Medicago* with dynamic regulatory module networks (DRMNs)
PRESENTER: Sara Knaack

L

Andrew Leakey: Innovations to the DayCent Biogeochemical Model to Better Simulate Carbon and Nitrogen Cycling in Bioenergy Crop Systems with Increasing Climate Variability
PRESENTER: Danielle Berardi

Andrew Leakey: Engineering Vegetative Lipids in a Fast-Growing and High-Biomass Arabidopsis Line
PRESENTER: Yingqi Cai

Andrew Leakey: Biorefinery of Lignocellulosic Carbohydrates: Production of Lipid and Ethanol through Engineered Microbial Conversion
PRESENTER: Ming-Hsun Cheng

Andrew Leakey: Deep Learning of Transcriptional Regulation in *Issatchenkia orientalis*
PRESENTER: Payam Dibaeinia

Andrew Leakey: Discovering Transcription Regulation Networks in Bioenergy-relevant Yeast Species
PRESENTER: Veronika Dubinkina

Andrew Leakey: Towards an Efficient Multiallelic Gene Editing Platform for Highly Polyploid Sugarcane
PRESENTER: Ayman Eid

Andrew Leakey: Nitrogen Delivery Efficiency in the Mississippi River Basin
PRESENTER: Theodore Hartman

Andrew Leakey: L-malic Acid Production from Xylose by Engineered *Saccharomyces cerevisiae*
PRESENTER: Nam Kyu Kang

Andrew Leakey: Testing Unifying Theories of Ozone Response in C_4 Bioenergy Grasses
PRESENTER: Shuai Li

Andrew Leakey: Optimizing Measurement Methods for N_2 Fixation in *Miscanthus* × *giganteus*
PRESENTER: Di Liang

Andrew Leakey: Engineering of the Non-Model Yeast *Issatchenkia orientalis* to Produce Organic Acids
PRESENTER: Teresa Martin

Andrew Leakey: The Genome of the Perennial Biomass Grass *Miscanthus sinensis*
PRESENTER: Therese Mitros

Andrew Leakey: Genetic Designs Targeting Accumulation of Vegetative Lipids, Multiplexing of Editing Reagents for "In Context Promoter Bashing" and Gaining Insight on Roots Exudates Impact on Shaping the Root Microbiome in Sorghum (*Sorghum bicolor* (L.) Moench
PRESENTER: Truyen Quach

Andrew Leakey: Responsiveness of Miscanthus and Switchgrass Yields to Stand Age and Nitrogen Fertilization: A Meta-regression Analysis
PRESENTER: Bijay Sharma

Andrew Leakey: Optimizing Miscanthus Regeneration and Transformation
PRESENTER: Anthony Trieu

Andrew Leakey: Quantifying Root Carbon Rhizodeposition from Bioenergy Cropping Systems in the Midwest United States
PRESENTER: Adam von Haden

Andrew Leakey: Improving Photosynthetic Efficiency of C4 Energy Crops: A Dynamic Modeling Analysis
PRESENTER: Yu Wang

Andrew Leakey: Engineering *Yarrowia lipolytica* to Produce 3-acetyl-1,2-diacyl-sn-glycerol
PRESENTER: Qiang Yan

Andrew Leakey: Comparison of Actively Growing (RNA) and Potentially Active (DNA) Soil Microbial Communities in *Miscanthus x giganteus*
PRESENTER: Jihoon Yang

Karen Lloyd: Metagenomic binning of high latitude (79°N), mineral permafrost active layer
PRESENTER: Karen Lloyd

Karen Lloyd: Contribution Evenness: A functional redundancy metric for microbially-mediated biogeochemical rates and processes
PRESENTER: Taylor Royalty

Karen Lloyd: Microbial organic matter degradation in the active layer of Svalbard permafrost
PRESENTER: Katie Sipes

Stephen Long: Optimizing Lipogenic Factors for Vegetative Oil Accumulation
PRESENTER: Sanket Anaokar

Stephen Long: Metabolic Engineering of Energycane for Hyperaccumulation of Triacylglycerol and Improved Biomass Production
PRESENTER: Viet Dang Cao

Stephen Long: Development of Sustainable Transformation of *Miscanthus* × *giganteus* to Improve Photosynthesis
PRESENTER: Kher Xing Chan

Stephen Long: Improving Photosynthetic Efficiency of C_4 Bioenergy Crops in Fluctuating Lights
PRESENTER: Moonsub Lee

Stephen Long: Extending the Use of Time-domain ^{1}H-NMR for Rapid and Non-invasive Quantification and Characterization of *In-situ* Lipids in Transgenic Bioenergy Crops
PRESENTER: Shraddha Maitra

Chaofu Lu: Deciphering genetic and physiological mechanisms that enhance nitrogen use efficiency and seed oil accumulation in camelina
PRESENTER: Chaofu Lu

Ting Lu: Co-consumption of mixed sugars through the division of labor (DOL) in a synthetic *Saccharomyces cerevisiae* consortium
PRESENTER: Jonghyeok Shin

Ting Lu: Probing the Stability of a Synthetic Symbiotic Microbial Consortium / Systematic Evaluation of the Stability of the synthetic LAB-Yeast Consortium
PRESENTER: Yongping Xin

M

Hiroshi Maeda: Constructing the Nitrogen Flux Maps (NFMs) of Plants
PRESENTER: Hiroshi Maeda

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

Chris Marx: Using Gene Editing and an Accumulated Bioproduct as a Reporter for Genotypic and Phenotypic Heterogeneity in Growth-vs-Production for *Methylobacterium extorquens* Conversion of Aromatics to Butanol
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: Discovering peptidic natural products by integrating computational mass spectrometry and genome mining
PRESENTER: Liu Cao

James Moran: Spatially resolved proteomic analysis to reveal controls on rhizosphere microbial recruitment and track carbon exchange between plants and microorganisms
PRESENTER: Jim Moran

Paula Mouser: Microbial Osmotolerance Mechanisms in Hydraulically Fractured Shale Elucidated Through Metagenomics Analysis
PRESENTER: Jishnu Adhikari

Paula Mouser: 'Omics Analyses of the Hydraulically Fractured Shale Isolate *Halanaerobium* Highlights Membrane Modifications that Underpin Adaptation Under Deep Subsurface Biogeochemical Drivers
PRESENTER: Fabrizio Colosimo

Wellington Muchero: The Plasminogen-Apple-Nematode (PAN) domain is a key negative regulator of Jasmonic Acid and Ethylene-mediated defense responses in plants.
PRESENTER: Wellington Muchero

N

Krishna Niyogi: Multifactorial Nutrient-Dependent Proteomics Elucidates Lipid Accumulation and Regulation of Photosynthesis
PRESENTER: Tim Jeffers

Krishna Niyogi: Metabolic Model of *Chromochloris zofingiensis*, an Emerging Model Green Alga for Sustainable Fuel Production
PRESENTER: Michelle Meagher

Krishna Niyogi: Structural Gene Organization in *Chromochloris zofingiensis* Can Drive Advancements in Synthetic Biology
PRESENTER: Jeffery Mosley

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: Catharine Adams

Trent Northen: TEAMS: Advancing microbiome science through high-throughput, automated EcoFAB experiments
PRESENTER: Peter Andeer

Trent Northen: Generating a Stable, Reproducible Rhizosphere Microbial Community
PRESENTER: Joanna Coker

Trent Northen: Techniques for *in situ* DNA delivery and targeted editing within microbial communities
PRESENTER: Spencer Diamond

Trent Northen: Towards integrated rhizosphere microbial community modeling through bottom-up COMETS genome-scale metabolic simulations and top-down data analysis
PRESENTER: Ilija Dukovski

Trent Northen: Phenotypic Characterization of *Brachypodium distachyon* and a Synthetic Community for Dissecting Plant-Microbial Community Interactions in Fabricated Ecosystems (EcoFAB)
PRESENTER: Hsiao-Han Lin

Trent Northen: Bioinformatic Guided Enrichments to Develop Representative Rhizosphere Communities
PRESENTER: Mon Yee

Trent Northen: Spanning laboratory ecosystem scales: insights into environmental complexity with EcoFABs andEcoPODs
PRESENTER: Kateryna Zhalnina

Erin Nuccio: Crosstalk: Interkingdom interactions in the mycorrhizal hyphosphere and ramifications for soil C cycling
PRESENTER: Erin Nuccio

O

Victoria Orphan: Microbial dynamics and syntrophic interactions at the pore scale: towards an integration of reactive transport and microbial cell models
PRESENTER: Nidhi Gupta

Victoria Orphan: Ecology, Diversity, and Biogeochemical Contribution of Viruses in Methane-Rich Sediments
PRESENTER: Alon Philosof

Victoria Orphan: Uncovering Syntrophies within Methane-Oxidizing Microbial Consortia in Sediments: Synthesizing Insights from the Subcellular to the Population Scales
PRESENTER: Kriti Sharma

P

Himadri Pakrasi: Systems analysis of a fast growing N_2-fixing cyanobacterium for production of advanced biofuels and nitrogen-containing petrochemical replacement compounds
PRESENTER: Anindita Banerjee

Terry Papoutsakis: Dynamic modeling of a synthetic *Clostridia* triple co-culture for producing medium-chain fatty acids from glucose
PRESENTER: Charles Foster

Terry Papoutsakis: Syntrophic Co-Cultures of *Clostridium* Organisms to Produce Higher Alcohols and Other C6-C8 Metabolites
PRESENTER: Jonathan Otten

Kabir Peay: Patterns of diversity in the North American Populus mycobiome
PRESENTER: Michael Van Nuland

Gary Peter: Enhanced Resistance Pines for Improved Renewable Biofuel and Chemical Production
PRESENTER: Gary Peter

Jennifer Pett-Ridge: Historic Precipitation Regimes Influence Microbial Population Dynamics in Response to Seasonal Rewetting in Mediterranean-Grassland Soil
PRESENTER: Steven J. Blazewicz

Jennifer Pett-Ridge: Cell Size Constraints on Microbial Ecophysiology and Implications for Soil Carbon Cycling
PRESENTER: Gianna Marschmann

Jennifer Pett-Ridge: Prevalence and Automated Curation of Local Errors from Metagenomic Assembled Genomes
PRESENTER: Rohan Sachdeva

Jennifer Pett-Ridge: The Influence of Drought on Carbon-Use Efficiency and Soil Carbon Formation in Rhizosphere and Detritusphere Microbial Communities
PRESENTER: Noah Sokol

Jennifer Pett-Ridge: Virus Activity in Soil Revealed Through SIP-Metagenomics
PRESENTER: Olivier Zablocki

R

Chris Rao: Transcriptomic and Metabolomic Analysis of Nitrogen and Carbon Metabolism in *Saccharomyces cerevisiae* and *Rhodotorula toruloides*
PRESENTER: William Woodruff

Sue Rhee: Nanoparticle-Mediated Transformation of Sorghum towards the Determination of a Subcellular Metabolic Network Map
PRESENTER: Christopher Jackson

Sue Rhee: High-Throughput Determination of a Subcellular Metabolic Network Map of Plants
PRESENTER: Jason Thomas

Simon Roux: An Integrated Machine-Learning Framework for Reliable Host Prediction of Uncultivated Phages
PRESENTER: Simon Roux

S

Howard Salis: Non-Repetitive Promoters and Ribosome Binding Sites to Control *Clostridium autoethanogenum* Gene Expression Levels during Syngas Fermentation
PRESENTER: Audrey Harris

Howard Salis: Multiplex Genome Engineering for Bioproduction of 3-Hydroxypropionic Acid and 1,3-Propanediol from Waste Gases
PRESENTER: Fungmin (Eric) Liew

Rodriguez Salvachua: Elucidation of aromatic catabolic pathways in white-rot fungi
PRESENTER: Davinia Salvachua

Jonathan Schilling: Using aggregated field collections data and the novel R package “fungarium” to investigate fungal traits
PRESENTER: Hunter Simpson

James Schnable: Meta-Analysis Identifies Pleiotropic Loci Controlling Phenotypic Trade-offs in Sorghum
PRESENTER: Ravi Mural

Danny Schnell: Overexpression of the Phosphatidylcholine: diacylglycerol cholinephosphotransferase (PDCT) Gene Has Increased Carbon Flux Toward Triacylglycerol (TAG) Synthesis in Camelina (*Camelina sativa*) Seeds
PRESENTER: Hesham Abdullah

Danny Schnell: The metabolic origins of non-photorespiratory CO_2 release during photosynthesis: A metabolic flux analysis
PRESENTER: Yair Shachar-Hill

Danny Schnell: Integration of a Synthetic CO_2 Fixation Cycle into *Camelina sativa*
PRESENTER: Nathan Wilson

Joseph Schoeniger: Computational Front End for Creating Universal Phage-based Vectors for Bacteria
PRESENTER: Catherine Mageeney

John Sedbrook: Developing mutant resources for pennycress to domesticate and improve crop resilience
PRESENTER: Ratan Chopra

John Sedbrook: Advancing field pennycress as a new oilseed biofuels feedstock that does not require new land commitments – Improving pennycress stand establishment
PRESENTER: Michael Marks

John Sedbrook: Interrogating pennycress natural and induced variation to improve abiotic stress tolerance and oilseed bioenergy crop resilience
PRESENTER: John Sedbrook

Jeremy Semrau: Oxygen Generation via Water Splitting by a Novel Biogenic Metal Binding Compound
PRESENTER: Alan DiSpirito

Jeremy Semrau: Methanotrophs Produce Diverse Chalkophores to Compete for Copper
PRESENTER: Christina Kang-Yun

Jeffrey Skolnick: Integrated deep-learning computational approach to proteome annotation
PRESENTER: Mu Gao

Claudia Solis-Lemus: Bayesian Conditional Auto-Regressive LASSO Models to Learn Sparse Networks with Predictors
PRESENTER: Yunyi Shen

Kevin Solomon: High quality anaerobic fungal genome assemblies and annotation for study and optimization of lignocellulose conversion
PRESENTER: Casey Hooker

Rhona Stuart: Characterizing Algal Exo-Metabolites and Their Impacts on Algae and Associated Bacteria
PRESENTER: Vanessa Brisson

Rhona Stuart: Integrated Experimentation and Trait-based Model Exploration of Algal-Bacterial Interaction Mechanisms
PRESENTER: Eoin Brodie

Rhona Stuart: Context Dependent Mycorrhizal Resource Exchange in Bioenergy Cropping Systems
PRESENTER: Rachel Hestrin

Rhona Stuart: Impacts of physical proximity and metabolite diffusion on algal-bacteria interactions
PRESENTER: Hyungseok Kim

Rhona Stuart: Characterizing algicidal bacteria and their mechanisms in antagonistic algal-bacterial interactions
PRESENTER: Megan Morris

Rhona Stuart: System-level analysis of metabolism of *Kordia algicida* and its interaction with *Phaeodactylum tricornutum*
PRESENTER: Ali Navid

Matthew Sullivan: Coupled Metabolomics and Transcriptomics Analyses Reveal Active Dynamics of Infection in Virocells
PRESENTER: Roya AminiTabrizi

Matthew Sullivan: Expanding standards in viromics: *in silico* evaluation of viral identification, taxonomy, and auxiliary metabolic genes (AMGs) curation
PRESENTER: Akbar Adjie Pratama

Matthew Sullivan: Viral ecology through time and along a permafrost thaw gradient
PRESENTER: Christine Sun

T

Bob Tabita: A Nitrogenase-like Methylthio-alkane Reductase Complex Catalyzes Anaerobic Methane, Ethylene, and Methionine Biosynthesis
PRESENTER: Srividya Murali

Bob Tabita: Engineering of Enhanced Microbial Anaerobic Ethylene Synthesis Through Predictive Modeling and Metagenomic Functional Gene Discovery
PRESENTER: Justin North

Michiko Taga: High-Throughput Isolation of Bacteria to Dissect Corrinoid Based Interactions in Soil
PRESENTER: Zoila Alvarez-Aponte

Michiko Taga: Corrinoids as model nutrients to probe microbial interactions in a soil ecosystem
PRESENTER: Zachary Hallberg

Michiko Taga: Reliable bioinformatic prediction of cobamide biosynthesis by core biosynthesis genes and taxonomy
PRESENTER: Alexa Nicolas

Baas Tas: Understanding the microbial controls on biogeochemical cycles in permafrost ecosystems
PRESENTER: Neslihan Tas

Gail Taylor: A new 15 acre field plantation for CBI harnessing the natural diversity of *Populus trichocarpa* and determining the genetic basis of drought tolerance
PRESENTER: Jack Bailey-Bale

Gail Taylor: Identifying traits that underpin the genetic basis of drought tolerance in bioenergy poplar
PRESENTER: Marie Klein

Malak Tfaily: Untargeted metabolomics by high resolution LC-MS/MS revealed different metabolic profiles of oaks (*Quercus* spp.)
PRESENTER: Nathalia Graf Grachet

Cong Trinh: Proteomic Analysis of Robust Yarrowia lipolytica Isolates Reveals Key Processes Impacting Sugar Utilization and Lipid Degradation during Growth on Biomass Hydrolysate
PRESENTER: Caleb Walker

Cong Trinh: Multi-OMICs Profiling Reveals Key Genes and Cellular Processes Underlying Ionic Liquid Robustness in *Yarrowia lipolytica*
PRESENTER: Caleb Walker

Danielle Tullman-Ercek: Employing Bacterial Microcompartments To Create Privileged Redox Pools for Biofuel Production
PRESENTER: Charlotte Abrahamson

Gerald Tuskan: Modeling and NMR Methods to Probe Spatial Arrangements in Biomolecules: Towards predictive models of plant cell wall structure
PRESENTER: Bennett Addison

Gerald Tuskan: Effects of *Claviceps spp./Epicoccum andropogonis* inoculation on switchgrass phenotypic traits
PRESENTER: Bochra Bahri

Gerald Tuskan: Techno-economic analysis and life cycle assessment of a biorefinery utilizing reductive catalytic fractionation
PRESENTER: Andrew Bartling

Gerald Tuskan: Engineering a Cytochrome P450 System for Oxidative Demethylation of Lignin-Related Aromatics
PRESENTER: Alissa Bleem

Gerald Tuskan: Thermodynamic analysis of *C. thermocellum* glycolysis using deuterated water (^2H_2O) during high substrate loading fermentations
PRESENTER: Melanie Callaghan

Gerald Tuskan: High Resolution Analysis of Recombination Rates in *Populus trichocarpa*
PRESENTER: Stephen DiFazio

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
PRESENTER: Jacob Fenster

Gerald Tuskan: Simulating landscape-scale impacts of switchgrass nitrogen use efficiency
PRESENTER: John Field

Gerald Tuskan: Disc Milling of Fermented Corn Stover to Increase its Accessibility to Fermentation by *Clostridium thermocellum*
PRESENTER: Sanchari Ghosh

Gerald Tuskan: Exploring Catalytic Conditions for C-C Bond Cleavage of Lignin-Based Compounds
PRESENTER: Nina Gu

Gerald Tuskan: Field experiments of switchgrass GWAS population for increased sustainability: nitrogen-use efficiency and rust disease resistance
PRESENTER: Mitra Mazarei

Gerald Tuskan: Using Iterative Random Forest to Predict the Progeny of Crosses in *Populus trichocarpa*
PRESENTER: Jonathon Romero

Gerald Tuskan: Identifying growth-coupled genetic loci using integrated metabolic modeling in populus
PRESENTER: Debolina Sarkar

Gerald Tuskan: Systems Metabolic Engineering of *Clostridium thermocellum* for Direct Conversion of Cellulosic Biomass to Designer C4-derived Esters
PRESENTER: Hyeongmin Seo

Gerald Tuskan: Continuous hydrodeoxygenation of neat poplar lignin oil to jet-range aromatic hydrocarbons with molybdenum carbide
PRESENTER: Michael Stone

Gerald Tuskan: A Rapid Semi-Automated Phenotyping System to Capture the Highly Useful Diameter at Breast Height In *Populus trichocarpa*
PRESENTER: Jared Streich

Gerald Tuskan: Engineering Plant Cell Wall Polysaccharide O-Acetyltransferases with Altered Specificity
PRESENTER: Hsin-Tzu Wang

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

Gerald Tuskan: Physiological responses of *Populus trichocarpa* genotypes to drought
PRESENTER: Marvin Wright

Gerald Tuskan: Application of Base Editing Technology in Poplar
PRESENTER: Guoliang Yuan

Keith Tyo: Characterizing Thiamine Diphosphate Dependent Enzymes for Promiscuous C-C Bond Formation Catalysis
PRESENTER: Tracey Dinh

U

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

V

Kranthi Varala: Novel seed lipid regulators identified by gene network inference
PRESENTER: Kranthi Varala

W

Jin Wang: Meta-Transcriptomic Network Analysis of Inter-Species Interactions in a Cyanobacterial-Methanotroph Coculture
PRESENTER: Alexander Beliaev

Jin Wang: Harnessing Methanotroph-Photoautotroph Interactions for Biogas Conversion
PRESENTER: Jin Wang

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

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

Ian Wheeldon: Developing the yeast *Kluyveromyces marxianus* as a thermotolerant bioproduction host
PRESENTER: Ian Wheeldon

Travis Wheeler: FATHMM: Frameshift Aware Translated Hidden Markov Models
PRESENTER: Genevieve Krause

Thea Whitman: Experimental pyrocosms demonstrate key features of the autecology of post-fire fungi
PRESENTER: Thomas Bruns

Thea Whitman: Life After Fire: Microbial Community Dynamics Over Time in Fire-Affected Soils
PRESENTER: Monika Fischer

Mari Winkler: Bridging Scales: Conceptualizing microbe-climate links in wetland ecosystems
PRESENTER: Pieter Candry

Mari Winkler: Hydrogel Beads to Encapsulate Sediment Microbes as a Strategy to Quantify Climate Impacts on Microscale Biogeochemical Activity
PRESENTER: Julie Johnston

Mari Winkler: Metabolic Modeling of Synthetic Estuarine Wetlands Microbial Communities in response to Climate Change
PRESENTER: Chongle Pan

Kelly Wrighton: From Genomes to Methane Emission: Targeting Critical Knowledge Gaps in Wetland Soils
PRESENTER: Adrienne Narrowe

Y

Todd Yeates: Advancing the molecular understanding of growth in algae and plants
PRESENTER: Siobhan Braybrook

Todd Yeates: Development and Deployment of Enabling Technologies at the UCLA-DOE Institute
PRESENTER: Marcus Gallagher-Jones

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 using natural language processing and machine learning
PRESENTER: Carlos Soto

Jamey Young: Rapid Strain Phenotyping and Metabolic Flux Analysis to Accelerate Engineering of Microorganisms.
PRESENTER: Berkley Ellis

Jamey Young: Mitigating guanidine toxicity and manipulating circadian clock for enhanced ethylene production in engineered cyanobacteria
PRESENTER: Bo Wang

Jamey Young: Development of untargeted metabolomics approaches to study bacterial-fungal co-cultures
PRESENTER: Baltazar Zuniga

Z

Karsten Zengler: Integration of Experimental and Modeling Approaches to Understand, Predict, and Modulate Rhizosphere Processes for Improved Bioenergy Crop Productivity
PRESENTER: Oriane Moyne

Philipp Zerbe: Improved Biofuel Production through Discovery and Engineering of Terpene Metabolism in Switchgrass

PRESENTER: Kira Tiedge

Huimin Zhao: Leveraging comparative population genomics to dissect the mechanisms of multi-stress tolerance of *I. orientalis*

PRESENTER: Yusuke Sasaki

Huimin Zhao: Metabolic regulation in *Issatchenkia orientalis* revealed by integrative omics

PRESENTER: Yihui Shen

Huimin Zhao: Model-driven Metabolic Engineering and 13C-Metabolic Flux Analysis for Non-model Yeast Organisms

PRESENTER: Patrick Suthers

Title: Establishment, Spread, and Impact of Non-Native Microbes in Non-Model Perennial Plants

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

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 community, and traits of the resident microbial community. We will investigate how genetics and ecology of non-model bacteria drive their establishment, persistence, and the productivity of biofeedstock perennial plants.

Bacillus species are ubiquitous bacteria in soils, sometimes forming beneficial mutualisms that promote plant growth (Xie, Zhang, and Pare, 2009). Establishment depends on gene-gene interactions between the bacteria and host plant. We will examine establishment success among *Bacillus* species and genotypes that are introduced to soils planted with either *Populus triochocarpa* (genotype 93-968), *P. deltoides* (ILL-101 and D124), a F1 hybrid cross (52-225), or a no-plant control. We will introduce multiple genotypes of three *Bacillus* species: *B. velezensis* (GB03, FZB42, BK100), *B. subtilis* (168, RO-NN-1), *B. amyloliquefaciens* (DSM7), and a mixture of all six genotypes mixed at equal proportions. This study will be conducted in the greenhouse in a fully factorial experiment: 8 *Bacillus* groups (6 genotypes, 1 mixture, 1 no-*Bacillus* control) x 5 plant groups (4 genotypes, 1 no-plant control) x 5 replicates, for a total of 200 pots. Establishment success will be measured using targeted qPCR on extracted DNA from soils. Plant growth parameters – photosynthetic rate, number of leaves, stem height, SPAD greenness index, foliar N content, and above and belowground biomass – will be measured to quantify the effect of microbial invaders on plants. We will quantify how microbial invaders affect the native soil microbiome by characterizing community composition via 16S and ITS2

amplicon sequencing and metaproteomics. We hypothesize that: (1) *B. velezensis* (GB03) will promote plant growth more than the other genotypes; (2) *Populus deltoides* (D124) will have the least successful microbial invasions because it is known to be resistant to other microbial invaders (e.g., *Septoria*); and (3) microbial invasions will be more successful in a diverse community than with individual genotype invasions. This controlled experiment will allow us to quantitatively disentangle how gene-gene interactions and genetic diversity drive microbial invader establishment and persistence.

In a concurrent field study, we will investigate the establishment and systemic spread of *Bacillus velezensis* (GB03). This genotype is applied as a commercial biofungicide in ornamental and fruit crops but has not yet been studied in a *Populus* system. It is applied to the soil and can spread from roots to leaves where it interacts antagonistically with fungal pathogens. We will study how the efficacy of *B. velezensis* varies among *Populus* species and genotypes (two *P. trichocarpa* genotypes, two *P. deltoides* genotypes, and two F1 hybrid genotypes). Invasion ecology theory suggests that propagule pressure increases establishment, and we will introduce *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). We will track establishment and spread of *B. velezensis* over time by periodically sampling soil, root, and leaf tissues and quantify abundance through targeted qPCR. We hypothesize 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 field study will be a fully factorial cross of 6 tree genotypes x 4 propagule pressure levels x 3 destructive harvest points (720 plants total). Harvests will continue non-destructively once plants are 6 months post-planting. This is the first temporal study of biofungicide systemic spread in a biofeedstock plantation and has applications for national energy security.

References:

Xie, X., Zhang, H., and Pare, P. 2009. Sustained growth promotion in *Arabidopsis* with long-term exposure to the beneficial soil bacterium *Bacillus subtilis* (GB03). Plant Signalling and Behavior 4(10):948-953.

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: Understanding the establishment of *Sphaerulina musiva* and development of genetic approaches to enhance poplar disease control

Authors: Joanna Tannous[1*] (tannousj@ornl.gov), Cole Sawyer,[1,2] Alyssa Carrell,[1] Miranda Clark,[1] Jessy Labbé,[1,2] and **Paul E. 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, USA

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.

The fungus *Sphaerulina musiva* has been reported to be both an agent of disease and a mutualistic partner of Poplar trees. While *S. musiva* is known to be the causal agent of one of the most detrimental diseases affecting susceptible hybrid poplar (e.g. *P. trichocarpa*) in North America, there have evidence that it can act as a passive endophyte when interacting with *P. deltoides*. This vast diversity of interactions is mainly related to the host genetic variation and susceptibility.

Breeding and cultivation of resistant plant species have been the main approaches used so far to control the damage caused by this pathogen on susceptible hosts. The limited amount of information about genetic functional molecular markers involved in *S. musiva*'s establishment and pathogenicity combined with the lack of optimized genetic tools have greatly hampered the development of disease mitigation strategies at the pathogen level. Therefore, we aim to quantify the genetic and molecular determinants that influence the establishment of *Sphaerulina musiva* within native *Populus* and *Salix* microbial communities.

For our first experiment, we have selected two *S. musiva* strains that represent the extremes of the virulence spectrum for the host *Populus trichocarpa*, with a long-term goal of extending and validating this work on the large collection of North America native population of 122 *S. musiva* isolates. Transcriptomics and proteomic data will be leveraged to identify functional markers that associate with successful establishment to interrogate the establishment and persistence of *S. musiva* in *Populus* and its associated microbiome. The collected omic datasets will be used to design and implement CRISPR RNA-guided gene drive systems to eradicate host-recognition mechanisms between *Sphaerulina* and *Populus* or *Salix* and attenuate its pathogenicity. Our ultimate objective of the CRISPR/Cas9 gene drive is to suppress leaf blight and stem canker

disease caused by *S. musiva* which has adverse outcomes on the production of the ecologically and commercially significant DOE biofeedstock *Populus*.

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: Bacterial quantitative trait-loci (QTL) mapping – a novel method for identification of genetic determinants affecting establishment of allopatric bacteria

Authors: Delyana P. Vasileva[1*] (vasilevadp@ornl.gov), Leah H. Burdick,[1] Jared C. Streich,[1] Dawn M. Klingeman,[1] Hari B. Chhetri,[1] Daniel A. Jacobson,[1] Joshua K. Michener,[1] and **Paul E. Abraham**[1]

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

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 tissues are intimately associated with hundreds of thousands of microbes. Members of these microbial communities can profoundly affect plant health and microbiome manipulation has the potential to improve the productivity of biomass plantations. However, microbial inoculants often perform inconsistently in field settings because the newly introduced bacteria fail to survive and proliferate in the presence of the plant and associated microbiome. These failures are attributed to poor understanding of the complex genetic networks and environmental factors involved in microbial establishment. Strains of *Bacillus velezensis* provide a good model system to study this process because they stimulate growth of a broad array of plant species and are currently available as commercial root and foliar inoculants.

To study the genetic architecture of *B. velezensis* establishment in plant ecosystems, we are using a novel genetic method, bacterial QTL mapping. QTL mapping techniques, which rely on crosses between distinct strains or lines, have a long history as powerful tools to untangle the genetic basis of complex phenotypes in eukaryotes. However, the lack of native systems for sexual recombination has prevented the use of this method for similar analyses in bacteria. Our approach uses genome shuffling by protoplast fusion to remove the barriers to recombination on a genome scale between genetically diverse bacterial cells.

As a proof-of-concept, we demonstrated the capacity of this method to produce QTL populations in *B. subtilis* where protoplast generation has been well characterized. Using *B. subtilis* strains with different genetic distances as parents, we showed that a single round of shuffling produces multiple random recombination events across the genome. Recursive protoplast fusion generated further genetic diversity. The resulting libraries of shuffled strains can be sequenced and phenotyped in a high-throughput manner to map causal genetic variants. We are currently expanding this approach to *B. velezensis* strains to identify genetic variants that correlate with successful establishment in plant-associated communities. This work will allow predictive control of microbial establishment, engineering of plant-microbe systems with optimized performance, and risk mitigation in the use of microbial inoculants.

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.

Real-time molecular detection systems for CRISPR-Cas genome engineering tools

Guoliang Yuan (yuang@ornl.gov), Md. Mahmudul Hassan, Tao Yao, Haiwei Lu, Michael Melesse Vergara, Jessy L. Labbé, Wellington Muchero, Jin-Gui Chen, Xiaohan Yang (yangx@ornl.gov), and **Paul E. Abraham**

Biosciences Division, Oak Ridge National Laboratory

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: CRISPR/Cas-based genome-editing has recently emerged as a powerful technology for genetic modifications in various organisms for a wide range of applications. Different CRISPR-Cas genome engineering tools, such as Cas9 nucleases, base editors, and prime editors, have been developed and widely used in 'loss-of-function' and 'gain-of-function' studies. Meanwhile, there are concerns about the risks associated with potential unwanted and unintended DNA changes that might accidentally arise from CRISPR-mediated gene editing. Potential off-target effects are currently unavoidable and very difficult to detect, requiring whole-genome sequencing for thorough evaluation, which is time consuming and costly. Therefore, it is critical to develop biosensor technologies for detecting the activities of CRISPR-Cas gene editing systems. Here, we aim to develop real-time detection systems that can indicate the presence of functional gene-editing systems for gene knockout, base editing, and prime editing in plants. To achieve this goal, we have created fluorescence-based molecular biosensors that are able to effectively detect (through fluorescence microscope) knockout, base editing and prime editing using protoplast transient expression and *Agrobacterium*-mediated leaf infiltration. Also, we are working on incorporating two visible reporters for real-time, noninvasive detection of Cas9 nucleases, base editors, and prime editors in plants without need for tedious analysis under fluorescence microscope. One of the reporters can indicate genome engineering under UV light. The other reporter can be visualized by naked eyes without the need of using special equipment or chemical treatments. These biosensors with visible reporters will not only simplify the detection procedures, but also be directly integrated with high-throughput plant phenotyping facility to link genes to traits.

Systems Biology-Based Optimization of Extremely Thermophilic Lignocellulose Conversion to Bioproducts

James Crosby* (jrcrosby@ncsu.edu),[12] Tania Tanwee,[2] Ryan G. Bing,[1] Tunyaboon Laemthong,[1] Gabriel Rubenstein,[2] Gina Lipscomb,[2] Ke Zhang,[3] Ying Zhang,[3] Dmitry Rodionov,[4] Robert M. Kelly,[1] and **Michael W. W. Adams[2]**

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

Project Goals: We are using systems biology-guided approaches to develop a non-model, microbial metabolic engineering platform based on the most thermophilic lignocellulose-degrading organism known, *Caldicellulosiruptor bescii* (T_{opt} 78°C). This work leverages recent breakthrough improvements in the molecular genetic tools for *C. bescii*, complemented by a comprehensive understanding of its metabolism and physiology gained over the past decade of study in the PIs' laboratories. We are applying the latest metabolic reconstruction and modeling approaches to optimize biomass to product conversion using switchgrass as a model plant, and acetone and other industrial chemicals as targets. The over-arching goal is to demonstrate that a non-model microorganism, specifically an extreme thermophile, can be a strategic metabolic engineering platform for industrial biotechnology using a systems biology-based approach.

Bioprocessing above 70°C has important advantages over near-ambient operations. These include resistance to contaminating organisms or phages, lower utility costs by using low-grade steam for heating and non-refrigerated water for cooling and reduced operating costs for maintenance of reactor and facility sterility. Additionally, generation of volatile products can reduce downstream separation costs, which typically account for a significant portion of operating costs. To develop *C. bescii* into a bioprocessing platform for conversion of unpretreated lignocellulose into industrially relevant chemicals, a comprehensive metabolic and regulatory reconstruction is necessary to provide a detailed description of this bacterium's physiology and metabolism and, more importantly, inform metabolic engineering strategies (1). This project leverages recent developments in *C. bescii* biomass deconstruction, genetic tools and strains and characterization of an alternate glycolytic pathway to experimentally validate ongoing modeling efforts (2,4-7). Recent metabolic engineering efforts in *C. bescii* demonstrated the ability to upgrade organic acids to alcohols through an aldehyde ferredoxin oxidoreductase and alcohol dehydrogenase (AOR-adhA pathway) (3) and the production of acetone from cellulose (8). Additional *C. bescii* strains are currently being constructed to study energy carrier usage and carbon flow and produce industrially relevant chemicals. High temperature continuous cultures and transcriptomic analyses are being employed to determine bioenergetic parameters and gene regulation patterns for growth of *C. bescii* strains. Additionally, fermentation profiling will be performed with lignocellulose-relevant sugars, including glucose, xylose, and cellobiose. These data will be used to refine the metabolic reconstruction and modeling analyses to inform metabolic engineering strategies that improve carbon and energy flow toward chemical production pathways, with the goal of demonstrating bioreactor scale industrial chemical production.

References

1. Crosby, J. R., Laemthong, T., Lewis, A. M., Straub, C. T., Adams, M. W. W. and Kelly, R. M. (2019) Extreme thermophiles as emerging metabolic engineering platforms. *Curr. Opin. Biotechnol.* **59,** 55-64
2. 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. Nature Comm.* **10,** 3548)
3. Rubinstein, G. M., Lipscomb, G. L., Williams-Rhaesa, A. M., 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. Indust. Micro. Biotech.* **47,** 585-597
4. Straub, C. T., Bing, R. G,, Wang, J. P., Chiang, V. L., Adams, M. W. W. and Kelly, R. M. (2020) Use of the lignocellulose-degrading bacterium *Caldicellulosiruptor bescii* to assess recalcitrance and conversion of wild-type and transgenic poplar. *Biotechnol. Biofuels* **13,** 43
5. 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
6. Zhang, K., Zhao, W., Rodionov, D., 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. (2020) "Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for ethanol production" *mSystems* (submitted 12/2020)
7. Rodionov, D. A., Rodionova, I. A., Rodionov, V. A., Zhang, K., Rubinstein, G. M., Tanwee, T. N. N., Crosby, J., Bing, R. G., Nookaew, I., Basen, M., Brown, S. D., Wilson, C., Klingeman, D. M., Poole, F. L., Zhang, Y., Kelly, R. M. and Adams, M. W. W. (2020) "Transcriptional regulation of biomass degradation and carbohydrate utilization in *C. bescii*" *mSystems* (submitted 12/2020)
8. 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. Bioengineer*. **117**: 3799– 3808.

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

Genome-Scale Metabolic and Regulatory Network Reconstruction of *Caldicellulosiruptor bescii* Leads to the Reconstruction of Predictive Models for Bioproduct Generation

Dmitry A. Rodionov,[1]*(rodionov@sbpdiscovery.org), Ying Zhang, [2](yingzhang@uri.edu), Weishu Zhao, [2] Ke Zhang, [2] James R. Crosby,[3] Ryan G. Bing,[3] Diep M. N. Nguyen,[4] Tania N. N. Tanwee,[4] Gabriel M. Rubinstein,[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: We are using systems biology-guided approaches to develop a microbial metabolic engineering platform for the non-model organism, *Caldicellulosiruptor bescii*, the most thermophilic lignocellulose-degrading organism known with an optimal growth temperature near 80°C. This work leverages recent breakthrough advances in the development of molecular genetic tools for this organism, complemented by a deep understanding of its metabolism and physiology gained over the past decade of study in the PIs' laboratories. We are applying the latest metabolic reconstruction and modeling approaches to optimize biomass to product conversion using switchgrass as the model plant and acetone and other fermentation products as targets. The overarching goal is to demonstrate that a non-model microorganism, specifically an extreme thermophile, can be a strategic metabolic engineering platform for industrial biotechnology using a systems biology-based approach.

Caldicellulosiruptor bescii is an extremely thermophilic, strictly anaerobic, gram-positive bacterium. It is the most thermophilic cellulolytic bacterium known to date (T_{opt} 78°C, T_{max} 90°C), and it can use a wide range of simple and complex carbohydrates. Its ability to degrade plant biomass without enzymatic or chemical pretreatments and at a high optimum growth temperature offers several advantages for industrial applications (1). Engineered *C. bescii* strains have been shown to produce desired bioproducts, such as ethanol, from un-pretreated plant biomass through consolidated bioprocessing (CBP). However, efficient metabolic engineering requires in-depth understanding of its metabolic and transcriptional regulatory networks. Previous experimental studies identified a variety of carbohydrate-active enzymes in *C. bescii* and related *C. saccharolyticus* species, while prior transcriptomic experiments identified their putative carbohydrate uptake transporters.

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 *C. bescii*. The reconstruction of carbohydrate utilization regulatory network includes the predicted binding sites for 34 mostly local transcription factors and points to the regulatory mechanisms controlling expression of genes involved in degradation of plant biomass (1). The Rex and CggR regulons control the central glycolytic and primary redox reactions. The identified transcription factor binding sites and regulons were validated with transcriptomic and transcription start site experimental data for *C. bescii* grown on cellulose, cellobiose, glucose, xylan, and xylose. The XylR and XynR regulons control xylan-induced transcriptional response of genes involved in degradation of xylan and xylose utilization. The reconstructed regulons informed the carbohydrate utilization reconstruction analysis and improved

functional annotations of 51 transporters and 11 catabolic enzymes. Using gene deletion, we confirmed that the shared ATPase component MsmK is essential for growth on oligo- and polysaccharides, but not for the utilization of monosaccharides (2).

The reconstruction of regulatory networks of carbohydrate utilization was complemented with a genome-scale model (GEM) of the *C. bescii* metabolism (3). The model was used to examine potential bottlenecks that could be encountered for metabolic engineering of *C. bescii* to produce bio-based chemicals from plant biomass. The model utilizes subsystems-based genome annotation, targeted reconstruction of carbohydrate utilization pathways, and biochemical and physiological based experimental validations. Specifically, carbohydrate transport and utilization pathways involving 160 genes and their corresponding functions were incorporated, representing the utilization of C5/C6 monosaccharides, disaccharides, and polysaccharides such as cellulose and xylan. The model predicted that optimal production from biomass-based sugars of the model product, ethanol, was driven by ATP production, redox balancing, and proton translocation, mediated through the interplay of an ATP synthase, a membrane-bound hydrogenase, a bifurcating hydrogenase and a bifurcating NAD- and NADP-dependent oxidoreductase. These mechanistic insights guided the design and optimization of new engineering strategies for product optimization, which were tested in the *C. bescii* model, showing a two-fold increase in ethanol yields (3).

References

1. Crosby, J. R., Laemthong, T., Lewis, A. M., Straub, C. T., Adams, M. W. W. and Kelly, R. M. (2019) Extreme thermophiles as emerging metabolic engineering platforms. *Curr. Opin. Biotechnol.* **59,** 55-64
2. Rodionov, D. A., Rodionova, I. A., Rodionov, V. A., Zhang, K., Rubinstein, G. M., Tanwee, T. N. N., Crosby, J., Bing, R. G., Nookaew, I., Basen, M., Brown, S. D., Wilson, C., Klingeman, D. M., Poole, F. L., Zhang, Y., Kelly, R. M. and Adams, M. W. W. (2020) "Transcriptional regulation of biomass degradation and carbohydrate utilization in *C. bescii*" *mSystems* (submitted 12/2020)
3. Zhang, K., Zhao, W., Rodionov, D., 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. (2020) "Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for ethanol production" *mSystems* (submitted 12/2020)

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

From natural carbon to a defined medium: Characterization of carbon utilization of FRC isolates

Markus de Raad[1*] (MdeRaad@lbl.gov), Xiaoqin Wu[1*] (xiaoqinwu@lbl.gov), Yifan Li[1], Suzanne M. Kosina[1], Benjamin P. Bowen[1], A.M. Deutschbauer[1], R. Chakraborty[1], Trent R. Northen[1], Adam P. Arkin[1,2], and **Paul D. Adams**[1,2]

[1]Lawrence Berkeley National Lab, Berkeley CA; [2]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 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.

Natural organic matter (NOM) plays important roles in biological, chemical, and physical processes within the terrestrial ecosystem. The water-soluble fraction of NOM, known as dissolved organic matter (DOM), is associated with high microbial activity and respiration. Improving our knowledge of DOM composition is therefore critical to understand the native substrates used to support soil and sediment microbial communities.

Previously, we analyzed DOM of a saprolite soil collected from the Oak Ridge Field Research Center (ORFRC) using untargeted characterization of water-soluble soil microbial metabolites [1]. With this approach, 96 metabolites were identified, including amino acids, sugars, mono- and di-carboxylic acids, nucleobases, and nucleosides. This pool of metabolites was used to build a soil defined medium (SDM). Although SDM is environmentally based, it only supported the growth of only half of the tested bacterial from the ORFRC field site [1] which we now attribute to elemental stoichiometry especially the low carbon to nitrogen ratio which was 1. Here, we describe the construction of a new defined medium, NLDM, which additionally accounts for soil elemental stoichiometry, soil microbe metabolite use, and the composition of the widely-used R2A rich medium. We found that NLDM supported the growth of 53 of 53 phylogenetically diverse isolates we tested and enabled rapid profiling of substrate use via LC-MS/MS. We anticipate that this media may have additional use in isolating soil and sediment microorganisms.

Using ultrahigh resolution mass spectrometry, we were able to characterize ORFRC sediment DOM and its transformation by indigenous microbes. Lignin-like carbon, is the major component of DOM along the entire depth of sediment core. Labile carbon such as carbohydrate and amino sugar-like compounds are present in DOM from shallow surficial sediment. Decrease in abundance in deeper sediment, which happens to be more abundant in mainly recalcitrant carbon such as lipid and condensed aromatics. The dissolved organic nitrogen (or protein-like compounds) fraction of sediment DOM is an alternative carbon source for microbes inhabiting ORFRC groundwater when labile carbon is limited [2]. With sediment DOM as sole carbon source we successfully enriched rarely cultured phyla including *Verrucomicrobia*, *Planctomycetes*, and *Armatimonadetes*, and obtained isolates of several novel and undescribed organisms [3].

With our novel understanding of DOM and their transformations by FRC isolates using NLDM are now being used to link our lab studies to field observations. This will result in a better understanding of the coupling between growth substrates, other environmental controls, and microbial community activity and structure.

References

1. References are not required but should be placed here if needed. Jenkins, S. *et al.* Construction of Viable Soil Defined Media Using Quantitative Metabolomics Analysis of Soil Metabolites. *Front. Microbiol.* **8**, 2618 (2017).
2. Xiaoqin Wu, Liyou Wu, Yina Liu, Ping Zhang, Qinghao Li, Jizhong Zhou, Nancy J. Hess, Terry C. Hazen, Wanli Yang, Romy Chakraborty. Microbial interactions with dissolved organic matter drive carbon dynamics and community succession. *Front Microbiol*. **2018**, 9, 1234.
3. Xiaoqin Wu, Sarah Spencer, Sara Gushgari-Doyle, Mon Oo Yee, Jana Voriskova, Yifan Li, Eric J. Alm, Romy Chakraborty. Culturing of "Unculturable" Subsurface Microbes: Natural Organic Carbon Source Fuels the Growth of Diverse and Distinct Bacteria from Groundwater. *Front Microbiol*. **2020**, 11, 610001.

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

Targeted Isolation Using Field-Informed Approaches

S. Gushgari-Doyle[1*] (gushgari@lbl.gov), M.O. Yee[1], J.V. Kuehl[1], H.J. Smith[5], M.P. Thorgersen[8], X. Ge[8], A.E. Otwell[10], T.L. Lie[10], K.A. Hunt[10], M.W.W. Adams[8], E.J. Alm[3], N.S. Baliga[4], J.-M. Chandonia[1], A.M. Deutschbauer[1], D.A. Elias[7], M.W. Fields[5], T. C. Hazen[6,7], T.R. Northen[1], A. Mukhopadhyay[1], G.E. Siuzdak[9], D.A. Stahl[10], P.J. Walian[1], J. Zhou[11], R. Chakraborty[1], A.P. Arkin[1,2] and **P.D. Adams[1,2]**

[1]Lawrence Berkeley National Lab, Berkeley; [2]University of California at Berkeley; [3]Massachusetts Institute of Technology, Cambridge; [4]Institute for Systems Biology, Seattle; [5]Montana State University, Bozeman; [6]University of Tennessee, Knoxville; [7]Oak Ridge National Lab, Oak Ridge; [8]University of Georgia, Athens; [9]Scripps Research Institute, San Diego; [10]University of Washington, Seattle; [11]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.

Isolation of microorganisms representative of key environmental metabolisms is essential to developing a fundamental understanding of ecological processes., yet, the scientific community estimates that we are able to culture fewer than 2% of microorganisms on Earth in the laboratory.[1] To recover field-relevant isolates from Oak Ridge Reservation Field Research Center (ORR FRC), we combined several statistical analyses using our environmental and sequencing metadata to A) identify high-priority targets for isolation based on abundance, community correlation, and other metrics, B) select samples and enrichments with increased probability to yield those targets, and C) inform growth medium composition as well as enrichment and isolation approaches. We have employed these targeted isolation techniques (in addition to high-throughput, untargeted approaches- ENIGMA SFA poster by Kuehl et al.) to not only increase isolate diversity in the ENIGMA culture catalogue, but also to recover key isolates from the fieldsite that are suspected to play integral roles in carbon and nitrogen cycling in the terrestrial subsurface.

Here, we present several isolation and enrichment successes resulting from this targeted approach, including organisms exhibiting nitrate- and sulfate-reducing metabolisms, complex carbon-transforming metabolisms, high metal tolerance, and acidophiles. Several of our isolates are novel and rarely cultivated/ previously uncultivated clades. We also present ongoing efforts in enrichment and isolation of ammonia oxidizing- and nitrous oxide reducing- microorganisms (see ENIGMA SFA poster by Hunt et al.). In addition, we have several ongoing efforts enriching and isolating for metabolisms and characteristics representative of field observations at the ORR

FRC. The results of this work will facilitate development of novel, tractable genetic systems, community interaction studies, and environmental simulations to connect phenotype and genotype to field observations.

References

1. Steen, A. D. et al. High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J.* 13, 3126–3130 (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

Resolving Biotic and Abiotic Controls of Nitrous Oxide Flux in a Subsurface Site Contaminated with High Nitrate Concentrations.

K. A. Hunt[1*](hunt0362@uw.edu), A. E. Otwell[1], S. Bowman[2], S. D. Wankel[2], K. F. Walker[3], E. R. Dixon[3], M. Rodriguez[3], K. A. Lowe[3], D. C. Joyner[3], A. Carr[4], L. Lui[5], T. Nielsen[6], N. S. Baliga[4], T. C. Hazen[3], D. A. Stahl[1] (dastahl@uw.edu), A. P. Arkin[5,7], **P. D. Adams**[5]

[1]University of Washington, Seattle; [2]Woods Hole Oceanographic Institution, Woods Hole; [3]Oak Ridge National Lab, Oak Ridge; [4]Institute for Systems Biology, Seattle; [5]Lawrence Berkeley National Laboratory, Berkeley; [6]DOE Joint Genome Institute, Walnut Creek; [7]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.

Abstract:

Linking field observations with laboratory studies, and vice versa, is essential for advancing predictive understanding of environmental systems and for stewardship of those systems. We are developing tools and identifying systems that capture phenomena observed in the field in a reproducible, minimally disruptive, and dissectible manner (see poster by Smith *et. al.*). In turn, those observations serve to direct more refined laboratory studies designed to more completely resolve the roles of microbial ecology and abiotic processes in observable systems-level processes (see poster by Valenzuela *et. al.*).

Microbial activity in the field can be a challenge to quantify due to spatial and temporal heterogeneity and the cost/methodological constraints of real time observations without perturbation. Stable isotopic analysis of biogeochemically active substrates and products can circumvent some of these limitations. The field research center (FRC) at Oak Ridge, TN contains a site that has been contaminated with low pH (3-7), heavy metal laden nitrate (~10 g/l) for decades from historical activities. To understand the microbial processes of this site, we analyzed stable isotopes of ground water as well as dissolved nitrate and nitrous oxide collected from a total of 27 different wells over 3 different areas. These isotopic analyses are well suited for this investigation given the strong process signals (denitrification vs nitrification vs dilution) imprinted on δ15N and δ18O of different nitrogen species. Previous measurements of nitrous oxide flux indicated major subsurface production of nitrous oxide, which are supported here by δ15N, δ18O, and site preference of nitrous oxide samples analyzed from groundwater. Analyses revealed a variable distribution of activity across the site (horizontally and with depth) and implicated both denitrification and chemodenitrification in nitrous oxide production. These data are being contextualized by complementary molecular, biological, and chemical characterization of groundwater and sediment traps recovered from the 27-well survey, together contributing to a framework for developing a more predictive understanding of biotic and abiotic controls of local and systems-level processes.

High concentrations and fluxes of nitrous oxide measured in the subsurface without associated surface emissions indicated significant microbial activity driven by both the production and consumption of this high energy electron acceptor. Two populations of nitrous oxide reducers (clade I and II) were observed to be stratified with depth, a distribution suggested to be controlled by either nitrous oxide concentration

or response to inhibitory factors, including competition for oxygen. To understand these processes and populations, nitrous oxide reducers are being isolated and will provide field relevant kinetic and physiological parameters to evaluate the roles of the activities and variable distribution of clade I and II nitrous oxide reducers in controlling nitrous oxide emissions from the subsurface (see poster by Gushgari-Doyle *et. al.*).

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: Combining Multi-omics with Random Walks to Explore the Function of Candidate Genes

Authors: David Kainer[1]* (kainerd@ornl.gov), Izaak Miller,[1] Mikaela C. McDevitt,[1] Daniel Jacobson,[1] and **Paul E. Abraham[1]**

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

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 SEED project will inoculate a *Populus trichocarpa* (Black Cottonwood) genotype with over one hundred strains of the fungal pathogen *Sphaerulina musiva* (formerly Septoria) in order to measure a phenotype that describes the success of microbial invader establishment. A wide array of omics data will be assayed in the Septoria population, including genomics, transcriptomics and proteomics, which will allow us to explore the factors that influence an invader's ability to become established. GWAS (Genome Wide Association Studies) will be used to associate Septoria genomic variation with variation in establishment phenotypes. However, GWAS for complex traits often provides cryptic results. To unravel the complexity of establishment dynamics we will apply network analysis techniques coupled with Explainable AI (X-AI) algorithms so that all aspects of omics data are incorporated jointly with the GWAS results.

Each available omics layer, including publicly available data on *Sphaerulina musiva* and related species, will be converted to a network format that connects genes to genes based on functional and experimental relationships. Multiple networks will be combined into a multiplex network[1] for use with machine learning algorithms such as Random Walk with Restart (RWR)[2]. The RWR approach enables the exploration of the function of candidate genes using network-topology-association without being limited to their immediate neighborhood. The multiplex network jointly uses the gene-gene connectivity from every omics layer while maintaining the information encoded within the unique topology of each layer. Together this is a powerful multi-omic analysis framework.

References/Publications

1. Battiston, Federico, Vincenzo Nicosia, and Vito Latora. "Structural measures for multiplex networks." *Physical Review E* 89.3 (2014): 032804.

2. Alberto Valdeolivas, Laurent Tichit, Claire Navarro, Sophie Perrin, Gaëlle Odelin, Nicolas Levy, Pierre Cau, Elisabeth Remy, Anaïs Baudot, Random walk with restart on multiplex and heterogeneous biological networks, *Bioinformatics*, Volume 35, Issue 3, 01 February 2019, Pages 497–505,

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.

Cone Penetrometer 3-D Characterization of Y-12 Site to Determine the Hydrological, Geological and Biogeochemistry Best Sites for ENIGMA Subsurface Observatories

T.C. Hazen[1,2], **E.R. Kelly**[1]* (ekelly14@vols.utk.edu), A. Putt[1], K. Walker[1], D.C. Joyner[1], K. Lowe[2], M. Rodriguez Jr[2], M.W. Fields[3], R. Chakraborty[4], X. Wu[4], D. Stahl[5], T. Lie[5], M.W.W. Adams[6], F. Poole[6], P.J. Walian[4], J. Zhou[5], J. Van Nostrand[5], T.R. Northen[4], J.-M. Chandonia[4], A.P. Arkin[4,8] and **P.D. Adams**[4,8]

[1]University of Tennessee, Knoxville; [2]Oak Ridge National Lab, Oak Ridge**,** [3]Montana State University, Bozeman; [4]**Lawrence Berkeley National Lab, Berkeley**; [5]University of Washington, Seattle; [6]University of Georgia, Athens; [7]University of Oklahoma, Norman ; [8]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. The Cone Penetrometer field study overarching aim was to get a detailed view of lithology of the subsurface in Area 3 next to S-3 ponds disposal site to select the best sites for installing subsurface observatories for follow-on ENIGMA studies. Additional aims were to: 1. Create groundwater flow model of Area 3 that shows nitrate concentration, 2. Analyze impacts of groundwater rate and flow direction on geochemical parameters and nitrate concentration in Area 3, 3. Analyze impact of subsurface lithology on nitrate concentration in Area 3, and 4. Analyze impact of subsurface lithology on DOC concentration in Area 3.

Over 16 days, a 131-push cone penetrometer grid was completed across the 2,600 square meters of Area 3 with subsurface lithology mapped from 1m to 11m below ground surface (mean=5.9m). In addition, 34 sediment samples were collected near select pushes. Each 0.6m core was collected within 2 to 9.5m below ground surface and consisted primarily of clay. Cores were subdivided into 292 subsamples for: metals, C, N, DOC, DON, biomass, pH, nitrate, nitrite, isotopic fractionation, 16S/18S sequencing, Geochip, respiration, and metabolomics. Groundwater in wells in this area were also measured during the cone penetrometer activity for: DO, pH, temperature, conductivity, redox, salinity, nitrate, depth to water and vector for groundwater flow and rate with colloidal boroscope. The cone penetrometer survey provided a detailed view of the unconsolidated sediment layers of Area 3. This study was conducted to evaluate the interaction between groundwater, different sediment types, and biogeochemistry in order to identify the locations of the future subsurface observatories for ENIGMA. Sediment types, cone pressures, and geochemical data can be mapped to find the extent of subsurface sediment layers with the intention of identifying pathways of flow and recharge. Pushes were

driven to refusal by stiff fine grained material or rock. In the northern section of Area 3, refusal was reached around 4m and the layers most likely contributing to flow are those high in sand and gravel with detectable radiation measured on an in-field Geiger counter. Moving southward, the distance to refusal increased linearly with some pushes extending to 12m indicating a southward slope in the direction of a low-order surface stream. The well-defined gravel layers of the north appear to be mixed in the southern and central section's where mixed sand layers are abundant. The sands include gravely sands, silty sands, and clayey sands with shifting sedimentary composition indicating heterogeneity within the layers. Initial models of the lithology suggest these layers are connected although the central section had a reduced number of pushes due to existing infrastructure. The cone penetrometer study also revealed a large number of localized discontinuous clay and silty clay lenses with limited horizonal and vertical extent. This high-resolution study of the sediment types will benefit our future investigations and current understanding of the transport, storage, and fate of both organic and inorganic substrates in Area 3.Water levels, colloidal borescope vectors, and geochemical measurements were gathered in combination with the stratigraphy data to investigate the possible locations for future multi-level subsurface observations. Because of the large number of wells (102) in Area 3 it allowed measurements before, during and after cone penetrometer pushes near these wells. This demonstrated that the cone penetrometer pushes had no effect on adjacent well water level, vector for flow direction and rate of flow or for DO, pH, temperature, conductivity, redox, salinity, and nitrate. To our knowledge this is the first time this has been demonstrated for a cone penetrometer survey.

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

High Throughput Isolation and Microbial Community Enrichment Platforms

J.V. Kuehl[1](JVKuehl@lbl.gov), V.V. Trotter[1], H.K. Carlson[1,] R. Chakraborty[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.

Generating models that can accurately predict how environmental perturbations will affect the microbial communities and their natural ecosystems will require a representative set of characterized microbial strains found in the field. To that end, we have developed microbial isolation strategies to quickly generate and catalogue thousands of isolates using a variety of standard growth media and a scalable characterization pipeline including high quality genome sequences, growth data over a standard set of conditions, and accompanying exometabolomics. To date, this approach has generated > 1000 isolates, from commonly isolated genera (i.e. *Pseudomonas, Cupriavidus*), to isolates assigned to a genus that is new to our ENIGMA strain collection (i.e. *Zoogloea, Xylophilus, Cellovibrio*). The commonly cultured genera, which in some cases are abundant at the site, can be used to pinpoint the genomic signatures that account for differences in their growth profiles, while the rare members provide a valuable strain resource to discover novel genes and pathways. In parallel, we are also using microbial enrichments, that expand the spectrum of growth conditions tested and microbial membership, to increase the culturable microbial diversity in the lab. This enrichment dataset provides a wealth of information about the microbial growth parameters ideal for capturing microbes important in the field, while at the same time generating the microbial community materials that can be used for generating isolates using the high throughput pipeline. In addition, the microbial enrichments-when done at the scale we are doing here- can provide insight into microbial interactions, the conditions that drive those interactions, and the raw materials to start dissecting those interactions. All of this work, complements the targeted approach of other ENIGMA researchers and provides them the bandwidth to go after the very difficult and slow growing microbes that are important at the field site. With the successful development of these platforms, we have begun the rountine processing of isolates and enrichments adding to the shared ENIGMA resource of a diverse and tractable microbial collection.

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

Using high-throughput technologies to understand mechanisms of predation

V.K. Mutalik[1] (vkmutalik@lbl.gov), S. Carim[2], A. Ramirez[1], J. Kuehl[1], A.L. Azadeh[2], A.E. Kazakov[1], M. N. Price[1], P.J. Walian[1], LM. Lui[1], T. Nielsen[1], R. Chakraborty[1], A.M. Deutschbauer[1], A.P. Arkin[1],[2] and **P.D Adams[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]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.

Abstract:

Microbial communities are highly dynamic impacted by diverse biotic and abiotic factors in an ecosystem. One of the key biotic factors that impact the community dynamics is the antagonistic behavior by the members of the community. Like bacterial viruses/phages, bacterial predators that kill and feed on other nearby prey bacteria account for large percentage of bacterial mortality and are widespread in diverse ecosystems. Similarly, phage tail-like bacteriocins (PTLBs) are highly potent, variable spectrum bactericidal agents known to impact microbial community dynamics. Though predator-prey interactions have been studied over deacdes, our molecular understanding of how prey-recognition and specificity of interaction works is still unclear[1]. In particular, there is a scarcity of such information on bacterial predators (phages and bacterial predators) and bactericidal agents in heavy metal contaminated soil and water bodies such as Oak Ridge FRC field site. For example, the metagenomic sequence data from Oak Ridge FRC site samples shows the presence of bacterial predators such as *Bdellovibrio* and *Bacteriovorax* species, however their role in structuring the community observed is less clear. The discovery of predator-prey interaction specificity determinants and predator resistance mechanisms would open new avenues for the dissection of ecosystem function.

We identified PTLB biosynthetic gene clusters in eleven of a set of twelve closely related Oak Ridge *Pseudomonas* isolates[2]. We focus on five killing interactions, and two resistant interactions, between PTLBs purified from producing strains, and genome-wide random bar code transposon-site sequencing (RB-TnSeq) mutant libraries[3] of target strains. PTLB production was confirmed and characterized by transmission electron microscopy and proteomics analyses. To survey host factors involved in PTLB mediated target killing, we performed pooled fitness assays with the RB-TnSeq libraries, using PTLBs as stressors. Initial analysis of our genome-wide fitness data suggests that specific lipopolysaccharide residues–likely serving as binding receptors–are involved in killing by different PTLBs. We aim to determine the atomic structure of the tailocin tail fibers in complex with the LPS in order to understand the recognition mechanism used by Tailocins to bind to and kill bacteria.

We have successfully enriched bacterial predator *Bdellovibrio* from the serially enriched water sample from Oak Ridge FRC field site. We are currently refining our methods to isolate clonal isolate of predators that grow on different target hosts. Our next steps will include characterization of their host-range and prey cognition determinants using genome-wide libraries of target isolates from the same environment.

References

1. Mutalik, et al., *PLoS Biology* 2021, 18(10): e3000877
2. Carim, et al., *BioRxiv* 2020, doi: https://doi.org/10.1101/2020.05.27.113977
3. Wetmore, et al., *mBio* 2015, 10.1128/mBio.00306-15

Funding Statement: *This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), 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*

Physical size matters in groundwater bacterial community assembly

D. Ning[1](ningdaliang@ou.edu), Y. Fan[1], L. M. Lui[2], J. P. Michael[1], Y. Fu[1], J. D. Van Nostrand[1], R. Tian[1], Y. Wang[1], K. F. Walker[3,4], E. R. Dixon[3,5], A. D. Putt[3,5], D.E. Williams[3,5], D. C. Joyner[3,5], K. A. Lowe[3], F. L. Poole[6], X. Ge[6], M. P. Thorgersen[6], M.W.W. Adams[6], R. Chakraborty[2], X. Wu[2], D. A. Elias[3], R. L. Wilpiszeski[3], J. Zhou[1,2], M.W. Fields[7], T. C. Hazen[3,4,5,2], A.P. Arkin[2,8] and **P.D. Adams**[2,8]

[1]University of Oklahoma, Norman; [2]Lawrence Berkeley National Laboratory, Berkeley; [3]Oak Ridge National Laboratory, Oak Ridge; [4]Bredesen Center of Interdisciplinary Research and Education, Knoxville; [5]University of Tennessee, Knoxville; [6]University of Georgia, Athens; [7]Montana State University, Bozeman; [8]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.

A central issue in microbial ecology is to identify key drivers shaping community assembly and functioning, which is also fundamental for ecological modeling. Various physical-chemical and microbial drivers have been extensively investigated, including climate (e.g. temperature), chemistry (e.g. pH, carbon sources), microbial functional genes, etc. However, a basic trait, physical size of cells or the particles they attached, was rarely studied. Considering the porous habitats in subsurface, physical size should be particularly important in groundwater microbial community assembly and functioning. Small-size microbes (e.g. ultramicrobacteria) can be easier to pass through soil pores along groundwater flow; but when flow rate is limited, they may be more difficult to escape from absorption, trapping, or flocculation. In addition, small size generally leads to higher specific surface area and slower growth rate (longer division period) of bacteria. Thus, we hypothesize that physical size plays an essential role in the variation of microbial diversity, assembly, and functioning.

The groundwater in the Oak Ridge Integrated Field Research Challenge site (FRC, Oak Ridge, TN) has large geochemical gradients and diverse subsurface conditions for microbial dispersal. In the spring of 2019, three areas under different contamination levels, three groundwater wells in each area, were selected for 9-week bi-weekly sampling from March to May. Each groundwater sample was filtered in succession through 10-μm filter mainly for large-size and particle-attached bacteria (so-called ‘large’), 0.2-μm mainly for normal-size free-living bacteria (so-called ‘medium’), and 0.1-μm for small free-living bacteria (so-called ‘small’). The results of bacterial communities from 16S rRNA gene sequencing well supported our hypothesis.

The large, medium, and small-size bacteria counted for 7%, 62%, and 31% of the DNA concentrations on average, and the small-size portion was higher than 20% (up to 32%) in some well in May, indicating these three types were all not negligible. **The size significantly affected bacterial diversity and community structure.** Alpha diversity of small-size bacteria (Shannon 3.8±1.2) was generally lower than medium (5.4±1.2) and large-size bacteria (5.1±1.1). The size difference explained a substantial proportion (24.4%, $P<0.001$; higher than location and time) in the variation of bacterial alpha diversity (Shannon index), and proportion (8.4%, $P<0.001$; similar to location and time) of beta diversity (Bray-Curtis index). The location, which reflects the influence of contamination and/or spatial distance, showed obvious effect on the alpha (21% and 29%) and beta diversity (26% and 29%) of large and medium-size bacteria, but had much lower impact on small-size bacteria (14% of alpha and 10% of beta). While large and medium-size bacteria from different areas always showed different community structure, small-size bacteria from different areas became convergent after late Apr. The results might be related to better dispersal of small-size bacteria and increased precipitation in late April and May. **The size also affected the dominant phylogeny and some key functional species.** While alpha- and/or beta-Proteobacteria generally dominated in large and medium-size bacteria, the phylum Bacteroidetes significantly increased or even predominate in small-size bacteria after late April, mainly attributed to the genus *Hydrotalea*. Sulfate-reducing bacteria, a relevant functional group in this site, were mainly detected in the orders *Syntrophobacterales* and *Desulfobacterales* and Class *Thermodesulfovibrio* in both large and medium-size bacteria, but were nearly undetectable in small-size bacteria. A newly developed framework based on phylogenetic-bin-level null model analysis (iCAMP) was applied to explore **the different assembly mechanisms of bacteria with different sizes**. Based on the results, dispersal limitation generally played more important roles in large-size (46±5%) than in medium (38±6%) and small-size bacteria (23±9%). In contrast, selection was obviously more influential in small (31±12%) and medium (23±5%) than in large-size bacteria (18±2%). The major assembly processes, selection, dispersal limitation, and drift, all showed dramatically higher (1.5-4.0 folders) temporal variations in small (CV 0.39-0.45) than in medium (CV 0.17-0.29) and large-size bacteria (CV 0.10-0.19).

In conclusion, the size of cells or attached particles is essential in shaping groundwater microbiome. Bacteria with larger sizes or particle-attached bacteria were more affected by dispersal limitation and showed more variation among different locations but higher resistance in temporal dynamics; in contract, smaller free-living bacteria are more sensitive to temporal change of environmental conditions, and can be easier to migrate under adequate flow rate.

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), 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

Mechanism across scales: integrating laboratory and field studies for microbial ecology as illustrated by the ENIGMA SFA

L. M. Lui[1,*], E. L-W. Majumder[2,*], H.J. Smith[3,*] (hjsmith12@gmail.com), H. K. Carlson[1], F. v. Netzer[4], T. N. Nielsen[1], M. Peng[5], X. Tao[5], A. Zhou[5], M. Price[1], J. V. Kuehl[1], A.J. Hendrickson[1], V. Trotter[1], S. Gushgari-Doyle[1], J. Valenzuela[6], A. Otwell[4], K. Hunt[4], A. Carr, K. Walker[7,8], E. Dixon[7,8], F. Poole[9], M. Thorgersen[9], X. Ge[9], M.W.W. Adams[9], E.J. Alm[10], N.S. Baliga[6], J.-M. Chandonia[1], A.M. Deutschbauer[1], D.A. Elias[7], M.W. Fields[3], T. C. Hazen[7,8], T.R. Northen[1], A. Mukhopadhyay[1], G.E. Siuzdak[11], D.A. Stahl[4], P.J. Walian[1], J. Zhou[11], R. Chakraborty[1], A.P. Arkin[1,12] and **P.D. Adams[1]**

*Contributed equally

[1]Lawrence Berkeley National Lab, Berkeley; [2]University of Wisconsin, Madison; [3]Montana State University, Bozeman; [4]University of Washington, Seattle; [5]University of Oklahoma, Norman; [6]Institute for Systems Biology, Seattle; [7]Oak Ridge National Lab, Oak Ridge; [8]University of Tennessee, Knoxville; [9]University of Georgia, Athens; [10]Massachusetts Institute of Technology, Cambridge; [11]Scripps Research Institute, San Diego; [12] University of California at Berkeley, 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

Over the last century, leaps in technology for imaging, sampling, detection, high-throughput sequencing and 'omics analyses have revolutionized microbial ecology to enable rapid acquisition of extensive datasets for microbial communities across ever-increasing temporal and spatial scales. The present challenge is capitalizing on our enhanced abilities of observation and integrating diverse data types from different scales, resolutions and disciplines to reach a causal and mechanistic understanding of how microbial communities transform and respond to perturbations in the environment. This type of causal and mechanistic understanding will make predictions of microbial community behavior more robust and actionable in addressing microbially-mediated global problems. To discern drivers of microbial community assembly and function, we recognize the need for coordinated, model-driven experiments that integrate the analysis of genomics data, biogeochemical parameters, and ecological and physical forces to rates of microbial growth at specific locations.

To link processes and factors from the gene scale to the ecosystem scale for subsurface microbiology, ENIGMA, a U.S. Department of Energy Science Focus Area, seeks to understand the biogeochemical and microbial processes in the Oak Ridge Reservation (ORR). To accomplish this, we are using a coordinated inter-laboratory framework to link processes and factors from the gene scale to the ecosystem scale. ENIGMA coordinates multiple studies at the

field scale, mesocosm scale, and molecular/species level and has major research thrusts aimed at field surveys, laboratory and bioreactor studies of isolates, syncoms, enrichments, improved isolation methods, genetic tool development, and bioinformatics analyses and tools. We describe how these ENIGMA efforts are being utilized to characterize and build a predictive understanding of the microbial subsurface communities of ORR and how we are generalizing this integrated approach to be applicable to other study systems and environments.

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

Technologies for High-Throughput characterization of environmental isolates

V.V. Trotter[1*] (vvtrotter@lbl.gov), J.V. Kuehl[1], M. De Raad[1], M.E. Garber[1], M.P. Thorgersen[3], E.L. Majumder[4], J. Xue[4], M.N. Price[1], M.W.W. Adams[3], A.M. Deutschbauer[1,2], T.R. Northen[1], A. Mukhopadhyay[1], G.E. Siuzdak[9], A.P. Arkin[1,2] and **P.D. Adams[1,2]**

[1]Lawrence Berkeley National Lab, Berkeley; [2]University of California at Berkeley; [3]University of Georgia, Athens; [4]Scripps Research Institute, San Diego;

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.

Ongoing large scale high-throughput cultivation and targeted lower throughput approaches aim to recover the broad diversity of microbes from the Oak Ridge site that are exemplars of relevant metabolism occurring in this specific environment. With the overarching goal to understand the interactions of this large set of isolated microbes with their environment, the ENIGMA team has developed high-throughput technologies that allow large-scale phenotypic investigation and in-depth characterization.

Some of the molecular genetic tools we have developed, such as RB-TnSeq, led to a considerable amount of gene function inferences from gene-phenotype measurements. We are working on extending the application of this approach to previously genetically challenging microbes. We developed an automated platform for DNA Affinity Purification (DAP) – seq, a biochemical assay used to characterize protein-DNA interaction sites by in vitro affinity-based protein purification, DNA binding, and next generation sequencing. When applied in parallel, DAP-seq has the power to elucidate the DNA binding sites of a bacterium's entire repertoire of transcription factors.

Another approach, exometabolomics, is based on the comparison of inoculated vs. uninoculated media to identify secreted products and depleted metabolites. This provides direct biochemical observations on consumed and secreted metabolites which can be used to predict resource competition and cross-feeding in microbial consortia and communities. This approach has enabled the rapid profiling of substrate use via LC-MS/MS by FRC isolates. These efforts will result in a better understanding of the coupling between growth substrates, and microbial activity.

We combined RB-TnSeq and metabolomics to explore the impact of stresses on a biological system and showed that the use of multi-omics techniques provides a way to probe complex interactions.

Finally, we developed computational tools capable of incorporating our large-scale experimentally generated data (Fitness browser – https://fit.genomics.lbl.gov/) and providing reliable analyses for genome annotation (Gap Mind - http://papers.genomics.lbl/gov/gaps) facilitating data communication and availability to the scientific community at large.

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 Development, Progress, and Cross-Campaign Investigation of the Abiotic Influences on Denitrification Processes Partitioned Between Synthetic Communities

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], X. Ge[4], M. Thorgersen[4], M. Wells[5], 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; [2]Montana State University, Bozeman; [3]University of Washington, Seattle; [4]University of Georgia, Athens; [5]Oak Ridge National Lab, Oak Ridge; [6]Lawrence Berkeley National Lab, Berkeley; [7]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 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 Modelling Campaign to simulate, model, predict, and characterize mechanistic underpinnings of N_2O emissions in varying ecological contexts of shifts in pH, oxygen, metal availability and community organization.* Insights from these studies will inform the efforts of additional campaigns when performing field studies to test mechanistic hypotheses that may explain observed field phenomenon.

Abstract:

Legacy activities at the Field Research Center (FRC) at Oak Ridge, TN has led to some of the highest recorded subsurface nitrate concentrations [>10g/L] on record. This concerningly large pool of subsurface nitrate is often reduced to the greenhouse gas nitrous oxide (N_2O) via denitrification or from dissimilatory nitrate reduction to ammonia (DNRA), which eventually can be converted to N_2O from a variety of microbial processes. It is important to understand the environmental conditions that favor complete denitrification (N_2 emission) or incomplete denitrification (e.g., N_2O emission) in the subsurface. 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 a rainfall events revealed a sudden decline in pH up to 1.5 units over a matter of hours. This phenomenon has been observed in other soil environments, and during such perturbations, denitrification processes may be stimulated, leading to the increased production of N_2O and N_2. We therefore, hypothesized that if the process of complete denitrification is partitioned among incomplete denitrifiers, it is subject to specific abiotic influence that may lead to increased N_2O off-gassing. To test this hypothesis, we have organized a cross-campaign effort to elucidate different mechanisms of abiotic control, in particular pH shifts, C/N ratios, microaerobic environments, and metal availability. Here we have established a synthetic community (SynCom) of two field isolates --*Rhodanobacter sp*. R12 and *Acidovorax sp*. 3H11-- which together can perform complete denitrification. 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. By

analyzing transcriptional profiles, we aim to model, predict, and characterize mechanistic underpinnings of N_2O emissions in varying ecological contexts. Ongoing experiments are focused on additional environmental controls such as, shifts in pH at differing C/N ratios, oxygen, and metal availability (e.g., Ni). Insights from these studies will inform field studies that may validate model-driven hypotheses.

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

Strain dynamics and functional diversity of 22 high-quality single cell genomes from ENIGMA ground water

A. Zhang[1] (anniz44@mit.edu), D. M. Needham[1,2], W. Zheng[3], S. Zhao[1], Y. Yin[1], D. A. Weitz[3], T. C. Hazen[4,5], N.S. Baliga[6], A.M. Deutschbauer[2], M.W. Fields[7], M.W.W. Adams[8], E.J. Alm[1], R. Chakraborty[2], J.-M. Chandonia[2], D.A. Elias[5], T.R. Northen[2], A. Mukhopadhyay[2], G.E. Siuzdak[9], D.A. Stahl[10], P.J. Walian[2], J. Zhou[11], A.P. Arkin[2,12] and **P.D. Adams[2,12]**

[1]Massachusetts Institute of Technology; [2]Lawrence Berkeley National Lab; [3]Harvard University; [4]University of Tennessee; [5]Oak Ridge National Lab, [6]Institute for Systems Biology; [7]Montana State University; [8]University of Georgia; [9]Scripps Research Institute; [10]University of Washington; [11]University of Oklahoma, [12]University of California at Berkeley

http://enigma.lbl.gov

Project Goals: We aim to obtain novel genomes with high quality (of completeness and contamination) from ENIGMA samples through single-cell sequencing and integrate them into KBase as good references for not-yet cultured bacteria in natural environments.

Microbial communities are diverse, dynamic ecosystems comprised of genetically diverse populations. Single-cell sequencing yields information about strain-level differences within a population that can hardly be obtained by metagenome sequencing. Thus, we aim to obtain high quality novel genomes from ENIGMA samples through single-cell sequencing and develop a pipeline to decode strain-level variations in metagenomes. Importing them into KBase as good references for ENIGMA community could have an outsize impact on environmental microbiology research.

We profiled 22 high-quality and high novelty single cell genomes (with >97% complete and < 1% contamination) from an ENIGMA groundwater well GW462 by droplet microfluidics (Microbe-seq). We observed high strain dynamics and functional diversity in these single cell genomes. To reveal the history of bacteriophage infections and adaptive immune system in ENIGMA bacterial species, we detected CRISPR loci from 4 of 22 single cell genomes (from *Aeromonas*, *Alkanindiges*, *Propionivibrio*, and *Rhodobacteraceae*) using crass crispr. Spacer sequences were extracted and searched agaisnt NCBI nt database to identify potential phages.

We identified a total of 154 unique spacers from 4 single cell genomes, indicating intense infections of bacteriophage and coevolution of bacterial host and phage. Among 154 spacers identified in single cell genomes, only 3 spacers in Aeromonas species were annotated as Aeromonas phages by NCBI. This indicates that our knowledge of bacteriophage targeting natural microbes is still quite limited. We also found the same spacer carried by multiple CRISPR.

We annotated the pathways in 22 single cell genomes by KEGG and EGGNOG databases using hmmsearch. We searched for pathways that exhibiting maximal functional diversity across single cell genomes with at least 2 fold-change and with the maximum of at least 10 genes found in a single cell. We observed high functional diversity in pathways involving antimicrobial resistance (to beta-lactam, vancomycin, and cationic antimicrobial peptide CAMP), energy metabolisms (nitrogen metabolisms, methane metabolisms, sulfur metabolisms, and oxidative phosphorylation), and xenobiotics degradation (benzoate degradation, toluene degradation, polycyclic aromatic hydrocarbon degradation, aminobenzoate degradation, nitrotoluene degradation, styrene degradation, and caprolactam degradation). This indicates distinct strategies of adaptation and diversification among bacterial species in ENIGMA environment. We further annotated nitrogen metabolisms pathways by Fama. We found a high diversity of *Burkholderiales* species in nitrogen metabolisms involving denitrification, nitrification, and dissimilatory nitrate reduction.

All single cell genomes were imported into KBase (https://narrative.kbase.us/narrative/65855) and annotated by Fama (https://narrative.kbase.us/#catalog/apps/FamaProfiling/run_FamaGenomeProfiling/), as good references for environmental microbiology research in ENIGMA community.

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

Optimal Experimental Design (OED) of Biological Systems

Francis J. Alexander,[1,*] (falexander@bnl.gov), Ian Blaby[2], Edward Dougherty[3], Puhua Niu, Maria Soto[2], Xiaoning Qian[3], and Byung-Jun Yoon[1,3]

[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 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.

Any future bio-economy likely will include a spectrum of engineered organisms. As sources of economically valuable products, prokaryotes offer many beneficial attributes (e.g., rapid growth and diverse metabolic capabilities), including the production of multiple value-added products that can offset the cost of bioenergy products. However, the biological complexity and diversity of these organisms impede development of genome-wide engineering strategies. Lack of knowledge about proteins that participate in or regulate given processes presents a barrier to predictive engineering. Consequently, despite recent molecular advances with Clustered Regularly Interspaced Short Palindromic Repeats associated (CRISPR Cas)-based tools, knowing what and how to engineer organisms to achieve a desired goal remains a bottleneck, resulting in many genome engineering projects that do not meet expected outcomes. Even with simple organisms such as prokaryotes, knowledge is highly uncertain and incomplete. Understanding how these systems respond to an intervention is even less exhaustive. Thus, such paucity of knowledge regarding complex biological systems requires robust optimization strategies.

While computing infrastructures can assist bench scientists in designing experiments that can effectively fill knowledge gaps in biological networks, designing and implementing these infrastructures remain significant tasks. Data derived from biological experiments are multifaceted, multidimensional, and originate from different sources (i.e., organisms), and interpretation often requires understanding and analyzing multiple fields of research. Consequently, engineered organisms may exhibit unanticipated outcomes.

Addressing these challenges requires a probabilistic framework for integrative modeling of heterogeneous omics data (especially transcriptomics and metabolomics data), quantification of the uncertainty affecting the objective (i.e., strain improvement to optimize metabolite yield), and

designing the optimal experiment that can effectively reduce this objective-based uncertainty. The MOCU (mean objective cost of uncertainty) concept and the MOCU-based OED framework proposed in this project are well suited for overcoming these challenges.

This project exploits the team's collective expertise in systems biology, high-performance computing, mathematical modeling, and control of uncertain complex systems to: (1) take advantage of existing models and data, even when there is uncertainty, to robustly predict optimal experiments; and (2) employ an OED framework to optimize the outcome in an efficient manner (i.e., fewer experiments and less guesswork), where optimization is achieved by optimally (most favorably) improving knowledge about the model (or the microbial system represented by the model) relevant to the objective.

We have developed a new flexible analysis pipeline, **TRIMER**: Transcription Regulation Integrated with MEtabolic Regulation, enabling integrative systems modeling of transcription factor (TF) regulated metabolism. In this workflow, we adopt a Bayesian network (BN) inferred from large-scale gene-expression compendia, rather TF-gene conditional probabilities, which enables the incorporation of prior relational knowledge when modeling TF regulations that affect metabolism. Consequently, our modeling framework can take advantage of pathway knowledge and quantify the impact of extending our current knowledge regarding transcription regulation via future experiments on the objective (i.e., optimize metabolite yield). Based on the constructed BN, we can infer the probabilities of gene states of interest, and consequently predict genome-scale metabolic fluxes of mutants by TF knockouts. Additionally, we have developed a simulation framework to mimic the TF-regulated metabolic network, which is capable of generating both gene expression states and metabolic fluxes, thereby providing a fair evaluation platform for benchmarking models and predictions. Here, we present progress on these computational pipelines as well as their applicability to both simulated and actual experimental data.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER).

Developing episome-based gene expression platforms in the model diatom *Phaeodactylum tricornutum*

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

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

Project Goals: We propose integration of genome-scale modeling with genome engineering to optimize energy and metabolte flux through subcellular compartments to promote efficient production of high value and fuel-related metabolites. We aim to construct streamlined artificial chromosomes encoding reprogrammed biological modules designed for *in vivo* optimization of electron flow efficiency, photosynthesis, and overall cellular growth while directing key metabolic precursors away from storage carbohydrates and into lipids or branched chain amino acids (BCAA). The underlying goal of the proposed research is to produce strains of diatoms encoding cellularly compartmentalized biosynthesis pathways on an artificial chromosome, with the natural genetic background altered to include knockouts of respective native genes as well as the installation of in vivo metabolite bioreporters. Specific goals and technical approaches are focused around four themes: 1) Modeling and Flux studies, 2) Photosynthetic efficiency, 3) Linking metabolic and regulatory networks, 4) Genome scale engineering methodology and application.

Developing genetic approaches for efficient and scalable disruption of gene expression can aid in dissecting mechanisms governing cellular processes, studying gene function, enable high throughput genome-scale screens and assist in redirecting metabolic flux towards high-value metabolites. In diatom algae, RNAi has been the most commonly used gene expression knockdown tool. The current state of the art for inducible gene expression is based on endogenous promoters that respond to different environmental conditions that also change host physiology. Here, we are developing highly efficient, target specific, tunable, and scalable transcriptional control system in the model diatom *P. tricornutum*. We designed and tested six episome-based chemical inducible gene expression systems that have been proven effective in other eukaryotic organisms. Using flow cytometry, we assess time- and dose-response dynamics of each expression system and its reporter protein (YFP). Addition of a chemical inducer/ligand to transgenic strains activates transcription with a dynamic range of up to ~450-fold. We demonstrate our transcriptional control system is tunable and reversible in a dose-and time-dependent manner. Using RT-PCR, we found that inducer dependent transcriptional activation starts within ten minutes of addition and saturates within two hours - without any detectible expression in controls. Upon inducer withdrawal, the expression of the reporter protein was nearly undetectable after 24hrs. The system described here will expand the molecular and synthetic biology toolkits in algae, facilitate gene discovery efforts and tailoring of organisms for maximum productivity.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018344.

Title: Cas9-mediated mutagenesis of GS-GOGAT genes in the diatom Phaeodactylum tricornutum

Authors: Mark Moosburner,[1,2]*, Pardis Gholami[2], James McCarthy[2], and **Andrew E. Allen**[1,2] **(aallen@jcvi.org)**

Institutions: [1]Scripps Institution of Oceanography, University of California, La Jolla, CA; [2]J. Craig Venter Institute, La Jolla, CA

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

Abstract text: Diatoms contribute tremendously to primary production in contemporary marine food webs partially due to their ability to quickly assimilate and out-compete other microbes for nitrogenous compounds. Nitrogen assimilation in diatoms is multifaceted in that it uniquely contains two complete glutamine synthetase - glutamate synthase (GS-GOGAT) cycles, one localized to the chloroplast and one to the mitochondria. In order to elucidate the function of genes involved in nitrogen assimilation and of entire GS-GOGAT cycles (chloroplast or mitochondria), CRISPR-Cas9 was employed in the diatom *Phaeodactylum tricornutum* to knock-out individual genes and two genes simultaneously. Episomal delivery of CRISPR-Cas9 was used to deliver an antibiotic-based selectable Cas9 and one or two guide-RNA expression cassettes via bacterial conjugation. To aid in the workflow in generating Cas9 episomes, a colorimetric cloning methodology was developed based on hierarchical Golden Gate assembly[1]. Using this Cas9 episome, mutant cell lines were generated for four genes, the chloroplastic GS (GSII) and GOGAT and the mitochondrial GS (GSIII) and GOGAT. Also, both the mitochondrial GS and GOGAT were knocked out together, which resulted in mutant cell lines with reduced growth rates. A paired knock out of the chloroplastic GS and GOGAT enzymes was not possible, potentially due to a severely decreased or abolished growth rate subsequently. The single-gene mutant cell lines for GSII and GSIII did not impair growth regardless of the nitrogenous compound supplemented for growth. Interestingly, GSII mutants appear to have elevated lipid content during early and mid-exponential phases compared to wild-type *Phaeodactylum* while growing comparably. Here, physiological growth measurements of mutant cell lines were used to hypothesize the respective functions of GSII and GSIII in the context of lipid biosynthesis.

References/Publications

1. Moosburner, M. A., Gholami, P., McCarthy, J. K., Tan, M., Bielinski, V. A., & Allen, A. E. (2020). Multiplexed knockouts in the model diatom phaeodactylum by episomal delivery of a selectable cas9. Frontiers in microbiology, 11, 5.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018344

Title: Mapping transcription factor-mediated remodeling of diatom metabolism in response to shifting environmental conditions

Authors: Sarah R. Smith[1], Hong Zheng[1], Pratap Venepally[1], and **Andrew E. Allen[1,2]* (aallen@jcvi.org)**

Institutions: [1]J. Craig Venter Institute, La Jolla, California; [1]Scripps Institution of Oceanography, University of California, San Diego

Project Goals: We propose integration of genome-scale modeling with genome engineering to optimize energy and metabolte flux through subcellular compartments to promote efficient production of high value and fuel-related metabolites. Through the proposed research activities, we aim to construct streamlined artificial chromosomes encoding reprogrammed biological modules designed for *in vivo* optimization of electron flow efficiency, photosynthesis, and overall cellular growth while directing key metabolic precursors away from storage carbohydrates and into lipids or branched chain amino acids (BCAA). The underlying goal of the proposed research is to produce strains of diatoms encoding cellularly compartmentalized biosynthesis pathways on an artificial chromosome, with the natural genetic background altered to include knockouts of respective native genes as well as the installation of in vivo metabolite bioreporters. Specific goals and technical approaches are focused around four themes: 1) Modeling and Flux studies, 2) Photosynthetic efficiency, 3) Linking metabolic and regulatory networks, 4) Genome scale engineering methodology and application.

Abstract text: Transcription factors regulate gene expression by binding DNA and promoting (activate) or block (repress) recruitment of RNA polymerase. Therefore, they are a key part of regulating the enaction of gene expression programs that govern regular growth (progression through the cell cycle) and allow cells to respond and acclimate to shifts in environmental conditions (such as light, and nutrient status) to optimally maintain homeostasis and survive. Despite the importance of diatoms in the marine environment, and their relevance for biofuels development, little is known about these mechanisms with any molecular detail. DNA affinity purification sequencing (DAP-Seq) is a high-throughput *in vitro* method to characterize transcription factor binding sites genome-wide. Using this method, we have mapped the transcription factor binding sites for >100 transcription factors in the *P. tricornutum* genome. 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. Notably, we identify a homolog of the fungal regulator of the nitrate regulon and link it to the regulation of nitrate metabolism in *P. tricornutum* and across diatom diversity. Further, we identify transcription factors involved in other key cellular events such as the coordination of metabolic shifts across diel cycles, and activation of the carbon concentrating mechanism. In diatoms, heat shock factors (HSF) have expanded massively into diatom-specific distinct classes, and we detail for the first time, variation in the binding sites of the different HSF classes providing insight into the functional

significance of their evolution in this group. Through linking the prevalence of binding sites in the promoters of condition-specific regulons, we develop the first genome-scale view of the metabolic regulatory network in diatoms.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018344

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), Graham Peers[3], Jamey D. Young[1,2],**Andrew Allen[4,5]**,

[1] Department of Chemical and Biomolecular Engineering, [2] Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN [3] Department of Biology, Colorado State University, Fort Collins, CO [4] Microbial and Environmental Genomics Group, J.Craig Venter Institute. La Jolla, CA [5] 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 store 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 light intensity 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 observed higher metabolic flux in the TCA cycle under high light relative to the low light. Further, we observed that the TAG profile at high light contained significantly more omega-3 fatty acids compared to low light.

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 (Pt-WT), Pt-WT without nitrate (Pt-WT_N-), and Pt-NR with nitrate (Pt-NR). We found enhanced accumulation of TAG in Pt-WT_N- and Pt-NR relative to Pt-WT. Concomitantly, the pool sizes of pyruvate (end product of glycolysis) and amino acids decreased significantly in Pt-WT_N- compared to those in Pt-WT, whereas the pool sizes of TCA cycle metabolites and urea increased dramatically. 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 N- conditions[3]. Interestingly, Pt-NR showed a similar trend of abundance of intracellular metabolites except a few in the TCA cycle. Our preliminary ^{13}C-MFA results have revealed remarkable differences in the metabolic fluxes between Pt-WT, Pt-WT_N- and Pt-NR.

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 storages. 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.

(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)

Diel Activity of Microbial Communities in Surface Soil Litter

Eric W Morrison[1*] (eric.morrison@unh.edu), Jennifer BH Martiny,[2] Steven D Allison[2], Michael L Goulden,[2] Alexander B Chase,[3] Zulema Gomez Lunar, [2] Claudia Weihe,[2] and **Adam C Martiny**[2]

[1]University of New Hampshire, Durham, NH; [2]University of California–Irvine, Irvine, CA; [3]University of California–San Diego, San Diego, CA

The overarching goal of this project was to quantify potential diel partitioning of metabolic functions in microbial communities of decomposing grass litter in a Mediterranean climate-grassland ecosystem. We measured litter respiration as an indicator of microbial activity overall, and used gene expression data to determine activity of specific taxa or functions. We ask 1) is there a diel signal in microbial gene expression, 2) what are the predominant abiotic drivers of short-term transcriptomic variation, 3) is there diel signal in the activity of dominant taxa, and 4) which genes are most responsive to diel cycling?

Diel fluctuations are known to be important for the functioning of photosynthesis-dominated microbial ecosystems – for example, aquatic communities, plant-associated rhizospheres, and desert crusts. In these cases, heterotrophic activity often appears to track light-dark cycles as nutrient and carbon availability fluctuates in relation to the activity of photosynthetic organisms. It is less clear whether there are prominent diel patterns in activity of soil microbes. In surface soil, including decomposing litter, microbes may be exposed to environmental variability that is diel in nature – including temperature, moisture, and UV radiation. Indeed, there is evidence that CO_2 flux from litter displays diel fluctuations in diverse ecosystems including wetlands, Mediterranean woodlands, and temperate forests. To the extent that diel variation is a predominant driver of microbial activity there is potential for specialization among soil microbes such that particular taxa may be active and/or perform different functions at different times. Our goal in this study was to quantify short-term variation in microbial activity and gene expression in litter decomposition. In particular we ask 1) is there a diel signal in microbial gene expression, 2) what are the predominant abiotic drivers of short-term variation, 3) is there diel signal in the activity of dominant taxa, and 4) which genes are most responsive to diel cycling? We addressed these questions using a metagenomic and metatranscriptomic sequencing approach in combination with litter respiration, temperature, and moisture measurements under a high-temporal resolution sampling over three continuous days. This study was performed at the Loma Ridge Global Change Experiment (LRGCE) in a Mediterranean climate, grassland ecosystem. Three plots were established within the larger experimental site, with 36 spatially randomized litterbag samples per plot, and one litterbag per plot sampled every two hours for three days. Metagenomic and metatranscriptomic sequencing was performed by the DOE Joint Genome Institute.

We observed diel fluctuations in temperature, moisture, and litter respiration over the three-day period. There was a gradient in terms of average moisture content across the three plots (P = 0.0015) ranging from 24% to 47%. This gradient resulted in the two plots with lower moisture content having stronger diel fluctuations in moisture content ($P < 0.01$), whereas no significant diel signal in moisture was observed in the third plot ($P = 0.12$). We also observed diel fluctuations in respiration in the two plots with lowest and highest moisture content (P = 0.01), with peaks in respiration coinciding with the peak in moisture content or temperature,

respectively. Community composition did not display a strong diel signal. However, there were diel shifts in the expression of specific metabolic functions correlated with moisture and temperature but varying by plot. Five of the ten most abundant taxa (identifiable to genus or family) displayed significant diel signal ($P < 0.05$) in the high-moisture plot, including the most abundant genus, *Curtobacterium*. In terms of functional genes, after correcting for multiple comparisons 101, 35, and 205 groups of KEGG orthologs displayed significant diel signal in the low-, intermediate-, and high moisture plots, respectively. Overall, our data demonstrate that gene expression of heterotrophic microbial litter decomposers responds to diel variability in the environment. Taxa appear to respond to this variation in unique ways, with some taxa responding more strongly to diel moisture variation and others responding to diel temperature variation. We further observed a diel signal in functional gene expression overall and in the expression of specific functional genes, indicating that microbes may perform particular functions in accordance with day-night cycles in environmental conditions.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-PS02-09ER09-25 and DE- SC0016410.

Title: Differential Response of Microdiversity to Simulated Global Change Within a Bacterial Genus

Authors: N.C. Scales*[1] (nscales@uci.edu), A.B. Chase[2], S. Finks[1], C. Weihe[1], S.D. Allison, **J.B.H. Martiny[1]**

Institutions: [1]Department of Ecology and Evolutionary Biology, University of California, Irvine, California, USA; [2]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California

Project Goals:

We sought to establish an amplicon sequencing method which would allow us to characterize within-genus diversity (microdiversity) of an ecologically important soil bacterium to simulated global change. *Curtobacterium* is one of the most abundant bacteria in the surface litter layer of soil at the Loma Ridge Global Change Experiment. We applied this method to determine the subclade-level response of the taxon to simulated drought and nitrogen addition. We did this within two ecosystems, a grassland and a shrubland, to allow us to elucidate differences in response to the global change treatments by ecosystem type. Lastly, we sought to explain these differences by analyzing the genomes of *Curtobacterium* strains to look for differences in carbohydrate-active enzyme (CAZyme) content.

Abstract text:

Global change experiments often observe shifts in soil bacterial composition based on 16S rRNA gene sequences, however, less is known about how global change might alter bacterial microdiversity, defined here as diversity below the genus level. In particular, it is not known whether the broad taxonomic shifts are consistent within a genus, or whether they represent the summation of divergent responses occurring at a finer scale, i.e., whether the 16S-level data masks the microdiversity response. To investigate the response of bacterial microdiversity to global change, we focused on *Curtobacterium*, a genus of gram-positive aerobic Actinobacteria highly abundant in leaf litter, the topmost layer of soil within the Loma Ridge Global Climate Experiment (LRGCE). Established in 2007, the LRGCE manipulates drought and nitrogen in two adjacent plant communities, grassland and coastal sage scrub (CSS), by intercepting approximately 50% of the rainfall, and adding soluble $CaNO_3$. To characterize fine scale diversity within the genus, we used an amplicon sequencing approach, designing *Curtobacterium*-specific primers for the *groEL* gene, a molecular chaperone found in all bacteria. This method revealed an enormous amount of diversity – more than 6,000 exact sequence variants (ESVs) that fall within at least 12 distinct phylogenetic lineages. The composition of *Curtobacterium* microdiversity varied significantly within the global change experiment, across the two ecosystems, and over time. Although added nitrogen did not alter ESV composition, the drought treatment did significantly alter ESV composition, explaining 9.79% of the variation (PERMANOVA: $p = 0.001$). Ecosystem (grassland or CSS) contributed

the largest amount of variation in *Curtobacterium* microdiversity, accounting for 19.6% (PERMANOVA: $p = 0.001$). Since most of the variation was explained by the two ecosystems, we further investigated the carbohydrate-active enzyme (CAZyme) content by subclade among sequenced isolates. The CAZyme genomic content of the strains differed significantly between subclades (ANOSIM: $R = 0.8044$; $p = 0.001$). These data support the idea that resolving fine-scale patterns of niche differentiation in microbes are key to understanding the response of microbial communities to global change.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant numbers: DE-PS02-09ER09-25, DE-SC0016410, and DE-SC0020382.

Unravelling the Role of Pennycress (*Thlaspi arvense* L.) Proteins in the Modulation of Neutral Lipid Droplet Abundance

Julius Ver Sagun[1*] (juliusver.sagun@unt.edu), Athanas Guzha[1], Cintia Arias[1], Tatiana Garcia[2], Allison Barbaglia[2], Erich Grotewold[2], Kent D. Chapman[1], and **Ana Paula Alonso**[1]

[1]Biodiscovery Institute, 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 candidate genes involved in lipid storage and stability using transient expression in *Nicotiana benthamiana*.

The finite nature of crude oil-derived fuels coupled with their adverse effects on the environment means the search for alternative, renewable sources of energy that are more environmentally friendly is paramount. 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 parts of North America, and produces seeds with high oil content (26-39%). It can be grown in a summer/winter rotation cycle with other conventional commodity crops such as maize and soybean and requires low agricultural inputs. 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 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 bioenergy oilseed crop. Transcriptional analysis of 22 pennycress accessions resulted in the identification of potential gene candidates whose expression levels were correlated with seed oil yield (DE-SC0019233). Here, we show that protein products of six of these candidate genes- a lipid transfer protein homolog (LTP6), a lipid droplet associated protein homolog (LDAP3), an annotated lipase (α/β hydrolase), a long-chain acyl-coA synthase protein (LACS1), an endomembrane regulatory protein (RABA3), and a lipid storage and packaging protein (Oleosin)- mainly localize to lipid droplets when transiently expressed in *Nicotiana benthamiana*. The overexpression of coding sequences for these six proteins in *N. benthamiana* leaves resulted in a proliferation of cytoplasmic neutral lipids appearing as droplets under confocal microscopy. Analysis of the infiltrated leaves using GC-MS indicated that the overexpression of these proteins increased the total neutral fatty acid content and somewhat altered the fatty acid composition of *N. benthamiana* leaves. Our data point to possible roles of these six candidate proteins in the compartmentalization and/or stability of pennycress lipid droplets and represent interesting targets for genetic manipulation of pennycress seeds with increased oil content.

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

Elucidating the Temporal and Spatial Organization of Storage Lipids using ^{13}C-labeling in Developing Embryos of pennycress, a Promising Source for Aviation Fuel

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

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

Project Goal: Assessing temporal and spatial regulation of storage lipid accumulations 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. With unpredictable prices, finite fossil fuel sources, and concerns over environmental impact, the discovery of sustainable alternatives to fossil-derived jet fuels is critical. In the last few years, pennycress (*Thlaspi arvense* L.) emerged as a promising oilseed crop especially suited for aviation fuel production due to its moderate 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 yields 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 temporal and spatial organization of storage lipid metabolism in pennycress seeds. This approach allows for the tracing of metabolism from carbohydrate sources to storage oils in near in vivo conditions. The fatty acid composition and growth rate of embryos harvested from these silique cultures were similar to *in planta*, which validates this method. Using this procedure, 100% [U-^{13}C]-glucose was supplied to 16 days after pollination (DAP) siliques for 120 h 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 results showed that the percentage of labeling in plastidic and cytosolic fragments was found to be significantly lower in the axis of the embryos in comparison to the cotyledons. A time course-experiment was conducted to monitor the temporal incorporation of ^{13}C-acetyl fragment in embryos from 14 DAP siliques incubated for 24 h, 48 h, 72 h, and 96 h; the labeling in acetyl fragments gradually increased up to 19%. Further, to evaluate the spatial organization of lipid metabolites in developing seeds, ^{13}C-isotopic labeling and mass spectrometry imaging (MSI) were coupled to analyze metabolic flux *in situ*. Combining isotopic labeling and MSI presents technical challenges ranging from sample preparation, label incorporation, data collection, and analysis. Using currently available software and techniques, ^{13}C-labeled isotopomers were analysed for the membrane lipid and storage oil lipid intermediate phosphatidylcholine (PC). Consistent with ^{13}C-isotopic labeling of fatty acids, MSI revealed greater ^{13}C-isotopic labeling of PC molecular species in the cotyledons than the embryonic axis of developing embryos. Moreover, greater isotopic enrichment in PC molecular species with more saturated and longer chain fatty acids suggest differences in flux related to fatty acid desaturation and elongation. Expanding the combination of ^{13}C-isotopic labeling and MSI to additional time points and to additional lipid intermediates will provide an opportunity to visualize the spatial differences in lipid metabolism during seed development. These data-intensive,

comprehensive approaches will offer greater insights into the internal organization and regulation of oil synthesis in pennycress seeds, which will ultimately support metabolic engineering efforts to produce higher yield of biofuels.

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.

In vivo thermodynamic analysis of glycolysis in *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum* using ^{13}C and ^{2}H tracers

Tyler B. Jacobson[1] (tbjacobson@wisc.edu), Travis K. Korosh[1], David M. Stevenson[1], Charles Foster[3], Costas Maranas[3], Daniel G. Olson[2], Lee R. Lynd[2], and **Daniel Amador-Noguez**[1]

[1]University of Wisconsin-Madison, Madison, WI; [2]Dartmouth College, Hanover, NH; [3]The Pennsylvania State University, University Park, PA

Project Goals: This project will integrate thermodynamic analysis with advanced mass spectrometry, computational modeling, and metabolic engineering to develop an approach for *in vivo* determination of Gibbs free energies (ΔG) in metabolic networks. This project will also investigate how the thermodynamics of biosynthetic pathways in microbial biofuel producers change dynamically as substrates are depleted or products accumulate. This research will result in the construction of computational models that quantitatively define trade-offs between energy efficiency of biosynthetic pathways and their overall catalytic rates. The approach developed in this project will be useful for identifying thermodynamic bottlenecks in native and synthetic pathways and pinpoint the enzymes whose expression levels will have the largest effect on production rates and final product yields.

Abstract: *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum* are thermophilic anaerobic bacteria with complementary metabolic capabilities that utilize distinct glycolytic pathways for the conversion of cellulosic sugars to biofuels. We integrated quantitative metabolomics with ^{2}H and ^{13}C metabolic flux analysis to investigate the *in vivo* reversibility and thermodynamics of the central metabolic networks of these two microbes. We found that the glycolytic pathway in *C. thermocellum* operates remarkably close to thermodynamic equilibrium, with an overall drop in Gibbs free energy 5-fold lower than that of *T. saccharolyticum* or anaerobically-grown *E. coli*. The limited thermodynamic driving force of glycolysis in *C. thermocellum* could in large part be attributed to the small free energy of the phosphofructokinase reaction producing fructose bisphosphate. The ethanol fermentation pathway was also substantially more reversible in *C. thermocellum* compared to *T. saccharolyticum*. These observations help explain the comparatively low ethanol titers of *C. thermocellum* and suggest engineering interventions that may be used to increase its ethanol productivity and glycolytic rate. In addition to thermodynamic analysis, we used our isotope tracer data to reconstruct the *T. saccharolyticum* central metabolic network, revealing exclusive use of the Embden-Meyerhof-Parnas (EMP) pathway for glycolysis, a bifurcated tricarboxylic acid (TCA) cycle, and a sedoheptulose bisphosphate bypass active within the pentose phosphate pathway.

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 *C. thermocellum* and *T. saccharolyticum* using ^{13}C and ^{2}H tracers. mSystems 5:e00736-19. doi: 10.1128/mSystems.00736-19.

This material is based upon work supported by the DOE Early Career Research Program under Award Number DE-SC0018998.

Genetic Improvement of Seed Yield and Oil Content in Field Pennycress

Zenith Tandukar[1] (tandu002@umn.edu), Ratan Chopra[1], Katherine Frels[1], Maliheh Esfahanian[2], Brice Jarvis[2], Liza Gautam[2], Danny Marchiafava[2], Nikhil Jaikulmar[2], John Sedbrook[2], M. David Marks[1], **James A. Anderson[1*]**
[1]University of Minnesota, St. Paul, MN, [2]Illinois State University, Normal, IL.

[*]**Principal Investigator**: ander319@umn.edu

Project Goals:

This project is aimed at understanding the genetic control of seed size and oil content in field pennycress (*Thlaspi arvense* L.; pennycress). This knowledge will be applied to improve pennycress seed and oil yields for its use as a new winter annual cash cover crop for the U.S. Midwest. Winter annual cover crops provide a continuous living cover on otherwise fallow agricultural soil. Genetic improvement of seed yield and oil content will aide in widespread adoption of pennycress as a profitable cover crop that can serve as a source of renewable feedstock for the biodiesel and biofuels industries as we transition to cleaner and more sustainable sources of energy.

Abstract:

Pennycress possesses extreme winter hardiness (up to -30^{o}C), high natural seed yields (1100 – 2250 Kg/hectare), highly oil-rich seeds (27%-39% on a dry weight basis), and protein for food, feed and other end uses (Sedbrook et al. 2014). Pennycress is planted in the fall and grows into a rosette that is dormant during the winter, flowers and sets seed in the spring, and is harvested in June. Fall planted pennycress acts as a continuous living cover on the otherwise fallow soil. This reduces pollution of our water sources through nutrient run-off and soil erosion from barren farmlands. In the spring, pennycress provides an early food source for pollinators, as well as helps in suppressing weeds. Cropping systems that can leverage optimal growing conditions for both the major summer annual cash crops (e.g. corn and soybeans) and winter annual pennycress can help maximize farmer profits and environmental benefits in the future.

The widespread adoption of pennycress will depend upon its economic viability as a second cash cover crop where the benefits outweigh the cost of inputs and farmer effort. Pennycress has spent most of its evolutionary history as a weed and has only recently been the subject of domestication and breeding efforts. The first generation of domesticated pennycress with improvements in reduced pod shatter, low erucic acid and glucosinolates concentration, and earlier flowering have already been achieved (Chopra et al., 2019). However, increased seed size and higher oil content are desirable for maximum yield and productivity and to make it amenable for mechanical processing. Therefore, our current research aims to i) identify loci that control seed size and oil content in a diverse collection containing wild germplasm via association mapping, ii) identify loci that control seed size and oil content in two biparental recombinant inbred line (RIL) populations, iii) characterize genes controlling seed size and oil content in pennycress EMS mutagenesis populations, and iv) generate pennycress CRISPR-Cas9 knockouts in genes known to regulate seed size and oil content in *Arabidopsis*.

The diversity panel has been grown in four locations over three years and genotyped with genotyping-by-sequencing to produce a rich dataset for genome wide association study. Two F_{6-7} RIL populations have been developed from parents divergent in seed size and oil content and planted in field conditions for the first time during fall 2020. These populations will be genotyped and phenotyped in 2021 for QTL mapping. We have generated and phenotyped over 15,000 EMS pennycress lines generated in MN106 genetic background and identified large seeded as well as high oil mutants which are currently being tested in field conditions. In addition, several single and stacked CRISPR-Cas9 mutants for both seed size and oil content have been generated in Spring32 genetic background. The results from these projects will help advance the knowledge of the genetic basis that underlies seed size and seed oil content in pennycress. Research relating to the development of large seeded and high oil pennycress can ensure the use of pennycress as a renewable feedstock for the biofuels industry, and in addition serve as a valuable resource for more translational research into other major oilseed crops that provide a significant portion of edible oil around the world.

References

1. Chopra, R., Johnson E. B., Emenecker, R., Cahoon, E.B., Lyons, J., Kliebenstein, D.J., …. Marks, M.D. Identification and stacking of crucial traits required for domestication of pennycress. *Nature Food.* 1(1): 84-91.
2. Sedbrook, J., Phippen, W.B., Marks, D.M. (2014). New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: Example pennycress. *Plant Sci*. 227:123-132.

Funding statement.
This research was supported by the Plant Feedstock Genomics for Bioenergy: A Joint Research Funding Opportunity: USDA, DOE. Grant No. DE-FOA-0001857, Award No. 2019-67009-29004

Creating representative microbial communities to promote biological nitrogen fixation on sweet sorghum

April MacIntyre[1]* (macintyre@wisc.edu), Madeline Hayes[1], Jaron Thompson[1], Vânia C. S. Pankievicz[1], **Sushmita Roy**[1], **Wilfred Vermerris**[2], **Ophelia Venturelli**[1], **Jean-Michel Ané**[1]

[1]University of Wisconsin-Madison

[2]University of Florida, Gainesville, FL

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 health, and growth.*

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

Biological nitrogen fixation (BNF) on cereal crops to increase crop production sustainability and reduce environmental damage from synthetic fertilizers has long been an agronomical Holy Grail. Cereal crops cannot form symbiotic relationships with nitrogen-fixing rhizobia. They rely on synthetic fertilizer inputs and less efficient associative nitrogen fixation or free-living nitrogen fixation from soil microbes[1]. Some indigenous maize accessions from Central America can obtain 29%-82% of their nitrogen from the air through BNF[2]. This 'nitrogen-fixing corn' recruits environmental diazotrophic microbes to live around its aerial roots, where root border cells produce mucilage that provides a low oxygen, sugar-rich niche suitable for BNF. Similar aerial root production and BNF have also been noticed in sorghum accessions but have not been characterized like maize. Since sorghum is a crop of interest for biofuel production, breeding this aerial root trait into more sorghum lines and optimizing BNF on aerial roots are target areas for research and development. On the microbial side, we aim to 1) *isolate and characterize microbes associated with sorghum aerial roots*, 2) *form representative communities from these isolates and observe community dynamics through model building,* and 3) *test the efficacy of these synthetic communities on protecting sorghum health and promoting biomass accumulation*. Currently, we have a diverse environmental isolate collection of ~90 strains selectively isolated on artificial mucilage media from aerial root producing corn and sorghum lines. With additional mucilage microbiome data, we may add more strains if we feel our selection missed key community members. All environmental isolates will be screened for plant growth-promoting phenotypes, such as nitrogen fixation, auxin and siderophore production, and polysaccharide catabolism. Further, we will examine the conditions under which these environmental diazotrophs release ammonia or nitrate to the plant via the mucilage and the survival of these strains around aerial roots through wet/dry cycles. Using high-throughput, combinatorial, *in vitro* community assembly experiments, we will measure temporal changes in community composition and ammonia production. These data will be used to build computational models at different resolutions to elucidate keystone bacterial species in community growth, survival, and

BNF, and decipher the microbial and metabolic interactions driving ammonia production. Our experimental data, coupled to computational models, will guide the selection of defined communities that can be studied in the context of sorghum accessions. With these defined communities we will confirm survival, nitrogen fixation, and promotion of sorghum nutrition, growth, and health in the field.

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

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, doi:10.1186/s12915-019-0710-0 (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, doi:10.1371/journal.pbio.2006352 (2018).

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

Emily Reed,[1*] Saddie Vela,[1*] Vânia C. S. Pankievicz,[2] Saptarshi Pyne,[2] April MacIntyre,[2] Sushmita Roy,[2] **Jean-Michel Ané,[2]** Wilfred Vermerris[1*] (wev@ufl.edu)

[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 is 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 that will be necessary to ensure reasonable yields raises the cost of production and 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 to 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. This symbiosis allows them to obtain a significant amount of N from the air, reducing the need for chemical fertilizers. Specifically, we are investigating the genetic basis of aerial root formation using a genome-wide association study of two panels of genetically diverse genotypes. We are also determining the dependence of aerial root formation on factors such as rainfall and soil composition by planting select genotypes in different geographic locations and with varying levels of fertilizer. In parallel, we are using single-cell 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. Furthermore, breeding populations derived from crossing bioenergy sorghums with accessions that support BNF are under development. 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 inputs.

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] Fatima Foflonker,[2] Michael Fonstein,[1] Sara Forrester,[1] Stephanie Greenwald, [1] Christopher S. Henry,[2] 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] Yasuo Yoshikuni, [4,5] Steven R. Fleming,[6] Ashty S. Karim,[6] Brenda Wang,[6] Michael C. Jewett,[6] 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 are creating the conditions 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.

Grammar and Language of CRISPR/Cas-Targetable Sites in *Escherichia coli*, *Shigella*, *Pseudomonas*, and *Salmonella*: A Comprehensive Survey

Carla M. Mann[1*] (cmann@anl.gov), Rebecca Weinberg,[2] Gyorgy Babnigg,[2] Peter E. Larsen,[2] Marie-Francoise Gros,[2,3] Philippe Noirot,[2,3] **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]National Research Institute for Agriculture, Food and Environment (INRAE), France

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 are creating the conditions for a paradigm shift in our understanding of the ways that cellular function can be designed at the level of bacterial communities.

The advent of CRISPR/Cas revolutionized precision genome editing and engineering. However, the technology has introduced its own set of challenges, particularly in choosing a target site within a genome that matches potentially stringent experimental parameters, such as efficiency of double-strand break (DSB) induction, DSB location, and number of off-target sites. Even when a suitable target site has been identified in a model organism, there is no guarantee that that particular target will be present in closely related species. These requirements demonstrate the importance of understanding the genomic landscape of CRISPR/Cas-targetable sites to better understand the rules governing the behavior of gRNAs. To that end, we surveyed the genomes and CRISPR/Cas-targetable sites within 16,171 *Escherichia coli*, 446 *Shigella*, 7,867 *Pseudomonas*, and 14,604 *Salmonella* genomes from the Pathosystems Resource Integration Center (PATRIC) [1,2] database.

We analyzed the quality of the genomes present in the PATRIC database for each species using the scoring system devised by Land et al. [3] and demonstrate a species quality bias – nearly 97% of *E. coli* genomes meet our quality threshold, while only 15% of *Shigella*, 75% of *Pseudomonas*, and 65% of *Salmonella* genomes meet that threshold.

Using the high-quality genome assemblies, we calculated genomic GC content, identified rare codons within each species, and identified the species' core genes. We further identified all SpCas9, St1Cas9, SaCas9, NmCas9, CjCas9, and AsCas12a-targetable sites in each genome. Using MASH [4], an implementation of the MinHash algorithm for estimating genomic distance, we found that these Cas-targetable sites are as well conserved among each species as the overall

genomes themselves, although those sites located within highly conserved core genes and operons are themselves highly conserved.

References

1. Davis J.J. et al. *The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities.* Nucleic Acids Res. 2020 Jan 8;48(D1):D606-D612. DOI: 10.1093/nar/gkz943.
2. Wattam A.R. et al. *Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center*. *Nucleic Acids Res*. 2017 Jan 4;45(D1):D535-D542. DOI: 10.1093/nar/gkw1017
3. Land, M.L. et al. *Quality scores for 32,000 genomes*. Stand Genomic Sci. 2014; 9:20 DOI: 10.1186/1944-3277-9-20
4. Ondov B.D., et al. *Mash: fast genome and metagenome distance estimation using MinHash.* Genome Biol 2016 17:132 DOI: 10.1186/s13059-016-0997-x

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 Under Diverse Physiological Conditions

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 are creating the conditions 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 modified and engineered organisms necessitates the creation of secure and efficient biocontainment systems to prevent these organisms from escaping the laboratory and endangering the environment, public health, and public perception of scientific research. Safeguards based on controlled activation of a self-targeting CRISPR/Cas9 "self-destruct" mechanism that activates outside of laboratory conditions are transferable between different organisms, cost-effective, and relatively easy to implement in comparison to other safeguard systems. 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 that variability is heightened under different environmental conditions. We have developed a machine-learning prediction method, CRISPRAct, that models this behavior across different environmental conditions to assist in identifying gRNAs for use in the proposed secure biosystem.

We hypothesized that dynamic gene expression responses to varying physiological conditions would influence the cell killing activity of the CRISPR/Cas9 system. We created a library of gRNAs composed of ~180,000 gRNAs corresponding to sites throughout the *Escherichia coli* MG1655 genome and control gRNAs (~20,000) that did not match. Using this library, we conducted screens for killing activity in three physiological conditions: rich (LB) media in exponential growth; defined (M9) media in exponential growth; and rich media in stationary phase. We then captured sequence data representing the transcriptomes corresponding to these conditions, and the killing activity of the gRNAs over time. Using our libraries, we identified ~6,000 guides that were statistically overrepresented (via ANOVA testing) for

physiology-specific function. A small subset of gRNAs (174) were "outlier switches" that exhibited outstanding killing activity in a specific physiological condition, while their high prevalence in another condition was consistent with providing a growth advantage. The library screens generated an extensive dataset of ~530,000 data points used to develop CRISPRAct to interrogate features influencing gRNA killing activity.

The CRISPRAct model combines a state-of-the-art natural language processing (NLP) model based on the Google AI ALBERT architecture [3] with conventional neural network (NN) models to predict the efficiency of gRNAs under various conditions. The NLP model treats genomic context as a machine-interpretable "language" by training embeddings on a representative set of twelve *E. coli* genomes [4] divided into "sentences" of seven "words" that each consist of three bases, and then fine-tuning on the gRNA activity dataset. The NN models use a set of 428 features including gRNA positional and physicochemical properties [1], as well as the *E. coli* growth phase in the form of optical density, and environmental factors including media growth concentrations. The results of these models are combined through a polynomial regression to predict the percent change in *E. coli* cell population after Cas9 induction. CRISPRAct achieves a Mean Square Error of 0.13 and Spearman Correlation Coefficient of 47.31% on a test dataset of 71,228 gRNAs. This level of model performance is comparable to other efforts to predict gRNA activity. However, due to our approach of leveraging physiological conditions to train models, CRISPRAct offers a more resilient basis for transfer learning in novel organisms and environmental conditions without having to undergo a costly re-training process. This strategy will decrease the time and effort needed by other researchers to identify gRNAs with desired behaviors in various physiological conditions.

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*. arXiv, 2019: 1909.11942
4. Abram, K., et al., *What can we learn from over 100,000 Escherichia coli genomes?* bioRxiv, 2020

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.

ModelSEED release 2: High Throughput Genome-Scale Metabolic Model Reconstruction and Analysis in KBase

José P. Faria[1*], Filipe Liu[1], Janaka N. Edirisinghe[1], Samuel M.D. Seaver[1], James G. Jeffryes[1], Qizh Zhang[1], Pamela Weisenhorn[1], Boris Sadkhin[1], Nidhi Gupta[1], Tian Gu[1] and Christopher S. Henry[1], Robert Cottingham[2], and **Adam P. Arkin[3]**

[1]Argonne National Laboratory, Lemont, IL; [2] Oak Ridge National Lab, Oak Ridge, TN; [3]Environmental Genomics and Systems Biology, Lawrence Berkeley National Lab, Berkeley, CA.
*presenting author

http://www.kbase.us

Project Goals: Short statement of goals. (Limit to 1000 characters)

The Department of Energy Systems Biology Knowledgebase (KBase) is a platform designed to solve the grand challenges of Systems Biology. KBase has implemented bioinformatics tools that allow for multiple workflows, including genome annotation, comparative genomics, and metabolic modeling. First released in 2010, the ModelSEED [1] genome-scale model reconstruction pipeline has now built over 200k draft metabolic reconstructions and supported hundreds of publications. Here, we describe the first major update to this model reconstruction tool with important new features including: (1) a dramatically improved representation of energy metabolism ensuring models produce accurate amounts of ATP per mol of nutrient consumed; (2) new templates for Archaea and Cyanobacteria; and (3) greatly improved curation of all metabolic pathways mapping to RAST and other annotation pipelines.

Abstract text

KBase has made several key improvements to the ModelSEED model reconstruction tool. ATP production was improved in our model reconstruction procedure by constructing core models, testing for proper ATP production from this core, then ensuring that ATP production does not incorrectly explode when expanding the core model to a genome-scale model. We similarly improved our gapfilling approach to ensure that gapfilling does not cause a model to start over-producing ATP. While other approaches aim to correct ATP overproduction in models, these new procedures in the ModelSEED pipeline aim to ensure that ATP overproduction does not happen in the first place. To handle the necessary expansion of templates, we developed machine learning (ML) classifiers to determine automatically which template most correctly applies to a new genome being modeled. The classifiers will also produce new template and biomass objective functions specific to archaea and cyanobacteria (modeled after existing published metabolic models of these species). This ML approach allows for the rapid introduction of additional modeling templates, enabling researchers working with unclassified species or metagenome-assembled genomes extracted to achieve more specific reconstructions.

To improve metabolic pathway annotation completeness and accuracy in ModelSEED models, we first updated our biochemistry database to include the latest reaction data from KEGG, MetaCyc, BIGG, and published models. Next, we manually curated the major pathways in our

reconstruction templates to reconcile pathway representation across these multiple databases. Finally, we curated our mapping of RAST functional roles to this reconciled biochemistry based on data mined from KEGG and published metabolic models.

Within the KBase platform, we demonstrate our improved model reconstruction pipeline on a phylogenetically diverse set of approximately ~6400 genomes (all Bacteria and Archaea complete genomes in KEGG) and constructed draft genome-scale metabolic models (GEMs). We show how the gene counts and modeling metrics (ATP production, biomass yields, reaction classification, pathway representation) are improved with this new release of the ModelSEED. We also selected genomes for which comprehensive Biolog data are available, and we compared model predictions of all data with experimental results. The comparisons showed significant improvement compared to models generated by the original ModelSEED. Finally, we compare the pathway content, gapfilling, size, and gene counts in our models with models constructed by the CarveMe method.

The listed improvements will be available as an update to the ModelSEED reconstruction pipeline. The new release will be made available across all web platforms currently supported; KBase (kbase.us), ModelSEED (modelseed.org), and PATRIC (patricbrc.org) resources.

References

1. Henry, Christopher S., et al. "High-throughput generation, optimization and analysis of genome-scale metabolic models." Nature biotechnology 28.9 (2010): 977-982.

This work is supported as part of the Genomic Sciences Program DOE Systems Biology Knowledgebase (KBase) 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 Partners with SFA Developers to Increase Functionality of Tools and Analysis

Elisha M. Wood-Charlson,[1 *] (elishaw@lbl.gov), Meghan Drake,[2] Chris Henry,[2] Paramvir Dehal,[1] Dylan Chivian,[1] Robert Cottingham,[3] and **Adam P. Arkin.** [1]

[1]Environmental Genomics and Systems Biology, Lawrence Berkeley National Lab, Berkeley, CA; [2]Argonne National Laboratory, Chicago, IL; [3] Oak Ridge National Lab, Oak Ridge, TN.
*presenting author

http://www.kbase.us

Project Goals: (Limit to 1000 characters)
Several SFAs were selected for training on the KBase Software Development Kit (SDK) to integrate tools and data into the KBase platform. Overall, the collaborative effort was viewed as highly successful. Every SFA project was able to achieve their proposed targets for training and functionality, while also gaining an appreciation for the role KBase plays/could play in their future science plans. Both the SFA and KBase teams felt that continued collaboration would be beneficial, especially if the programs were able to jointly prioritize scope and readiness of future projects. This poster will outline the scope of functionality added by SFA community developers, and highlight points of intersection across these efforts.

Abstract text

In total, 5 SFAs were awarded additional funds to support developer time, with the goal of integrating analysis tools into KBase to support the SFA's science. Across the board, this strategy seemed to be well-received, as several SFAs confirm it can be challenging to justify resources for technology-related goals alongside the often complex and ambitious SFA science goals. Two SFAs (LLNL Biofuels and Soil) reported that they are already actively using the new functionality inside KBase, and Stuart acknowledged that the integration has greatly improved their model development. Other SFAs recognized that this effort focused mainly on the core implementation of longer-term analysis goals (LANL Fungal, LLNL Soil, ORNL Structure). All SFAs recognized the value in these efforts to enable the broader research community to also benefit from the additional functionality in KBase.

Overall, the SFAs reported significant progress towards their proposed goals during the additional funding period. For example, KBase now has the first phage analysis pipeline in the system (LLNL Soil), which was presented as an online webinar to the broader user community in November 2020. KBase has several ways for researchers to improve their fine-scale evaluations of microbial communities by exploring strain-level variation (ENIGMA), improve metabolic models by combining various annotation methods (LLNL Biofuels), and screen small-molecule

ligand binding to protein structures (ORNL Structure). Still coming are additional measures to assess quality of contigs prior to assigning taxonomic and/or functional annotations (LLNL Soil), and a joint SFA effort to provide functional annotation for viral sequences (LLNL Biofuels and Soil). Many of the SFAs reported an added bonus to these efforts - increased knowledge of KBase's resource enhancements and community investment, making them ambassadors for KBase in their own communities (LANL Fungal, LLNL Biofuels and Soil).

This poster will outline the major functional components added by SFA developers. The presenter will be available to answer any questions about possible future collaborations.

This work is supported as part of the Genomic Sciences Program DOE Systems Biology Knowledgebase (KBase) 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.

Systems Biology of Isobutanol Production in *Saccharomyces Cerevisiae*

José Montaño López[1*] (jdjml@princeton.edu) and **José L. Avalos**[1,2]

[1]Department of Chemical and Biological Engineering; [2]The Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ.

Project Goals:

The overall goal of the project is to carry out a comprehensive systems biology study of branched-chain higher alcohol (BCHA) production and tolerance in yeast. We will leverage the genetically encoded biosensor of BCHA production described in this presentation to screen yeast genomic libraries to measure the effects of genetic perturbations on BCHA production and tolerance. Introducing this biosensor in strains engineered with optogenetic circuits that control BCHA production with light will enable us to establish a closed-loop control system to study these metabolic pathways. This includes measuring transcriptomic changes in steady state or dynamic production systems. Ultimately, we will use these genomic and transcriptomic data to discover the key cellular networks involved in BCHA production and tolerance, which will be instrumental in developing better producing strains.

Branched-chain higher alcohols (BCHAs), including isobutanol and isopentanol, have been identified as key biofuels by the Office of Energy Efficiency & Renewable Energy of the U.S. Department of Energy[1]. These alcohols have better fuel properties than bioethanol, including higher energy density and better compatibility with current gasoline infrastructure. The yeast *Saccharomyces cerevisiae* is a preferred host organism for BCHA production because of its relatively high tolerance to their toxicity, and the potential to retrofit existing bioethanol plants (most of which use this yeast) with strains engineered to produce these advanced biofuels.

Existing efforts to commercialize these types of biofuels are challenged by limited productivities, as well as the toxicity that these alcohols have on strains engineered to produce them. Significant progress has been made in boosting yields and titers, particularly of isobutanol, through extensive metabolic and enzyme engineering based on detailed knowledge of branched chain amino acid metabolism and the structure and function of the enzymes involved[2,3]. In contrast, virtually nothing is known about the interplay between different cellular networks and BCHA production and tolerance, leaving a basic question unexplored: What are the key cellular networks that influence BCHA production?

Using our recently described genetically encoded biosensor for isobutanol production[4], we are currently carrying out a genome-wide analysis of isobutanol production in yeast. We have previously shown that genomic screens are invaluable to discover the role of cellular networks in complex traits, such as tolerance to isobutanol[5]. Our biosensor reports on the activity of isobutanol biosynthesis by expressing a fluorescence reporter (GFP), allowing us to quantify this metabolic activity. By optimizing a protocol to transform yeast genomic libraries[6], we have completed the construction of a new library, in which we introduced this isobutanol biosensor into the entire yeast gene deletion collection[7]. We kept this library arrayed in a 96-well plate format, which allows us

to measure how each gene deletion affects isobutanol production and gives us maximum flexibility to develop high-throughput assays.

We have validated this library by verifying the correlation between the biosensor output signal and isobutanol production in a subset of strains. We have also confirmed the effect of gene deletions known to affect isobutanol production[8,9]. We will present the results of initial screens, which revealed previously unknown gene deletions that increase production. The effects of some of these deletions are consistent with what is known about yeast metabolism and physiology. However, the effects of other deletions cannot be explained by known yeast biology, constituting new scientific discoveries and areas of future inquiry. We are currently in the process of completing our screens of the entire collection, which will provide quantitative measurements of the effect of each non-essential gene on isobutanol biosynthesis. This will allow us to build a full genetic interaction map relating cellular networks involved in isobutanol production. Altogether, this project will produce invaluable knowledge towards understanding and developing strains for improved isobutanol production.

References

1. Farrell, J., Holladay, J. & Wagner, R. *Fuel Blendstocks with the Potential to Optimize Future Gasoline Engine Performance: Identification of Five Chemical Families for Detailed Evaluation.*Technical Report. U.S. Department of Energy, Washington, DC. 2018. DOE/GO-102018-4970.
2. Hammer, S. K. & Avalos, J. L. Uncovering the role of branched-chain amino acid transaminases in Saccharomyces cerevisiae isobutanol biosynthesis. *Metab. Eng.* **44**, 302–312 (2017).
3. Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* **31**, 335–341 (2013).
4. Yanfei Zhang, Sarah K. Hammer, J. L. A. Genetically encoded biosensor for branched-chain alcohol production in S. cerevisiae to improve strain selection and enable pathway enzyme engineering [bioRxiv].
5. Kuroda, K. *et al.* Critical Roles of the Pentose Phosphate Pathway and GLN3 in Isobutanol-Specific Tolerance in Yeast. *Cell Syst.* (2019). doi:10.1016/j.cels.2019.10.006
6. Liu, G., Lanham, C., Buchan, J. R. & Kaplan, M. E. High-throughput transformation of Saccharomyces cerevisiae using liquid handling robots. *PLoS One* **12**, (2017).
7. Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science (80-.).* **285**, 901–906 (1999).
8. Wess, J., Brinek, M. & Boles, E. Improving isobutanol production with the yeast Saccharomyces cerevisiae by successively blocking competing metabolic pathways as well as ethanol and glycerol formation. *Biotechnol. Biofuels* **12**, 173 (2019).
9. Ida, K., Ishii, J., Matsuda, F., Kondo, T. & Kondo, A. Eliminating the isoleucine biosynthetic pathway to reduce competitive carbon outflow during isobutanol production by Saccharomyces cerevisiae. *Microb. Cell Fact.* **14**, (2015).

This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0019363.

Phenotyping for water use efficiency and related traits in C_4 grasses *Setaria* and *Sorghum*

Jennifer Barrett[1], Darshi Banan[3], Allen Hubbard[1], Shrikaar Kambhampati[1], Hui Jiang[1], Xiaoping Li[1], Rachel Paul[3], Charles Pignon[3], Parthiban Prakash[3], Erica Agnew[1], Jennifer Brophy[4], José Dinneny[4], Todd Mockler[1], Patick Ellsworth[2], Asaph B. Cousins[2], Andrew D.B. Leakey[3] and **Ivan Baxter[1]**

[1]Donald Danforth Plant Science Center, St. Louis, MO1
[2]Washington State University, Pullman, WA 99163
[3]University of Illinois at Urbana-Champaign, Urbana, IL 61801
[4]Stanford University, Palo Alto, CA

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.

url: www.foxmillet.org

Abstract: Plant growth and water use are interrelated processes influenced by the genetic control of both plant morphological and biochemical characteristics. Improving plant water use efficiency (WUE) to sustain growth in different environments is an important breeding objective that can improve crop yields and enhance agricultural sustainability. However, genetic improvements of WUE using traditional methods have proven difficult due to low throughput and environmental heterogeneity encountered in field settings. To overcome these limitations we have utilized a combination of high throughput approaches that measure physiological and biochemical properties of plants applied to populations of the C4 species *Setaria viridis* and *Sorghum Bicolor*. Across multiple experiments, we have combined greenhouse and field based methods to control water availability, and leveraged image based phenotyping of plant growth along with methods to visualize root crown initiation, stomatal patterning, leaf physiology, and biochemical measurements of metabolites, transcripts, carbon isotopes and elemental accumulation. By combining these approaches with quantitative genetics populations, we are able to get a more comprehensive understanding of the factors contributing to water use efficiency in these important species.

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

Title: Engineering enhanced photosynthesis and water use efficiency in Sorghum

Ryan Wessendor[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.

The high rates of photosynthesis and the carbon concentrating mechanism in C_4 plants is initiated by the enzyme phospho*enol*pyruvate carboxylase (PEPC). A decrease in the K_m for HCO_3^- (K_{HCO3}) has been proposed as a selective advantage for maintaining high rates of C_4 photosynthesis, particularly when carbon substrate availability is low due to restricted stomatal conductance under drought. We will present our work showing significant variability in K_{HCO3} among PEPC enzymes from different C_4 grass species. Additionally, our data provides new insight into the structural components responsible for K_{HCO3} that can be used to engineer increased photosynthetic efficiency in C_4 plants. This research builds the foundation for engineering a kinetically enhanced PEPC and the next step is using gene editing to enhance the kinetic properties of PEPC in Sorghum.

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 leave 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 characterization 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.

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

Title: The Development of Transgenic Lines and Improved Technologies for the Analysis of Photosynthetic and Water Use Efficiencies in Sorghum

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

Institutions: [1]Department of Cell and Molecular Biology, University of Rhode Island, 530 Liberty Lane, West Kingston RI 02892, [2]School of Biological Sciences, Washington State University, 367 Eastlick Hall Pullman WA 99164, [3]Department of Genetics, Cell Biology and Development, University of Minnesota, 1500 Gortner Avenue, Saint Paul, MN, [4] Institute for Genomic Biology, University of Illinois, 1206 West Gregory Drive, Urbana Il, 61801, [5]Donald Danforth Plant Science Center, 974 North Warson Road, Saint Louis, MO, 6313

Website: http://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 text: 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 harness this capability to help meet the Overall Project Goal to develop novel technologies to redesign the bioenergy feedstock *Sorghum bicolor* and to enhance its water use and photosynthetic efficiencies. We have established reliable protocols for the *Agrobacterium*-mediated introduction of experimental genetic constructs into sorghum cv BTx430, and deliver the viable transgenics required for the ongoing investigations with our collaborators on this project. We show the timeline of stable lines of sorghum that have been and are currently being produced to investigate the selected target genes for the analysis of photosynthetic and water use efficiencies. For example, 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 transformation method using MMRT technology with the potential to increase through-put and decrease time for our projects. We will be working with the Voytas lab that will allow MMRT to be more broadly utilized within the group. We also report the development of qPCR methods for the quantification of transgene insert copy number which further improves our capabilities for molecular analysis of the transgenic sorghum lines prior to shipment to our collaborators. Our program continues to support the central and essential aspects to provide the transgenic lines to investigate photosynthetic and water use efficiencies in sorghum.

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

Targeted Mutagenesis and Programmed Transcriptional Regulation in Setaria and Sorghum

Yang Liu,[1] Matt Zinselmeier, [1] Chunfang Wang, [1] Elena Gamo, [1] Colby Starker, [1] Albert Kausch,[2] Dan Voytas[1*] (voytas@umn.edu) and **Ivan Baxter**[3]

[1]University of Minnesota, St. Paul, MN; [2]University of Rhode Island, West Kingston, RI; [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 edit genes and manipulate gene expression *in vivo*. We are optimizing mutagenesis strategies using CRISPR/Cas and CRISPR/Cpf1 nucleases to achieve targeted gene knockouts, gene replacements and transgene insertions. Further, we are implementing base editor technology to achieve precise sequence changes without the need for a DNA double strand break. To achieve regulated gene expression, we are optimizing the use of programmable transcription factors (activators and repressors) derived from nuclease inactive dCas9 and dCpf1. The programmable transcription factors will be deployed in an innovative strategy for biocontainment of transgenes. To achieve genetic containment, we will identify genes (target genes) that compromise viability of *Sorghum bicolor* when overexpressed by the programmable transcription factors. We plan to introduce mutations into the target gene so that it is no longer recognized by the transcription factors. We will then combine all components of the synthetic circuit needed for genetic containment and test efficacy.

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

Mathematical modeling strategies to study the impact of drought on plant growth and metabolism in Setaria and sorghum.

Cheng Zhao[1], Renee Dale[2], Pascal Schläpfer[1], Elena Lazarus[1], Danny Ginzburg[1], Olivia MacDonald[1], Jennifer Barrett[2], Allen Hubbard[2], Shrikaar Kambhampati[2], Hui Jiang[2], Xiaoping Li[2], Erica Agnew[2], **Ivan Baxter[2]**, Seung Y. Rhee[1,*] (srhee@carnegiescience.edu)

[1]Carnegie Institution for Science, Stanford, CA; [2]Donald Danforth Plant Science Center, St. Louis, MO

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.

URL: www.foxmillet.org

Mathematical modeling is useful to determine how components function in a system to produce different phenotypes. We apply two different modeling strategies at two scales of plant growth in *Setaria* and sorghum. In the first modeling strategy, we abstract the process of the above ground *Setaria* plant tissue growth to develop a top-down trait based dynamical model. This model describes events that occur within the growth and development of above-ground tissue. The model is applied to a diversity panel of over 250 Setaria lines in drought and well-watered conditions. The parameters in our model are intended to describe growth and developmental decisions of the plant. The heritability of the parameter estimates were calculated for both wet and dry conditions. Heritable parameters will be used to identify loci that control processes critical to drought response.

In the second modeling strategy, we present a computational framework of multiscale metabolic modeling to investigate how plants allocate metabolic resources for biomass production in response to drought. The framework is centered on a cell type-specific genome-scale metabolic network model of *Sorghum bicolor,* generated from a generic sorghum model constrained by cell type-specific RNA-seq data. A C4 photosynthesis biochemical model was then integrated with the cell type-specific model to simulate dynamic environments by controlling carbon and energy sources of the metabolic network model. We collected a variety of physiological and omics data to inform the multiscale model, such as photosynthetic rates, metabolomic data, and RNA-seq data for sorghum under well-watered and drought conditions at multiple time points. Using the computational framework, we predicted that high light could alleviate biomass production induced by drought stress, which was validated by experimental testing. Moreover, a list of candidate genes were identified as potential targets to engineer for a higher biomass production.The trait based dynamical method provides a novel way to identify plant phenotypic traits for identifying new genes that control dynamic processes. This novel framework will be used in the future to understand if phenotypic variability may be emergent

from the interaction between environmental space searching strategies, biomass allocation strategies, and genotype. The multiscale metabolic model provides a novel platform for the analysis of inter-cellular metabolism with a dynamic constraint of carbon and energy exchange.

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 noDE-SC0018277.*

Systems analysis of the beneficial associations of sorghum with arbuscular mycrorrhizal fungi studied with genetics, genomics, imaging and microbiomics

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

[1]University of Georgia, [2]Northern Arizona University

Project Goals: The objective of this project is to discover the sorghum genes that recruit and sustain the root microbial community, especially the AMF community, that promotes high sorghum biomass yield under a known range of environmental conditions, and to build/test predictive models derived from these insights. These studies are designed to discover the interactions between genetic and environmental factors that cause the variable performance of mycorrhizal symbioses in field conditions, using a series of field and greenhouse experiments with genetically diverse sorghum lines in Georgia and Arizona. Through six integrated sub-projects, we will collect tens of thousands of phenotypic measurements, including root morphology, AMF morphology, AMF taxa abundance, other microbial abundances, plant root gene expression, AMF gene expression, plant mineral content and biomass yield. Promising candidate associations of these phenotypes with each other and with the sorghum genotype will be identified under a multiple testing scenario and then confirmed by independent experiments to find validated causal links (metadata) for increased biomass production.

Abstract: The roots of most plants form symbioses with arbuscular mycorrhizal fungi (AMF), which can enhance nutrient uptake, drought tolerance and disease resistance for host plants. Harnessing these symbioses could reduce irrigation and fertilizer inputs for more sustainable production of biofuel crops like sorghum. Despite decades of effort, successful use of AMF as bioinoculants has been elusive, largely because mycorrhizae are complex systems and their effects are context dependent. The positive contributions of AMF and other microbes to crop productivity have been broadly documented, but the same microbes can vary in their contributions, from highly beneficial to functionally parasitic, depending on the field environment. Though the contributions of some plant genes in plant-AMF interactions have been studied in controlled laboratory conditions, none of these interactions are well understood in field conditions, and it is likely that whole categories of interactions are completely unknown at this stage. The project teams' laboratories have recently identified plant genome segments that determine which AMF and other microbial taxa are abundant in the root/soil environment. This suggests that sorghum genetics can be a key to unravelling the plant-microbe-environment system and optimizing plant-microbe interactions for yield. These predictions will be more than just biomass yield associated with particular sorghum genotypes, AMF communities and input regimens, but also predictions of the other phenotypes that will be commensurate. These predictions will be tested both in small-scale controlled-environment experiments, and in analyses of agricultural field data produced during the project but not used in the initial model building. These data and metadata associations will be utilized to create systems models that predict in detail the performance of novel sorghum genotypes under a variety of environmental conditions. Mutagenized sorghum will also be employed to test the impact of specific sorghum genes on AMF type, abundance and functions. Confirmations of predicted associations will have tremendous value in uncovering the mechanisms that determine beneficial plant-AMF-environment interactions. Moreover, the iteratively improved models will not only have exceptional impact for inspiring and guiding future experimental design, but also for applied use by farmers to guarantee optimal sorghum biomass yield by using beneficial microbial associations that require only the appropriate sorghum genotype and agricultural inputs, not any artificial microbial inoculum.

Revealing Metabolic Exchange and Optimizing Carbon Transformation in Co-Culture for Applications to Sustainable Biosynthesis

Pavlo Bohutskyi,[1] Jackson Jenkins[2]* (jjenki71@jhu.edu), Foteini Davrazou,[3] Jon Magnuson,[1] Michael Guarnieri,[3] and **Michael Betenbaugh**[2]

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Johns Hopkins University, Baltimore, MD; [3]National Renewable Energy Lab, Golden, CO

Project Goals: The goal of this project is to develop and characterize synthetic lichen communities of autotrophic and heterotrophic microbes as a novel sustainable symbiotic platform for the production of biofuels and biochemicals. Carbon-fixing autotrophs provide oxygen, organic carbon substrates and other metabolites to their heterotrophic neighbors, which in turn generate carbon dioxide and diverse interemediate compounds. By optimizing and enhancing these interactions, we can create a robust, sustainable and more efficient synthetic lichen community to transform carbon dioxide into bioproducts. Multi-omics driven genetic engineering will enhance metabolite exchange and product generation capabilities with the microbial co-culture.

Lichens are communities that collect sunlight and carbon dioxide and apply it to power the group's activities. They potentially represent a novel biotechnology platform that can transform CO_2 and sunlight into valuable energy-related biochemicals, eliminating the need for costly substrate feeding. Unfortunately, natural lichens have slow growth rates, making them impractical for most industrial applications. In this project, our goal is to enhance the exchange of metabolites between autotrophs and heterotrophs, creating superior synthetic lichens able to generate useful products of interest to the energy and chemical industries. Understanding the metabolite exchange between photoautotrophic and heterotrophic members of co-cultures is a complex but crucial task in optimizing a synthetic lichen for more efficient bioproduction. Two major challenges include recognizing the species origin of secreted metabolites and uncovering key rapidly-consumed exchange metabolites that may only be present in trace concentrations. To overcome these challenges, we implemented a membrane-separated photobioreactor (mPBR) allowing co-cultivation of photoautotrophic and heterotrophic microbes and capturing of critical exchanged metabolites. Profiling of extracellular metabolome in different compartments of the mPBR provided insights into the origin of exchange metabolites and direction of their flux, and suggested exchanged compounds crucial for a mutualistic microbial relationship. In addition, to explore the potential of improved carbon utilization in heterotrophs, our team has engineered an industry-relevant yeast *Yarrowia lipolytica* strain to grow on sucrose as the only carbon source by constitutively expressing invertase, the enzyme responsible for sucrose hydrolysis. Interstingly, invertase expression improved the utilization of several other carbon substrates compared to the wild-type strain. Improving the range of consumable substrate in heterotrophic co-culture partners opens the potential for more resilient and flexible synthetic lichens allowing generation of a wide array of biochemicals.

Supported under award DE-SC0019388, Biological and Environmental Research

Sustainable production of biofuels by consortia of *Synechococcus elongatus* and *Aspergillus* species.

Cristal Zuniga (crzuniga@eng.ucsd.edu)[1*], Bo Wang (bo.wang.2@Vanderbilt.Edu)[2*], Jamey Young [2], and Karsten Zengler[1], **Michael Betenbaugh**[3]

[1] University of California San Diego, La Jolla, CA. [2] Vanderbilt University, Nashville, TN. [3] Johns Hopkins University, Baltimore, Maryland

The goal of this project is to combine phototrophic and heterotrophic microorganisms as a novel platform for the sustainable production of biofuel and its precursors. Here we study the consortia of *Synechococcus elongatus* cscB$^+$ and *Aspergillus* species. *S. elongatus* cscB$^+$ is providing sucrose and oxygen to the heterotrophic *Aspergillus* species, and in exchange, the fungi are producing CO_2 for *S. elongatus* for carbon fixation via photosynthesis. Synthetic microbial communities of cyanobacterium-fungus pairs were evaluated for productivity through genome-scale metabolic modeling. For this, we manually curated and updated transport capabilities in the model of the cyanobacterium *S. elongatus* PCC 7942 (*i*JB792) and reconstructed new metabolic models for two *Aspergillus* strains, *i.e. A. nidulans* and *A. niger*. Subsequently, two community metabolic models were created by pairing the cyanobacterium model with the new *Aspergillus* models.

The experimental success of working with these communities relies on the characterization of sucrose secretion by the phototroph. For this, we performed ^{13}C-MFA under photoautotrophic conditions for three *S. elongatus* cultures: wild type, wild type + NaCl, and cscB$^+$ + NaCl. ^{13}C-labelled bicarbonate served as tracer to quantify labeling of more than a dozen intracellular metabolites over time. We identified significant changes in intracellular metabolite pool sizes and metabolic fluxes of these three cultures, unveiling that expanding the sucrose sink via overexpressing the *cscB* gene increases carbon fixation in *S. elongatus*. Using the metabolic model for *S. elongatus* we evaluated the solutions space of the three cultures using random sampling. We found that sucrose secretion is linked to a change in the flux through the photosystem, reducing photon absorption. These energetic changes augmented asparagine hydrolysis and decreased glutathione, glutamate, and glycine synthesis. We reconstructed the models of *A. nidulans* (*i*ANid1230) and *A. niger* (*i*ANig1153) using semi-automated algorithms. These knowledge base tools were extensively validated using high-throughput phenotypic growth data on 190 carbon and 95 nitrogen sources for each species.

The highly curated metabolic models of *S. elongatus, A. nidulans,* and *A. niger* enabled the development of two community metabolic models (CM-models). CM-models were integrated using the shared metabolite pool (SMP) approach, which includes metabolites that can potentially be exchanged by each synthetic community. Potential exchanges were refined by high-throughput phenotypic data. Co-cultures were characterized by a) predicting growth rates and population proportions (constraints-based choice to achieve experimental growth rates), b) determining metabolic interactions (theoretical interchange of metabolites, SMP analysis), c) co-culture medium optimization (robustness analysis), and d) syntrophic pathway inclusion (metadata contextualization). The reconstructed CM-models helped elucidating syntrophic interactions between community members. We furthermore deployed the CM-models to evaluate the capability of the consortia to produce organic chemicals and biofuels. Growth and flux distribution predictions will be validated by physiological data, as well as untargeted metabolomics, and gene expression.

Supported under award DE-SC0019388, Biological and Environmental Research

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], Ko-Hsuan Chen[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, which control the biogeochemistry of forest soils.

https://genomicscience.energy.gov/carboncycle/2019Awards.shtml

Interactions between soil microbes can drastically alter ecosystem processes both above and belowground, but the mechanisms by which these microbes interact, and their impacts on soil biogeochemistry remain elusive and difficult to parameterize in existing ecosystem models. We aimed to characterize interactions between coniferous plants, their major root fungal symbionts (ectomycorrhizal fungi, EMF), and free-living saprotrophic decomposers (SAPs) in soil. We performed a greenhouse-based synthetic ecosystem experiment with *Pinus taeda* seedlings growing with and without their EMF symbiont (*Suillus cothurnatus*), under high and low levels of soil carbon (C), soil nitrogen (N), and plant C (ambient vs. elevated carbon dioxide -CO_2). We expected that under low soil C, EMF prime decomposer activity and increase the release of soil C as CO_2, while under high soil C, EMF slow decomposition and reduce soil CO_2 release, with EMF competing with SAPs for access to soil organic matter (i.e. the Gadgil effect). These processes would be exacerbated under high plant C availability to EMF, but suppressed under high soil N. We found that EMF prime decay of soil organic matter under low soil C, but slow decay under high soil C. Elevated soil N suppressed the EMF effect on soil C-derived CO_2 losses. Elevated CO_2 might increase plant-EMF uptake of soil N by outcompeting SAPs, but total N uptake may depend on soil C availability. Together, our results suggest that the direction of EMF-SAP interactions is highly dependent on soil C and N availability to SAPs and might change according to plant C level.

Funding statement: The authors acknowledge DOE BER awards DE-SC0020403, AN_360581, and JGI/EMSL FICUS 51536.

Delivering the goods: leveraging functional genomics to understand cofactor trafficking

Miriam Pasquini[1], Nicolas Grosjean[1], Lifang Zhang[2], Meng Xie[1], Doreen Ware[2], **Crysten E. Blaby-Haas (cblaby@bnl.gov)***[1]

[1]Biology Department, Brookhaven National Laboratory, Upton, NY.
[2]Cold Spring Harbor Laboratory, Cold Spring Harbor, 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.

Transition metals occupy numerous and often essential positions within the biochemical framework of the cell. 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.

By leveraging phylogenomic and data-mining analyses combined with an inter-disciplinary experimental approach, we have discovered a novel transferase 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, where duplication has resulted in analogous pathways in the cytosol and chloroplast. Based on biochemical and phenotype analysis, this ancient and conserved mechanism is required during Zn insufficiency to ensure protein translation fidelity.

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.

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

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 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.

Up to 90% of the iron found in leaves is located in the chloroplasts, where every membrane-spanning photosynthetic complex has an absolute requirement for iron cofactors, such as heme. Specialized biogenesis pathways involved in heme trafficking and insertion have been described specifically for cytochrome *b*- and *c*-type hemoproteins. These pathways exist to ensure fidelity of cofactor synthesis and limit potential oxidative stress caused by free heme. However, the existence of a generalized heme chaperone that can interact with the labile heme pool in plant cells to protect and deliver heme has yet to be identified.

Using a phylogenomic approach, we identified a large protein family consisting of uncharacterized or putative heme-binding proteins. Our analysis suggested that distinct members of this family have evolved discrete functions as heme-sensing regulators, heme oxidases, and heme chaperones. Specifically, we identified three distinct, but related, subfamilies of phototroph-specific homologs. The first subfamily includes the previously characterized AtGluTRBP that binds to GluTR and plays a pivotal role in heme biosynthesis regulation. The second subfamily, which we predict contains novel heme oxidases, is composed of uncharacterized plant and algal proteins, and the third subfamily, which we predict contains novel heme chaperones, is composed of uncharacterized cyanobacterial proteins.

To test these computationally derived hypotheses, we purified protein homologs from the green alga, *Chlamydomonas reinhardtii*, the bioenergy feedstock *Populus trichocarpa*, and the cyanobacterium, *Synechocystis* sp. PCC 6803. We demonstrated that the algal and land plant proteins can bind and degrade heme *in vitro*, suggesting that these proteins, which are localized in the chloroplast, present a new family of plant heme oxidases. In contrast, the cyanobacterial homolog can bind but not degrade heme in the presence of an exogenous electron donor, suggesting that this subfamily may function as either heme storage or heme chaperones.

Determination of crystal structures of the cyanobacterial homolog in the presence and absence of heme revealed unprecedented features. In the presence of heme, the protein forms a dimer where the heme is saddled by two zinc ions. By analogy with the protective axial histidines found in a bacterial heme transporter, we propose that the zinc saddle found here protects heme from oxidation. By homology modelling, this structure also helps to understand the structure-function relationship of plant homologs where this protective function is not conserved. The discovery of this family of novel heme oxidases and putative heme chaperones provides new 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.

Optimizing hydroponic growth system for metal stress studies of bioenergy crops

Meng Xie (mxie@bnl.gov)[1*], Jeremiah Anderson[1], Nicholas Gladman[2], Michael Regulski[2], Doreen Ware[2], Timothy Paape[1], and **Crysten E. Blaby-Haas**[1]

[1]Biology Department, Brookhaven National Laboratory, Upton, NY.
[2]Cold Spring Harbor Laboratory, Cold Spring Harbor, 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.

Populus is one of the primary bioenergy crops that DOE has been focusing on for decades. With the state-of-the-art genomics resources, *Populus* has emerged as an ideal model system to study woody perennial plants. There is general agreement that the study of *Populus* is required to understand the specialized physiology of woody perennials, especially adaptation to environmental stresses. The transition metals iron (Fe) and zinc (Zn) are indispensable cofactors for numerous critical aspects of plant growth, including metabolism, signaling, gene expression, genome stability, and the assimilation of other nutrients. However, in excess, these metals can be cytotoxic for plants. Imbalanced Fe and Zn concentrations in soils can cause physiological stress es that negatively impact crop health and yield. How woody perennials deal with Fe or Zn stress (deficiency and excess) is poorly understood. We are combining transcriptomics, ionomics, and non-destructive physiological measurements in a time-series study of *Populus* treated by various Zn and Fe stresses to characterize the genome-wide response of *Populus* towards transition metal stresses. Our studies will enhance the understanding of the specialized physiology of long-lived perennial plants and the genetic improvement of *Populus* as a cost-effective sustainable biomass feedstock.

To perform the time-series study, we have optimized a hydroponic system for *Populus* growth that allows us to precisely manipulate Fe and Zn bioavailability. Using this versatile platform, we observed stress phenotypes within two weeks. Additionally, the hydroponic system is inexpensive to build and can easily be adapted for other bioenergy crops like sorghum.

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.

Contribution of Serine Biosynthesis and Degradation to Carbon and Nitrogen Metabolism During Salinity stress in Poplar.

Maria Del Mar Rubio Wilhelmi[1*](mmrubiowilhelmi@ucdavis.edu), Vimal K. Balasubramanian[2], Samuel Purvine[2,] Ying Zhu[2], Sarah Williams[2], Yongil Yang[3], Amir H. Ahkami[2], Jonathan Cumming [4], Stephen DiFazio [5] Neal Stewart, Jr.[3] and **Eduardo Blumwald[1]**

[1]Dept. of Plant Sciences, University of California, Davis, CA; [2]Pacific Northwest National Laboratory (PNNL), Richland, WA; [3]Dept. of Plant Sciences, University of Tennessee, Knoxville, TN; [4]Dept of Natural Resources, Univ. Maryland Eastern Shore; MD; [5]Dept. of Biology, West Virginia University, Morgantown, WV;

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 (1) identify stress-responsive genes and proteins in specific cell-types of poplar leaves and roots; (2) discover novel *cis*-regulatory elements; (3) construct stress-responsive synthetic promoters; and (4) use these promoter-gene fusions to develop abiotic stress-tolerant poplar. The transgenic poplar trees will be evaluated under both controlled and field conditions.

Plant responses to environmental stress are dynamic and involve complex cross-talk between different regulatory pathways[1], including metabolic adjustments and gene/protein expression at the cellular level for physiological and morphological adaptation at the whole-plant level[2]. However, each cell-type in plant tissues is defined by specific transcriptional, protein, and metabolic profiles that determine its function and response(s) to stress[2]. Thus, determining the plant responses to environmental changes requires the study of the cell/molecular properties of specific single cell-types within a tissue to effectively reveal the underlying mechanisms regulating developmental processes and plasticity under suboptimal conditions.

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 salinity stress was 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, including leaf mesophyll, xylem/phloem using cryo-sectioning and laser-capture microdissection (LCM) techniques. Plants were exposed progressively to 50 mM, 100 mM, and 150 mM NaCl for 10 days (Early Salt Stress, ESS) and maintained at 150 mM NaCl for another 10 days (Late Salt Stress, LSS), followed by a period of recovery from stress where the plants returned to control growth conditions (Recovery, R). Leaf palisade and vascular cell types were collected and processed using nanoPOTS (nanodroplet processing in one-pot for trace samples) platform, which includes a nanoliter-scale liquid handling robotic station in which the cells were lysed for protein extraction[3]. The proteins were then alkylated, digested, and the peptide samples were loaded onto nano-LC coupled to a Tribrid Lumos Mass spectrometer. The collected peptide abundance values were filtered, normalized, and converted to protein abundance values and used to limit reliable detection (LOD) and Z-score calculation.

Proteins detected at least in either control or stress conditions were used to calculate total Palisade and vascular identified proteins. The relative abundance values (salt vs. control) of

identified proteins were used for Z-score assessment analysis, and proteins altered under stress (+2<Z-SCORE<-2) were used to calculate %cell type unique and shared proteins. A higher percentage of proteins significantly altered under stress were found to be explicitly unique to both Palisade (ESS-44.5%, LSS-34.6%, R-34.2%) and vascular (ESS-49.4%, LSS-55.2%, R-60.2%), suggesting that cell type unique proteins might regulate the leaf proteomic responses to salt stress and recovery.

Our results showed that a higher number of proteins associated with the 'phosphorylated pathway' for Serine (Ser) biosynthesis and Ser degradation accumulated in the vascular tissue at LSS. The production of Ser from 3-P-glycerate in plastids, together with glycine (Gly) catabolism in the mitochondria, provide cytosol with Ser and generating intermediaries that play important roles in pathways associated with energy production, carbon/nitrogen (C/N) balance, and lignification. Under stress conditions, 2-OXG may be used in the TCA cycle, providing energy supply in the vascular tissue for the increasing demand brought about the stress and contributing to the transport and reallocation of resources to sink tissues. 2-OXG is associated with ammonia re-assimilation, supplying the carbon skeleton needed by the GS/GOGAT cycle. Glycine Decarboxylase (GDC) carries Gly catabolism in the mitochondria yielding Ser. This reaction also produces high amounts of ammonia that are re-assimilated to amino acids (via Asparagine synthetase (AS) and Carbamoyl phosphate synthase (CPS)), to avoid toxicity. Ser catabolism to Gly, mediated by Ser hydroxymethyl transferase (SHMT) and the recycling of Gly to Ser in the mitochondria by GDC, are the main Carbon source of C-1 metabolism in plants. Thus, Ser is crucial for tetrahydrofolate (THF) metabolism, providing methyl group donors through S-adenosylmethionine (SAM) cycle, an important process for synthesizing lignin in vascular tissues.

Phosphorylated and non-phosphorylated pathways of Ser biosynthesis are important processes linking Carbon and Nitrogen metabolism, maintaining energy levels under stress conditions[4]. Salinity stress-induced changes in protein levels of these pathways in vascular and Palisade tissues will be presented and discussed.

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. Igamberdiev, A.U.; Kleckowski, L.A. (2018). The glycerate and phosphorylated pathways of serine synthesis in Plants: The branches of plant glycolysis linking Carbon and Nitrogen metabolism. *Frontiers in Plant Science,* 9:318.

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

Rational design and testing of abiotic stress inducible synthetic promoters from poplar *cis*-regulatory elements

Yongil Yang[1,2 *](yiyang98@utk.edu), Jun Hyung Lee[1,2], Magen R. Poindexter[1,2], Yuanhua Shao[1,2], Wusheng Liu[2,3], Scott C. Lenaghan[1,4], Amir H. Ahkami[5], Stephen DiFazio[6], Jonathan Cummings[6], **Eduardo Blumwald**[7], and C. Neal Stewart, Jr.[1,2]

[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

The project goal is to attain robust biomass of trees under abiotic environmental stress conditions via genetic engineering. Abiotic stress-resistance genes, especially to water deficit-, salt-, and temperature-stress, need to be under inducible regulatory control. We are designing, building, and testing stress-inducible synthetic plant promoters to drive resistance genes. –Omics data are used to discover *cis*-regulatory DNA motifs that may be used to construct synthetic promoters. The synthetic promoters are then tested under appropriate stimuli in engineered plants to use toward development of environmentally-robust poplar.

Abstract text:

Abiotic stress resistance traits may be especially crucial for sustainable production of bioenergy tree crops. Here we show the performance of a set of rationally designed osmotic-related and salt stress-inducible synthetic promoters for use in hybrid poplar. *De novo* motif-detecting algorithms yielded 30 water-deficit- (SD) and 34 salt-stress (SS) candidate DNA motifs from relevant poplar transcriptomes. We selected three conserved water-deficit-stress motifs (SD18, SD13, and SD9) found in 16 co-expressed gene promoters, and we discovered a well-conserved motif for salt-response (SS16). We characterized several native poplar stress-inducible promoters to enable comparisons with our synthetic promoters. Fifteen synthetic promoters were designed using various SD and SS subdomains, in which heptameric repeats of five-to-eight subdomain bases were fused to a common core promoter downstream, which, in turn, drove a green fluorescent protein (GFP) gene for reporter assays. These 15 synthetic promoters were screened by transient expression assays in poplar leaf mesophyll protoplasts and agroinfiltrated *Nicotiana benthamiana* leaves under osmotic stress conditions. Twelve synthetic promoters were induced in transient expression assays with a GFP readout. Of these, five promoters (SD18-1, SD9-2, SS16-1, SS16-2, and SS16-3) endowed higher inducibility under osmotic stress conditions than native promoters. These five synthetic promoters were stably-transformed into *Arabidopsis thaliana* to study inducibility in whole plants. Herein, SD18-1 and SD9-2 were induced by water-deficit stress, whereas SS16-1, SS16-2, and SS16-3 were induced by salt stress. The synthetic biology design pipeline resulted in five synthetic promoters that outperformed endogenous promoters in transgenic plants.

Funding statement:

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018347.

Title: Multiscale approaches characterize microbial metabolic feedbacks to hydrological perturbation

Authors: Stephany S. Chacon[1], Aizah Khurram[1], Markus Bill[1], Hans A. Bechtel[2], Jana Voriskova[1], Liang Chen[1], Lee H. Dietterich[3], Ulas Karaoz[1], Hoi-Ying N. Holman[1,4], Daniela F. Cusack[3,5], **Nicholas Bouskill[1,*]** (njbouskill@lbl.gov)

Institutions: [1]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA. [2]Advanced Light Source Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA. [3]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, 80523, USA. [4]Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA. [5]Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama

Website: njbouskill.com/research

Project Goals: The microbial feedbacks project seeks to characterize and predict microbial (fungal, bacterial, and archaeal), and viral responses to rapid environmental fluctuations. The project examines responses across multiple scales, from analysis of metabolic response to environmental change at the single-cell level to whole community responses to drought along precipitation gradients in a humid tropical forest. This approach integrates multiple streams of data, including multi-omic datasets, infrared spectroscopy, isotope geochemistry, and gas flux data with an end goal of linking soil biogeochemistry, and organic matter composition to the traits and trade-offs selected for during drought.

Abstract: Model projections predict that climate change impacts on the tropics will include an increased frequency of drought and precipitation cycles. Such environmental fluctuations at the soil pore scale play an important role in shaping microbial adaptive capacity, and trait composition of a community, which feeds back on to the breakdown and formation of soil organic matter (SOM). Humid tropical forest soils contain vast soil carbon stocks due to high productivity. Therefore, developing a predictive understanding of the factors controlling soil carbon balance remains a social imperative. Improved characterization of microbial feedbacks to the composition and stability of SOM pools is critical. Herein, we examine the microbial response to drought perturbations across 3 different, but complementary scales. At the largest scale, we explored the impacts of drought across a 1 m precipitation gradient spanning four sites from the Caribbean coast to the interior of Panama. At each site, 4 throughfall exclusion plots (10 x 10 m) were constructed to reduce precipitation by 50 %. We sampled these exclusion plots, and the corresponding controls, 18-months into the treatment to characterize the traits selected for across the gradient and under throughfall exclusion. To increase the signal: noise ratio intact cores from one site were taken at mid-rainfall from infertile soils and subject to 3 different hydrological treatments (control, drought, rewetting-drying cycles) in the laboratory over a 5-month period. For the field and meso-scale experiments, we evaluated changes imparted by hydrological perturbations using multi-omic approaches, and physico-chemical measurements. Finally, to improve our holistic understanding of traits expressed by microorganisms under either osmotic or matric stress (both characteristic of drought), we reduced the complexity again, by isolating a range of gram-positive and negative bacteria and subjected them to acute stress at the scale of the single-cell and simple communities. Single-cells

were subjected to osmotic or matric stress and short-term physiological responses determined using non-destructive synchrotron radiation-based Fourier Transform-Infrared spectromicroscopy. Through this approach we identified changes in metabolic allocation within different cells, in particular to the secondary metabolome of the different bacteria. Our contribution will discuss the outcomes of these multi-scale experiments. Specifically focusing on how shifts in the microbial community and physiological changes may influence tropical soil carbon stability under future scenarios of altered drought and precipitation cycles.

Advanced Metabolic Modeling of Diurnal Growth in *Chlamydomonas reinhardtii*

Alexander J. Metcalf and **Nanette Boyle**

Chemical and Biological Engineering, Colorado School of Mines, Golden, CO 80401, USA

Project Goals: The main goal of this project is to develop more predictive metabolic models of diurnal growth for algal systems. We are using a multi-paradigm multi-scale approach which enables us to include phenomenon not previously integrated into metabolic models, such as diel light, diffusion of metabolites/nutrients, cell-cell interactions, as well as temporal and spatial tracking of cells. This model will further be enhanced with experimental data collected over 24-hour diel growth for transcript abundance and changes in biomass composition. Validation and improvement of the model will be performed by comparing predictions to ^{13}C-MFA of cells grown in the lab as well as in large outdoor ponds.

Photosynthetic microorganisms have the potential to become economical and sustainable sources of fuels, as the energy required for the cell to grow can be sourced from natural sunlight alone; however, we have yet to harness their full power due to a general lack of tools for engineering their metabolism. Metabolic models have been shown to drastically reduce the development time for commercial production strains of heterotrophic bacteria; however, these models are less applicable to photosynthetic systems due to the transient nature of diurnal (day/night) growth. Current metabolic models are not capable of accurately predicting growth rates in day/night growth cycles, let alone genetic changes which would lead to increased yields. Our work is focused on constructing an approach to diurnal modeling that allows for extension of current metabolic models into a transient space, using organism specific circadian information. We have used circadian gene expression data from *Chlamydomonas reinhardtii* to cluster gene expression and convert discrete data into continuous functions. We then implemented these functions as additional constraints on our metabolic model and are currently integrating this constrained model into an agent-based framework. We will present the result of this work and demonstrate how these modeling techniques are able to further improve the model and better predict growth in diurnal light. Ultimately, the availability of such models will introduce a new frontier in the ability to use *in silico* tools to investigate the metabolism, growth and phenotype of photosynthetic microorganisms. It will enable us to gain insight into why photosynthetic organisms have drastically different productivities when grown in continuous light compared to diurnal cycles and how to circumvent this.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0019171

Title: Functional and comparative genomics of pyrophilous fungi

Authors: Andrei S. Steindorff[1*], Kyungyong Seong[1,2], Akiko Carver[1,2], Sara Calhoun[1], Monika Fischer[2], Kyra Stillman[2], Haowen Liu[2], Anna Lipzen[1], Guifen He[1], Mi Yan[1], Bill Andreopoulos[1], Jasmyn Pangilinan[1], Kurt LaButti[1], Vivian Ng[1], Matthew Traxler[2], **Thomas D. Bruns**[2], **Igor V. Grigoriev**[1,2]

Institutions: [1]US DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA
[2]Plant and Microbial Biology, UC Berkeley, Berkeley, CA

Website: https://mycocosm.jgi.doe.gov/

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 pyrolyzed organic matter, 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: Understanding post-fire soil systems are essential because they have significant direct and indirect effects on global carbon storage. Fires result in a large amount of carbon that remains resident on the site as dead and partially "pyrolyzed" (i.e., burnt under low oxygen) material with long residency times and constitutes a significant pool in fire-prone ecosystems. Besides, fire-induced hydrophobic soil layers, caused by condensation of pyrolyzed waxes and lipids, increase post-fire erosion and lead to long-term productivity losses. Soil microbes are likely involved in the degradation of all these compounds, yet little is currently known about the organisms or metabolic processes involved. So far, we sequenced and annotated four pyrophilous Basidiomycetes and seven Ascomycetes genomes. In our previous work on Basidiomycetes fungi (Steindorff et al., 2021), we found expansion of genes potentially involved in the degradation of the hydrophobic layer, pyrolyzed organic matter, and mushroom formation. In this work, we focused on the seven ascomycetes genomes and compared them with other 12 non-pyrophilous in the same order and also with 124 genomes at a larger scale, including pyrophilous Basidiomycetes and other organisms with heterogenous lifestyles. Additionally, we explored enriched Pfam domains and CAZymes to identify patterns associated with these organisms' 'charcoal-loving' lifestyle. Our analyses uncovered gene families related to degradation of pyrolyzed organic matter, but these gene families were distinct from those expanded in the pyrophilous fungi in Basidiomycota. . The enrichment analysis revealed families like peritrophin-A, arthropod defensin, aminopeptidases, beta-glucosidase, heat shock proteins,

and fungal fucose-specific lectin. These families might be involved with the phyrophilous fungi' capacity to survive in a toxic environment like post-fire soil. We found a CAZyme CBM14 expanded exclusively in the Pyronemataceae family. This family is mainly found in metazoans, and in fungi, it is only found in some Eurotiomycetes. Since it's a chitin-binding domain, this suggests that secreted CBM14 domain proteins might protect the fungus from microbial attacks in its soil habitat. Another interesting finding is that pyrophilous fungi have larger proteins than non-pyrophilous, being in an intermediate state to thermophiles. Pyrophilous fungi are commonly found fruiting after fire events, therefore, passing through their sexual stages in this process. To make an in-depth comparison of these conditions, we analyzed the available transcriptomic data of *Pyronema domesticum* grown in charcoal and during sexual development. We performed a co-expression network analysis and found two modules with the most differentially expressed genes in charcoal and sexual development. Gene Ontology categories like chitin/carbohydrate/lipid/superoxide metabolism and transport were found in both modules, showing that such processes are likely required to grow in the presence of charcoal and sexual development. The transcription factor STE12, known to be involved in sexual development in yeast and filamentous ascomycetes, was up-regulated in both conditions when compared with their respective controls. This study will improve our understanding of this unique lifestyle of pyrophilous fungi and their role in post-fire carbon cycling.

References/Publications

Steindorff AS, Carver A, Calhoun S, Stillman K, Liu H, Lipzen A, He G, Yan M, Pangilinan J, LaButti K, Ng V, Bruns TD, Grigoriev IV. Comparative genomics of pyrophilous fungi reveals a link between fire events and developmental genes. *Environmental Microbiology*. 2021 Jan;23(1):99-109.

Funding statement:

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. 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: Using Metagenomic Stable Isotope Probing to Identify Genomic Signatures of Bacterial Life History Strategies

Authors: Samuel E. Barnett[1]* and Daniel H. Buckley[1]

Institutions: [1]Cornell University, Ithaca, NY

Project Goals: The goal of this project was to identify genomic signatures of life history strategies of soil bacteria that influence their activity in soil carbon cycling.

Abstract text: Life history strategies are important determinants of microbial activity in soil carbon cycling. Life history strategies are defined by tradeoffs in energy allocation to population growth, nutrient acquisition, and cellular maintenance. In microbes, these tradeoffs influence the rate and efficiency by which cells utilize carbon substrates as well as the fate of that carbon. Measuring life history strategies *in situ* is difficult due to complex microbe-microbe and microbe-environment interactions, though metagenomics may provide a solution. We hypothesized that genomic signatures of life history strategies exist in soil bacteria and that these signatures relate to bacterial activity in soil carbon cycling. We used metagenomic-stable isotope probing (metagenomic-SIP) to link genes to patterns of carbon assimilation from diverse ^{13}C-labeled organic matter substrates over time. These carbon assimilation dynamics have previously been linked to life history strategies along the copiotroph-oligotroph life history continuum. This gene to function linkage allowed us to examine 11 genomic signatures predicted to vary along this copiotroph-oligotroph continuum. Out of these 11 signatures, we found that the abundance of transcription factor genes and secretion signal peptide encoding genes in the genomes explained significant variation in genome position along the copiotroph-oligotroph continuum. The other 9 genomic features evaluated were not significantly correlated with the copiotroph-oligotroph continuum but may distinguish bacterial lifestyles within these strategies. Our results demonstrate that genomic signatures can be used to distinguish life history strategies of soil bacteria active in soil carbon cycling.

Funding statement: This project was funded Supported by the Department of Energy, Office of Biological & Environmental Research Genomic Science Program, award numbers DE-SC0010558 and DE-SC0004486 and from the Joint Genome Institute Community Science Program project number 503502

Spatiotemporal mapping of the origin of linolenic acid in signaling transduction cascades during leaf deconstruction

Dušan Veličković[1], Rosalie K. Chu[1], Corinna Henkel[2], Annika Nyhuis[2], Nannan Tao[3], Lily Khadempour[4], Jennifer E. Kyle[1], Bobbie-Jo M. Webb-Robertson[1], Joshua N. Adkins[1], Christopher R. Anderton[1], Carrie D. Nicora[1], Vanessa Paurus[1], Kent Bloodsworth[1], Mary S. Lipton[1], Cameron R. Currie[4], Lisa M. Bramer[1], Dale S. Cornett[5], Wayne R. Curtis[6], **Kristin E. Burnum-Johnson**[1] (Kristin.Burnum-Johnson@pnnl.gov)

[1]Pacific Northwest National Laboratory, Richland, WA, USA; [2]Bruker Daltonik GmbH, Bremen, Germany; [3]Bruker Daltonics, San Jose, CA, USA; [4]University of Wisconsin-Madison, Madison, WI, USA; [5]Bruker Daltonics, Billerica, MA, USA; [6]Chemical Engineering Department, Penn State University, PA, USA

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 leaf-cutter 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.

Naturally evolved microbial systems that are capable of efficient deconstruction of plant cell wall biomass exist. Biomass deconstruction in these natural communities is often dependent on bacterial-fungal symbiosis, yet the molecular underpinnings of these interactions are poorly understood. An excellent example of such a system is the leaf-cutter ant fungal garden ecosystem, which employs inter-kingdom interactions to liberate energy rich carbohydrates from plant lignocellulose biomass. Unfortunately, the microbial community dynamics of the leaf-cutter ant fungal garden ecosystem are a challenge to assess because of the high heterogeneity of species composition and phenotype occurring across space and time during plant biomass deconstruction.

To understand how the fungal garden is able to degrade plant matter with such efficiency, it is necessary to study the metabolic interactions and biochemical pathways utilized by its microorganisms in each microscopic region of the fungal garden. This research will accomplish that with novel microscale metabolomics, lipidomics, and proteomics approaches that can analyze very small samples, providing detailed information on the location and function of fungal and bacterial molecules. In this initial study, we evaluated the lipidomic differences between the leaves feeding the gardens and spatially-resolve regions of the fungus garden at initial to advanced stages of leaf degradation. Lipids containing alpha-linolenic acid (18:3) from the leaves were enriched in the top of the gardens, where lysophosphatidylcholines (LPC) provided evidence for phospholipase activity and 18:3 signaling which adversely impact fungus health and plant biomass degradation.

When leaves are wounded, the polypeptide systemin is emitted from the damaged cells into the apoplast, signaling the liberation of 18:3 from plant membrane lipids into the cells. 18:3

begins the defense pathway by being converted to 12-oxophytodienoic acid and then through beta-oxidation is converted into jasmonic acid. The main defense mechanism in these leaves is jasmonic acid, and 18:3 is crucial to its synthesis. Despite extensive study of the cascade of molecular events in response to plant wounding, there is limited knowledge on the contribution and fate of individual membrane lipids, and specificity of phospholipase enzymes in this process. Which is the reason why we still do not know the exact origin of linolenic acid in the signaling pathway. To visualize lipid composition at wound sites with micrometer-scale resolution, we used two complementary matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI) platforms. We performed MALDI-Fourier transform ion cyclotron resonance (FTICR)-MSI and MALDI-trapped ion mobility spectrometry time-of-flight (timsTOF)-MSI experiments across wounded leaf sections of *Solanum lycopersicum* which is a model system for studying plant defense signaling. With concurrent mapping of phospholipids with 18:3 fatty acid composition and lysolipid spatial behavior we obtained insight into the possible origin of linolenic acid in the wounding process. Among lysolipids, LPC species strongly co-localize with the injured zone of wounded leaflets in all bioreplicates. We observed the highest spatial correlation between LPC (16:0): LPC (18:3), LPC (16:0): LPC (18:2) and LPC (16:0): LPC (18:0) ion image pairs, while lower correlation is observed between individual 18C LPCs.

Here we explore how lipid levels change in leaves of dicot plants during microbial degradation and mechanical wounding. Both studies suggest that linolenic acids are predominantly released from phosphatidylcholines (PCs). Our micrometer-scale co-localization analysis in wounded zones suggests that linolenic acids are predominantly released from PCs with 16_18 fatty acid composition. A better understanding of plant molecular signaling pathways at a spatial and molecular level can aid in devising new approaches for the production of fuels and chemicals in bioenergy crops.

This research is supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER) under the Early Career Award Program. A portion of the research has been performed using EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

Spatiotemporal Dynamics of Photosynthetic Metabolism in Single Cells at Subcellular Resolution

Colin Gates (colin.gates@colorado.edu)[1,2], Taewoo Lee[2], Ivan Smalyukh[2], and **Jeffrey C. Cameron[1,2,3,4]**

[1]Department of Biochemistry, University of Colorado, Boulder, CO 80309, USA.
[2]Renewable and Sustainable Energy Institute, University of Colorado, Boulder, CO 80309, USA.
[3]Department of Physics, University of Colorado, Boulder, CO 80309, USA.
[4]National Renewable Energy Laboratory, Golden, CO 80401, USA.

Project Goals: The project objective is to design and build a multimodal nanoscopy system to generate adaptive 3D images with high-resolution, and real-time, dynamic label-free chemical imaging of metabolic processes in photosynthetic organisms.

Abstract:

Metabolism is highly organized in space and time. In bacteria, this spatial and temporal organization of metabolism enables multiple, often competing, reactions to occur simultaneously in the same cell. However, the architectural principles of metabolic reaction networks and underlying cellular complexity of bacterial cells is only beginning to be appreciated. To unlock the true potential of synthetic biology and design novel microbial systems, signaling pathways and metabolic networks, the subcellular environment must be considered (in space and time). Our interdisciplinary research team harnesses cutting-edge synthetic biology tools, advanced live-cell imaging modalities, quantitative image analysis, and an integrated theoretical framework to investigate the regulatory and physical design principles underlying the spatiotemporal modulation metabolism in single bacterial cells.

Cyanobacteria are major primary producers and are unique in their ability to perform oxygenic photosynthesis, nitrogen fixation, and CO_2 fixation using light energy; these reactions are naturally optimized through spatial and temporal separation. These attributes make cyanobacteria ideal platforms to investigate and modulate cellular architecture and metabolism. We have developed a new imaging system to enable multidimensional measurements of photosynthetic metabolism in vivo and will describe the new types of measurements that can now be made in single cells. Understanding the design principles that enable robust functionality of the photosynthetic and carbon-fixing machinery is a fundamental challenge to improve native and heterologous metabolic pathways.

References

1. Nicholas C. Hill, Jian W. Tay, Sabina Altus, David M. Bortz, Jeffrey C. Cameron. (2020). Lifecycle of a cyanobacterial carboxysome. *Science Advances* 6 (19): eaba12692.

2. Kelsey K. Dahlgren, Colin Gates, Thomas Lee, Jeffrey C. Cameron**.** Proximity-based proteomics reveals the thylakoid lumen proteome in the cyanobacterium *Synechococcus* sp. PCC 7002. (2020). *Photosynthesis Research*. https://doi.org/10.1007/s11120-020-00806-y

3. Sarah J. Hurley*, Boswell A. Wing, Claire E. Jasper, Nicholas C. Hill, Jeffrey C. Cameron*. Cyanobacterial oxygenic photosynthesis dominated Mid-Proterozoic primary productivity. (2021) *Science Advances* Vol 7. No. 2, eabc8998, DOE: 10.1126/sciadv.abc8998 (*Corresponding Author)

This study was financially supported by the U.S. Department of Energy (DOE) DE-SC0019306 (to J.C.C.).

A Multiscale Model of Fungal Growth & Metabolism

Jolene Britton*[1,2], Alireza Ramezani[1,2], Mark Alber[1,2], Dale Pelletier[3], Jessy Labbé[3], and **William R. Cannon**[4]

[1]Department of Mathematics, University of California, Riverside, Riverside, CA 92521
[2]Interdisciplinary Center for Quantitative Modeling in Biology, University of California, Riverside, CA 92521
[3]Oak Ridge National Laboratory, Oak Ridge, TN, 37830
[4]Pacific Northwest National Laboratory, Richland, WA, 99352;

Project Goals

The goals of this project are to develop hybrid machine learning/simulation models of *Pseudomonas fluorescens/ Laccaria bicolor interactions and dynamics*. These hybrid data-analytic/simulation models will be used to carry out virtual experiments and develop fundamental understanding of the interactions between *Pseudomonas fluorescens* and *Laccaria bicolor.* At the same time, we will carry out experiments aimed at developing and testing quantitative assays to measure the same interactions, and whose data will inform the virtual experiments. We are:

- Evaluating the impacts of (1) thiamine and phenazines and (2) trehalose, produced respectively by *P. fluorescens* and *Laccaria*, on the metabolisms of each other. Metabolic exchange is an emerging theme in bacterial-fungal and bacterial-bacterial interactions.
- Characterizing *Laccaria*-stimulated chemotaxis of *P. fluorescens* by coupling trehalose signaling and metabolism to chemotaxis *P. fluorescens*.
- Experimentally investigating (1) *Pseudomonas fluorescens* chemotaxis and metabolism of *Laccaria* produced metabolites, and metabolism of *P. fluorescens* produced metabolites in *Laccaria*.

Abstract

Bacterial-fungal interactions play a fundamental role in many processes including crop biofuel development and biosystem design. In this work, we focus on the interactions between the fungi *Laccaria bicolor* and the bacterium *Psuedomonas fluorescens*, which play an integral role in the fitness of the roots of the Populus tree, an organism of interest as a biofuel crop. *L. bicolor* synthesizes trehalose which stimulates growth and chemotaxis of *P. flourescens*. Furthermore, *P. flourescens* provides *L. bicolor* with thiamine thereby increasing fungal mass. We developed a multiscale computational model to investigate these interdependent interactions. Our focus of this presentation is on the development of the *L. bicolor* structure and characterizing the energetic costs of growth and maintenance.

The hyphae of the filamentous fungi *L. bicolor* are modeled as a series of connected line segments using an off-lattice model coupled with a grid representing nutrient concentrations in the external environment. Nutrients in the environment diffuse while hyphae absorb the nutrients in the substrate at a rate of uptake following Michaelis-Menten kinetics that is dependent on both the current concentration of nutrients inside the hyphae and the local concentration of nutrients external to the hyphae. Nutrients that have been absorbed into the mycelium are transported throughout the colony to either fuel further growth or contribute to maintenance of the fungi. Nutrients used for maintenance are converted to nutrients used for growth at a rate following Michaelis-Menten kinetics. Both forms of nutrients are transported throughout the mycelia structure between hyphae segments. Passive short-range translocation of nutrients for maintenance move by convection and diffusion. Additionally, nutrients for growth undergo longer range active translocation towards the hyphal tips. Nutrients due to maintenance experience a loss in concentration due to biomass maintenance [1]. The changes in concentrations of nutrients due to maintenance and growth are represented by time-dependent ordinary differential equations.

L. bicolor grows by means of apical elongation of the hyphae and sub-apical branching. The length of the hyphal tip segment exhibits a rate of change following Michaelis-Menten kinetics which is dependent on the concentration of nutrients for growth. The energetic cost of growth is dependent on the rate of extension and

results in a loss of nutrients used for maintenance and for growth. Since hyphae are observed to grow in straight directions, the change in angle of the direction of elongation from one hyphae segment to the next differs by a normally distributed random number with mean zero and a small standard deviation. While the biological mechanisms underlying lateral branching is not well understood, some theorize it is due to a build up of vesicles far from the tip [4]. We model this phenomena by increasing the likelihood for a branch to emerge as a function of the internal concentration of nutrients for growth[2]. A hyphae segment is eligible for branching if it has not previously branched, the internal nutrient concentrations are greater than the costs of growth, and the segment is located behind a septa. We assume compartments separated by septa are of uniform length, hence septa are located at every *N* segments on each hyphal branch. The angle of branching follows a normal distribution where the mean and standard deviation are determined from experimental data [3]. After apical growth or sub-apical branching occurs, a check is performed for anastomosis or hyphal fusion. In the model presented, this occurs when two hyphae segments intersect. In the instance of intersection, the endpoint of the hyphae segment that experiences new growth is redefined to be the point of intersection.

To begin studying the energetic costs of growth and maintenance of *L. bicolor*, the impact of three different rates on mycelia development are tested. The three rates tested were the maximum uptake rate, the maximum rate of conversion of nutrients for maintenance to nutrients for growth, and the maximum rate of elongation at the tip. Preliminary results show that each rate has a large impact on the development of the mycelium, in particular, the amount of branching, colony size, and concentration of internal nutrients.

Our future work aims to couple the fungal model with two other biologically relevant models. First, the fungal development model will be coupled with an thermodynamic-kinetic maximum entropy ODE model for metabolism. The metabolism model contains over 200 reactions including protein and nucleic acid synthesis, from which the costs of growth and maintenance can be calculated. Second, we will couple the fungal model with a bacteria movement model. Trehalose secretion at the tips of the hyphae acts as a source of diffusive chemoattractant for *P. fluorescens*. The bacteria are described by a subcellular element submodel, a coarse-grained approach for describing biological properties of bacteria with great flexibility. Each bacterium is represented by multiple nodes connected to one another with linear and rotational springs with parameters calibrated using experimental measurements of mechanical properties of the bacteria. Bacterial motion is governed by potential functions and a propulsive force, calibrated using cell tracking data, which determine the movement and rearrangement of the nodes in an overdamped regime. We use simulations to evaluate the orientation and turning angle of the bacteria as they reverse their direction of motion in response to chemotactic signaling. The novelty of this multiscale model is that it takes into account bacteria-bacteria, bacteria-external nutrient, and fungal-bacteria interactions in addition to providing specific predictions to be tested in experiments. For example, we plan to test the impact of variations in fungal structures and nutrient excretion rates on bacterial chemotactic behavior.

References:

1. Balmant, W., M. H. Sugai-Guérios, J. H. Coradin, N. Krieger, A. Furigo Junior, and D. A. Mitchell, 2015. A Model for Growth of a Single Fungal Hypha Based on Well-Mixed Tanks in Series: Simulation of Nutrient and Vesicle Transport in Aerial Reproductive Hyphae. *PLOS ONE* 10:e0120307.
2. Hopkins, S., and G. P. Boswell, 2012. Mycelial response to spatiotemporal nutrient heterogeneity: A velocity-jump mathematical model. *Fungal Ecology* 5:124–136.
3. Labbé, J. L., D. J. Weston, N. Dunkirk, D. A. Pelletier, and G. A. Tuskan, 2014. Newly identified helper bacteria stimulate ectomycorrhizal formation in Populus. *Frontiers in Plant Science* 5.
4. Prosser, J. I., 1993. Growth kinetics of mycelial colonies and aggregates of ascomycetes. *Mycological Research* 97:513–528

Funding:

This project is supported by the U.S. Department of Energy's Office of Biological and Environmental Research under contract 78460.

A Physics-Based Model of Fungal Metabolism and its Regulation

Sam Britton[1,2], Jolene Britton[1,2], Alireza Ramezani[1,2], Mark Alber[1,2], Dale Pelletier[3], Jessy Labbe[3], and **William R. Cannon***[4]

[1]Department of Mathematics, University of California, Riverside, Riverside, CA 92521
[2]Interdisciplinary Center for Quantitative Modeling in Biology, University of California, Riverside, CA 92521
[3]Oak Ridge National Laboratory, Oak Ridge, TN, 37830
[4]Pacific Northwest National Laboratory, Richland, WA, 99352;

Project Goals

The goals of this project are to develop hybrid machine learning/simulation models of *Pseudomonas fluorescens/Laccaria bicolor interactions and dynamics*. These hybrid data-analytic/simulation models will be used to carry out virtual experiments and develop fundamental understanding of the interactions between *Pseudomonas fluorescens* and *Laccaria bicolor.* At the same time, we will carry out experiments aimed at developing and testing quantitative assays to measure the same interactions, and whose data will inform the virtual experiments. We are:

- Evaluating the impacts of (1) thiamine and phenazines and (2) trehalose, produced respectively by *P. fluorescens* and *Laccaria*, on the metabolisms of each other. Metabolic exchange is an emerging theme in bacterial-fungal and bacterial-bacterial interactions.
- Characterizing *Laccaria*-stimulated chemotaxis of *P. fluorescens* by coupling trehalose signaling and metabolism to chemotaxis *P. fluorescens*.
- Experimentally investigating (1) *Pseudomonas fluorescens* chemotaxis and metabolism of *Laccaria* produced metabolites, and metabolism of *P. fluorescens* produced metabolites in *Laccaria*.

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 develop a complete physics-based model of metabolism, protein expression and gene expression and to couple this metabolic model to the filament model discussed in the poster, *A Multiscale Model of Fungal Growth & Metabolism*.

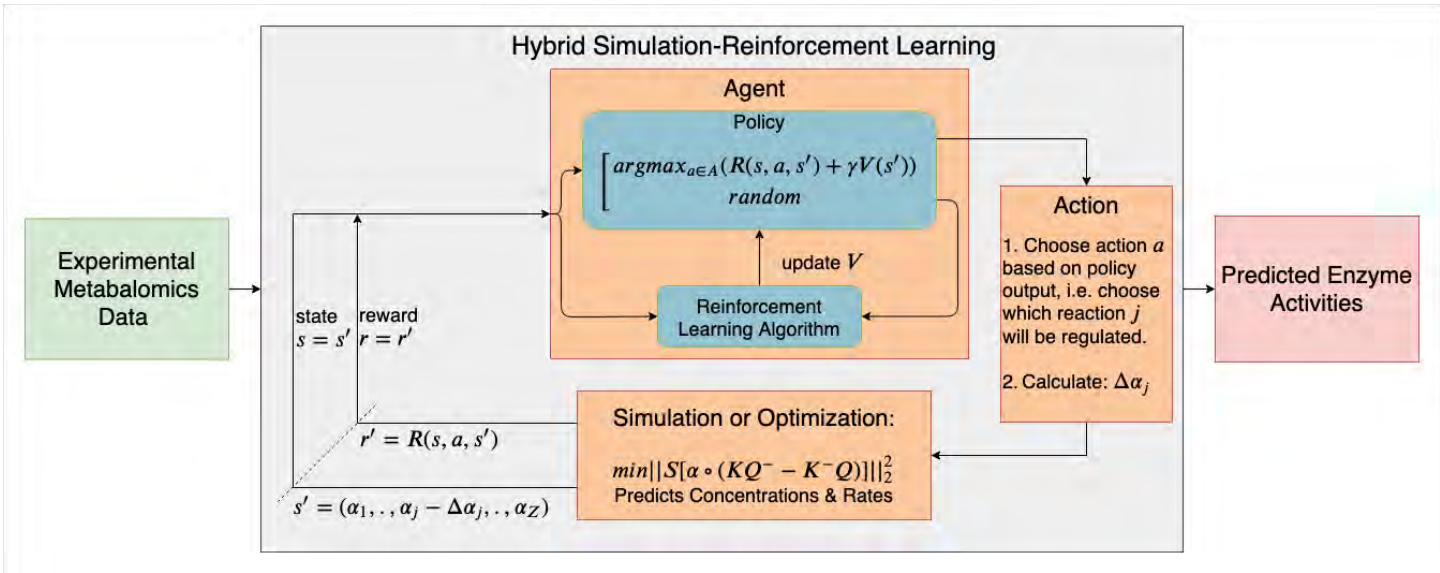
Our Laccaria bicolor metabolic model currently includes all reactions of central metabolism, synthesis of all 20 proteogenic amino-acids, synthesis of ATP, GTP, UTP, CTP and TTP and all deoxy-nucleic acid forms, synthesis of generic proteins, synthesis of generic DNA and growth on minimal media of glucose, ammonia, sulfate, and phosphate. Currently the model contains over 200 reactions and over 225 chemical species.

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 all of 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.

The ODEs are solved using optimization methods, which provides steady state solutions in seconds, but can also be solved using time-dependent ODE solvers when non-steady state simulations are needed.

Experimentally determined constraints on growth and metabolism are included in the model (for instance, see Control and Regulation below).

Control and Regulation of Metabolism. Experimental measurements or computational model predictions of the post-translational regulation of enzymes needed in a metabolic pathway is a difficult problem. Consequently, regulation is mostly known only for well-studied reactions of central metabolism in various model organisms. We utilized two approaches to predict enzyme regulation policies and investigate the hypothesis that regulation is driven by the need to maintain the solvent capacity in the cell [2]. The first predictive method uses a statistical thermodynamics and metabolic control theory framework. The second method is performed using a hybrid optimization-reinforcement learning approach:



Efficient regulation schemes were learned from experimental data that either agree with theoretical calculations or result in a higher cell fitness using maximum useful work as a metric. As previously hypothesized, regulation was shown to control the concentrations of both immediate and downstream product concentrations at physiological levels. The model predictions provide the following two novel general principles: (1) the regulation itself causes the reactions to be much further from equilibrium instead of the common assumption that highly non-equilibrium reactions are the targets for regulation; and (2) the minimal regulation needed to maintain metabolite levels at physiological concentrations maximizes the free energy dissipation rate instead of preserving a specific energy charge. The resulting energy dissipation rate is an emergent property of regulation which may be represented by a high value of the adenylate energy charge. In addition, the predictions demonstrate that the amount of regulation needed can be minimized if it is applied at the beginning or branch point of a pathway, in agreement with common notions. The approach is demonstrated for three pathways in the central metabolism, gluconeogenesis, glycolysis-TCA and Pentose Phosphate-TCA, which each require different regulation schemes. It is shown quantitatively that hexokinase, glucose 6-phosphate dehydrogenase and glyceraldehyde phosphate dehydrogenase, all branch points of pathways, play the largest roles in regulating central metabolism.

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.

"Breeding Resilient, Disease-Resistant Switchgrass Cultivars for Marginal Lands"

John Carlson[1] (jec16@psu.edu), Ryan Crawford[2], Jeremy Sutherland[1], Christopher Tkach[3], Terrence Bell[1], Stacy Bonos[3], Marvin Hall[1], Julie Hansen[2], Jesse Lasky[1], and Donald Viands[2]

[1]Pennsylvania State University, University Park, PA; [2]Cornell University, Ithaca, NY; and [3]Rutgers University, New Brunswick, NJ

Project Goals: This project is expanding on previous projects for the development of disease-resistant switchgrass cultivars for cultivation in the northeastern US, while expanding knowledge of underlying genotype-by-environment interactions. This involves: 1) Expanding the selection and testing of superior, disease-resistant switchgrass cultivars for marginal environments in the Northeast; 2) Mapping QTL for anthracnose resistance, Bipolaris resistance, and yield; 3) Identifying associations of SNPs and candidate genes with anthracnose and Bipolaris disease ratings; and 4) Identifying genome-wide and metagenome-wide variations associated with Genotype-by-Environment interactions affecting yield characters and disease susceptibility in switchgrass.

Abstract: Switchgrass is a fast growing, perennial, warm-season grass, native to North America with great potential as a bioenergy crop. In the humid Northeast, fungal diseases are prevalent, and these can reduce the yield and quality of harvests. We are building upon existing research, populations, and genomics tools from previous projects to accelerate the development of superior, disease-resistant, climate-resilient switchgrass (*Panicum virgatum* L.) cultivars for expanding the range of biomass cultivation in the Northeast. The project focuses on improvement of resistance to anthracnose (caused by *Colletotrichum navitas*), *Bipolaris* leaf spot (caused by *Bipolaris oryzae*), and environmental stress. In addition to cultivar development, we strive to better understand the sources of genetic and environmental variation affecting yield and disease susceptibility in switchgrass, including soil metagenome variation. To study Genotype-by-Environment interactions in=depth, we have established both cultivar yield trials and a GWAS family at 3 agronomically distinct field trial sites in NY, PA, and NJ.

Expanded selection and testing of disease-resistant switchgrass cultivars: A *seeded yield trial established at the NY, NJ, and PA field trial sites* in year 1, consisting of 13 cultivar selections replicated 4 times in randomized block design, was evaluated for disease incidence and biomass yield. Little anthracnose was observed, with exception of the NJ site. Vigor of plants at the mine reclamation site in PA was poor compared to the NY and NJ sites. A subset of plants from each site were prepared for evaluation of biomass composition by NIR. The *mature nursery at Ithaca* NY, consisting of 178 half-sibling progeny from two populations (one upland ecotype and one lowland ecotype) of switchgrass were evaluated for disease incidence. Leaf damage from unseasonably early frost in September made a second disease evaluation impractical. *Large progeny trials of 5,760 seedlings* from 180 advanced breeding lines, established in year 1 at the Rutgers and Ithaca field trial sites, were scored for vigor and disease incidence. Results of phenotypic data in years 1 and 2, including comparisons of years and sites, will be presented.

Three years of phenotypic data collection was obtained from the *QTL mapping population* of 240 full-sib plants at Rutgers. Overall, the mapping family exhibited a relatively normal distribution

of anthracnose disease severity, with a skew towards more disease susceptibility within the population. The family was transferred from the field site to the greenhouse over winter to provide healthy leaves for DNA isolation. After delays due to the pandemic, ddRADseq sequencing of the samples is underway. The data will soon be available for map construction and identification of QTL for disease-resistance.

Association mapping panel: A second year of phenotypic data (height, circumference, vigor, and anthracnose severity ratings) was collected from all 552 genotypes of the [Lu et al. 2013] association panel planted in 3 replicates at the NY, NJ, and PA trial sites in 2018. Bipolaris infections were not observed in 2019, however smut infections were wide-spread and thus severity ratings were also collected at all 3 trials sites. In general, growth, volume and vigor values were lower at the PA mine reclamation site relative to the NJ and NY field sites, although several genotypes did perform well there. Normal distributions of data were generally observed, with similar means in both 2019 and 2020. Anthracnose infection levels in the GWAS family were found to be negatively correlated with growth across the 3 test sites, as expected.

GWAS analyses were conducted to identify SNPs associated with anthracnose disease ratings and growth. A preliminary GWAS analysis for SNP-trait associations conducted with Tassel v5.0 software (Mixed Linear Model, using GBS genotypes (Lu et al 2013) and minimum SNP presence of 25%) was conducted for volume and anthracnose disease ratings collected in year 1 from all 540 association panel genotypes, replicates, and field sites in NY, NJ and PA. At a significance threshold of 5x10-8, 6 SNP markers were associated with anthracnose disease ratings, and 2 SNP markers were associated with calculated plant volumes. Several additional SNP markers just below the threshold are also being investigated. After year 2 growth phenotypes and disease ratings were collected, new GWAS studies were conducted to compare significance of associations among years and sites (Method = FDR, MAF >= 0.05, nSNPs = 67,721) using the statgenGWAS package in R and exome-capture genotypes (Evans et al., 2017). For severity of anthracnose infection, 92 SNPs were found to be significant above the fixed threshold of LOD = 4. Genes harboring significant SNPs included those related to known plant immunity and stress response gene families; such as NAC TranscriptionFactor-Like 9, Quinolinate Synthase, and GDP-D-mannose 3', 5'-epimerase (GME). Additionally, genes including those belonging to the Divergent CCT motif (relating to flowering time) and NAALAD (relating to plant cell wall regulation) were found to be significant and represented by a high LOD score (>4.5).

Rhizosphere microbiome composition: A further component of the assessment of genotype-by-environment interactions involves analysis of the rhizosphere microbiome composition and if the microbiome changed after transfer of the GWAS family to the 3 trial sites. The first step was characterization of the rhizosphere microbiome at the initial common garden site of the GWAS family at Cornell. Metagenomic sequence data (16S and ITS amplicons) from 382 rhizosphere soil samples revealed a rich rhizosphere microbial community of at least 493 bacterial genera and 57 fungal genera, distributed across 8 major phyla. Alpha diversity analyses indicated an influence of host switchgrass genotypes on rhizosphere bacterial diversity, and overall that rhizosphere diversity differed along three modes of switchgrass stratification - ploidy, ecotype, and populations. To assess microbiome composition after transplant and establishment at the 3 trial sites, soil core samples were taken for 128 selected switchgrass genotypes at all 3 trial sites

(Ithaca NY, Freehold NJ, and Philipsburg PA). Bacterial 16S amplicon libraries were generated from each sample. However, sequencing was delayed by covid pandemic issues. Sequencing and metagenomic analyses should be completed early in project year 3, barring further interruptions.

Reference cited:

1. Lu, F., Lipka, A.E., Glaubitz, J., Elshire, R., Cherney, J.H., Casler, M.D., Buckler, E.S. and Costich, D.E., 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS genetics*, *9*(1), p.e1003215.

2. Evans, J., Sanciangco, M.D., Lau, K.H., Crisovan, E., Barry, K., Daum, C., Hundley, H., Jenkins, J., Kennedy, M., Kunde-Ramamoorthy, G. and Vaillancourt, B., 2018. Extensive genetic diversity is present within North American switchgrass germplasm. *The plant genome*, *11*(1), p.170055.

Funding statement: This project is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2019-67009-29006. Seed support was provided by the Northeast Woody/warm-season BIOmass consortium, funded by USDA-AFRI Grant #2012-68005-19703. Program support to JEC was provided through the USDA National Institute of Food and Agriculture Federal Appropriations under Project PEN04532 and Accession number 1000326.

Adaptive hydrophobic and hydrophilic surface response of fungi to changing growth conditions

Saskia Bindschedler[1*] (saskia.bindschedler@unine.ch), Guillaume Cailleau[1], Ilona Palmieri[1], Matteo Buffi[1], Silvia Schintke[2], Eleonora Frau[2], Olivera Scheuber[2], Jana Reichenbach[3], Lukas Y. Wick[3], Pilar Junier[1], **Patrick Chain[4]**

[1]Institute of Biology, University of Neuchâtel, Neuchâtel Switzerland; [2] COMATEC-LANS, University of Applied Sciences Western Switzerland, Yverdon-les-Bains; [3]Department of Environmental Microbiology, Helmholtz Centre for Environmental Research UFZ, Leipzig Germany; [4]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico

https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-bacteria-fungal.php

Project Goals: Bacterial-fungal interactions (BFI) drive a multitude of process in the environment. In the context of soil, fungi are able to cope with heterogenous environmental constraints thanks to their filamentous lifestyle. This has a crucial impact on *fungal highways*, a BFI describing the ability of bacteria to spread in soil by using the fungal filamentous network. The formation of an aqueous film at the surface of hyphae appears to be a key parameter for the fitness of fungal highways. It is assumed that hydrophilic surfaces are more conducive to this fungal-driven bacterial dispersal. Therefore, in this study, we aimed at assessing how hyphal growth in hydrophobic or hydrophilic conditions impacts hyphae surface properties (i.e. hydrophilicity) by comparing two methods: a colony-scale approach (contact angle measurements) and a hypha-scale approach (atomic force microscopy).

At the microbial scale, soil is a heterogeneous environment composed of air pores, liquid patches, and solid particles. This results in the co-existence of numerous micro-environments with contrasting physicochemical characteristics that vary over space and time. This clearly represents a major influence to the development and activity of microbial life. In soil, filamentous fungi build 3D networks that may represent up to 10^4 m per g of soil. In addition to this, due to their microscopic dimensions, fungal hyphae have a high surface:volume ratio. As a result, a large fraction of the soil volume consists in fungal surfaces interacting with both biotic and abiotic soil components. This represents a central aspect of soil functioning. To be able to cope with the ever-changing conditions of soil microenvironments, while building and maintaining an extensive network, fungi need to dynamically adapt their surface properties. For instance, it is known that fungal hyphae are able to escape the water-air interface with amphiphatic peptides known as hydrophobins. In this study, we assessed the influence of growth conditions (either on hydrophobic or hydrophilic surfaces) on three fungal strains with two different approaches: contact angle (CA) and atomic force microscopy (AFM). CA measures surface hydrophilicity at the scale of fungal colonies, while AFM allows assessing surface properties at the scale of a single hypha. Two different types of behaviors were observed with CA: some fungi had a mycelial network that remained hydrophobic (or hydrophilic), whatever the growth conditions; while others dynamically adapted to the growth conditions. AFM showed that on hydrophobic surfaces, fungal hyphae tended to minimize the contact with the surface. On

the contrary, on hydrophilic surface, fungal hyphae tended to be flat by adhering to the surface and secreted EPS-like compounds. Likewise, the topographical surface characteristics of hyphae varied depending on the growth surface, but species-dependent patterns were still prominent. Overall, these results corroborate the ability of fungi to adapt their network to the highly dynamic conditions of soil by quickly adapting their surface properties. However, this adaptation occurs at the scale of single hyphae. Therefore, this aspect should be taken into account when considering fungal highways in which bacteria directly depend on the surface properties of fungal hyphae to spread.

This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T.

Endo-hyphal microbiome: a core of bacteria associated with the nitrogen cycle

Debora F. Rodrigues[1]* (dfrigirodrigues@uh.edu), Simone Lupini[1], La Verne Gallegos-Graves[2], Aaron J. Robinson[2], Julia M. Kelliher[2], Demosthenes Morales III[2], Armand E. K. Dichosa[2], Jean F. Challacombe[2], Jamey D. Young[4], Fabio Palmieri[3], Saskia Bindschedler[3], Andrea Lohberger[3], Pilar Junier[3], and **Patrick S. G. Chain[2]**

[1]Civil and Environmental Engineering, University of Houston, Houston, Texas; [2]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico; [3]Institute of Biology, University of Neuchâtel, Neuchâtel Switzerland; [4]School of Engineering Vanderbilt University, Nashville, Tennessee

https://genomicscience.energy.gov/research/sfas/lanlbfi.shtml

Project Goals: The interactions that occur in the rhizosphere between fungi and bacteria are fundamental for the ecological network of the soil. Over the past few years, scientist have described the diversity of microbial communities in a variety of soils associated with plants, yet little is known about the specific ecological significance and how these interactions are established. Bridging this knowledge gap could address multiple DOE priorities, as the effect of these interactions can have positive repercussions on the reduction of chemical fertilizers and increment of plant resistance towards abiotic and biotic stresses. Here we present preliminary results on the composition and possible function that these fungal-bacterial interactions could have on the soil ecosystem. Based on the isolation and identification of possible fungal-bacterial association from the soil, and combining bioinformatics and molecular tools, we are on the verge of understanding the importance of these associations.

Over the past several years, the scientific community has described the diversity of microbial communities in a variety of soils associated with plants, but at present, little is known about the specific diversity of the soil fungal microbiome, involving bacteria colonizing the surface of fungi (i.e., exo-bacteria) or existing within fungal hypha (i.e., endobacteria). This study aimed to collect, identify, and characterize several fungi and their associated (endo- and exo-) microbiome collected from the rhizosphere of several plants. Microcosm devices called fungal highway columns, containing one of four plant-based media as attractants, were placed in the rhizosphere of six different plants. The isolated fungi and their associated endo- and exo- bacteria were identified by sequencing of the ITS (fungi) or 16S (bacteria) rRNA regions, followed by Scanning Electron Microscope (SEM) and fluorescence in situ hybridization (FISH) imaging. Most of the fungi recovered are known plant pathogens, such as *Fusarium, Pleosporales,* and *Cladosporium* together with species associated with the soil, *e.g. Kalmusia.* The bacteria recovered were previously described as plant promoters, such as *Bacillus, Rhizobium, Acinetobacter* or *Ensifer*. The interactions between fungi and exo-bacteria recovered from any single medium were further investigated via confrontation assays. From the reconstruction of the potential co-occurring bacterial-fungal association in the rhizosphere, we discovered that the most promiscuous exo-

bacterium group (associated with diverse fungi) was *Bacillus*, which presented either neutral or negative interactions with different fungi. Furthermore, some endobacteria and exo-bacteria identified seemed to be host-specific. These findings suggest a complex interaction between fungi and bacteria in the rhizosphere.

This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T

Resolving Intracellular Chloroplast in Fungi from Sequence to Slide.

Demosthenes Morales, III,[1,2*] (dmorales@lanl.gov), Geoffrey House,[2] Julia Kelliher,[2] Aaron Robinson,[2] Pilar Junier,[3] Jim Werner,[1] and **Patrick Chain**[2]

[1]Center of Integrated Nanotechnologies, Los Alamos National Laboratory; Los Alamos, NM
[2]Biosecurity and Public Health, Los Alamos National Laboratory, Los Alamos, NM
[3]Université de Neuchâtel, Switzerland

https://genomicscience.energy.gov/research/sfas/lanlbfi.shtml

Project Goals: This project intends to inform the use of soil microbiomes to address DOE priorities in overcoming energy and environmental challenges. We are focusing on understanding the role of bacterial:fungal interactions in ecosystem development by connecting microbial diversity to actionable phenotypic responses. To do so, genomic and metagenomic sequencing of soil microbes will be combined with advanced imaging techniques and metabolomics to determine a mechanistic route in which these organisms associate to augment soil fertility and plant growth.

Abstract: Surveying the metagenomes of soil microbiomes has led to the understanding that there is an intricate network of microbes that interact with each other both extra- and intracellularly. Fungi contribute largely to the complexity of soil ecosystems and 16S amplicon sequencing revealed that they house their own diverse microbiomes comprised of previously unreported bacterial endosymbionts. More impressively, 16S signatures of plant chloroplasts have also been observed and appear to persist within the fungi across generations. Here, we sought to investigate the existence of chloroplasts within fungi at the cellular and molecular level. Using amplified fluorescence *in situ* hybridization techniques we were able to visualize abundant signals closely corresponding to chloroplast spatially distributed across the fungal hyphae of several environmental isolates. Understanding how chloroplasts can be utilized and maintain within Fungi may greatly change our perspective to the role of chloroplasts across kingdoms

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 by the Center of Integrated Nanotechnologies, a DOE user facility at Los Alamos National Laboratory.

LA-UR-20-20727

Exploring Mechanisms of Bacterial-Fungal Interactions Using *Ralstonia pickettii* Genomes Obtained from Diverse *Monosporascus* Isolates

Aaron J. Robinson[1,2*] (arobin@lanl.gov), Donald O. Natvig,[2] Demosthenes Morales,[1] Julia M. Kelliher,[1] La Verne A. Gallegos-Graves,[1] Karen W. Davenport,[1] and **Patrick S. G. Chain**[1]

[1]B-10 Biosecurity and Public Health, Los Alamos National Laboratory, Los Alamos, New Mexico; [2]University of New Mexico, Albuquerque

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 presents 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. Herein we discuss continued efforts towards establishing experimental models to examine and compare bacterial-fungal interactions.

Interactions between fungi and bacteria are both common and diverse. Descriptions of these interkingdom interactions generally fall into two categories: internal and external associations. The total number of descriptions of internal associations, where the bacterial partner is present and often maintained within cells of the fungal host, as well as the taxonomic diversity of these descriptions continue to steadily increase over time. However, the mechanisms that allow or promote these internal associations, as well as their diversity remain largely unknown due to the currently limited number of descriptions and genetic resources.

Genome sequencing results from several *Monosporascus* (Ascomycota; Xylariales) isolates obtained from the roots of plants in the Southwestern U.S. contained a substantial number of bacterial reads, despite the isolates being grown on diverse antibiotics and having been sub-cultured several times before sequencing. The majority of these sequences were classified as *Ralstonia pickettii* (Burkholdariaceae) at both the read and contig levels. Here we show fluorescence *in situ* hybridization (FISH) imaging indicates that *Monosporascus* is capable of harboring and maintaining *R. pickettii* as a bacterial endosymbiont. Attempts at isolating *R. pickettii* from these fungal hosts have been unsuccessful and the persistence of this bacteria in fungal isolates which have been maintained in culture for several years suggest an intimate relationship. The genome sequencing results for three distinct *Monosporascus* isolates each contained enough reads for the *de-novo* assembly of near complete *R. pickettii* genomes. Phylogenetic comparisons indicate that these *R. pickettii* genomes recovered from *Monosporascus* represent three distinct lineages that are closely related to previously identified environmental isolates. Broad-scale evolutionary comparisons conducted with these genome

assemblies also suggest differences both among the endosymbiotic *R. pickettii* lineages and between the fungal-derived and non-fungal-derived *R. pickettii* lineages. The observed diversity of these associations involving closely related bacterial endosymbionts recovered from closely related fungal hosts indicate this association could serve as a valuable model for studying bacterial-fungal interactions.

This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T.

LA-UR-20-20697

Identification of the tyrosine- and phenylalanine-derived soluble metabolomes of sorghum

Jeffrey Simpson,[1] Jacob Olson[1 *] (olson169@purdue.edu), Brian Dilkes, [1] and **Clint Chapple[1]**

[1]Purdue University, West Lafayette, IN

Project Goals: Improving our understanding of plant genomes and metabolomes is critical to understand the function of genes, unlock higher plant productivity, develop new strategies to protect crops from biotic and abiotic stress, and identify sources of new plant-based products. Progress towards these goals is limited by the fact that we do not know the identity of most plant metabolites, their biochemical origins, or the function of most of the genes involved in their synthesis and regulation. We will address these challenges through our recently developed stable isotope feeding/LC-MS/genome wide association strategy. This will identify functional gene-metabolite relationships for metabolites that are derived from amino acids in Arabidopsis and sorghum and authenticate them using reverse genetics. When complete, these data will identify known and unknown metabolites within untargeted LC-MS analyses, and characterize the genes involved in their synthesis.

The synthesis of small organic molecules, known as specialized or secondary metabolism, is one way plants resist and tolerate biotic and abiotic stress. Many specialized metabolites are derived from the aromatic amino acids phenylalanine (Phe) and tyrosine (Tyr). Improved characterization of the specialized metabolites derived from these amino acids is necessary to inform strategies for developing crops with improved stress resilience and traits for the biorefinery. *Sorghum bicolor* (L.), a drought tolerant monocot, is widely cultivated for feed and food and is an attractive crop for biofuels. Unlike dicots, sorghum and other monocots possess Phe and Tyr ammonia-lyase activity (PAL and TAL, respectively), which generate cinnamic acid and *p*-coumaric acids, respectively. Cinnamic acid can, in turn, be converted to *p*-coumaric acid by cinnamate 4-hydroxylase. Thus, Phe and Tyr are both precursors of common downstream products. Not all derivatives of Phe and Tyr are shared. Each amino acid acts as the precursor for unique metabolites relevant to sorghum adaptation, such as the anti-herbivore cyanogenic glycoside dhurrin derived from Tyr. In this study we used ^{13}C isotopic labeled precursors, and our recently developed PODIUM analytical pipeline, to identify MS-features derived from Phe and Tyr in sorghum. Over 600 MS-features were identified from Phe and/or Tyr across the roots, stems, and developing leaves of sorghum seedlings. These features comprised 20 percent of the MS signal collected. Ninety percent of the labeled mass features were derived from both Phe and Tyr. The ratio of incorporation of Phe and Tyr varied considerably between metabolites and tissues, suggesting the existence of multiple pools of *p*-coumaric acid that are fed by the two aromatic amino acids. Phe incorporation was greater for many known hydroxycinnamate esters and flavonoid glycosides. In contrast, mass features derived exclusively from Tyr were the most abundant in every tissue. The Phe and Tyr-derived metabolite library was also utilized to retrospectively annotate soluble MS-features in two *brown midrib* mutants (*bmr6* and *bmr12*), locating several MS features that change significantly in each mutant.

Funding statement. This work was supported by the U.S. Department of Energy, Office of Science (BER), Grant DE-SC0020368.

Metabolic Source Isotopic Pair Labeling and Genome-Wide Association Are Complementary Tools for the Identification of Metabolite-Gene Associations in Plants

Jeffrey Simpson[1] (jsimpso1@purdue.edu), Cole Wunderlich[1], Brian Dilkes[1] and **Clint Chapple**[1]

[1]Purdue University, West Lafayette, IN

Project Goals: Improving our understanding of plant genomes and metabolomes is critical to understand the function of genes, unlock higher plant productivity, develop new strategies to protect crops from biotic and abiotic stress, and identify sources of new plant-based products. Progress towards these goals is limited by the fact that we do not know the identity of most plant metabolites, their biochemical origins, or the function of most of the genes involved in their synthesis and regulation. We will address these challenges through our recently developed stable isotope feeding/LC-MS/genome wide association strategy. This will identify functional gene-metabolite relationships for metabolites that are derived from amino acids in Arabidopsis and sorghum and authenticate them using reverse genetics. When complete, these data will identify known and unknown metabolites within untargeted LC-MS analyses, and characterize the genes involved in their synthesis.

The optimal extraction of information from untargeted metabolomics analyses is a continuing challenge. Here we describe an approach that combines stable isotope labeling, LC-MS, and the development of a computational pipeline (named PODIUM- https://github.com/chapple-lab/podium) to automatically identify metabolites produced from a selected metabolic precursor. We identified the subset of the soluble metabolome generated from phenylalanine (Phe) in Arabidopsis thaliana, which we refer to as the Phe-derived metabolome (FDM) In addition to identifying Phe-derived metabolites present in a single wild-type reference accession, the FDM was established in nine enzymatic and regulatory mutants in the phenylpropanoid pathway. To identify genes associated with variation in Phe-derived metabolites in Arabidopsis, MS features collected by untargeted metabolite profiling of an Arabidopsis diversity panel were retrospectively annotated to the FDM and natural genetic variants responsible for differences in accumulation of FDM features were identified by genome-wide association. Large differences in Phe-derived metabolite accumulation and presence/absence variation of abundant metabolites were observed in the nine mutants as well as between accessions from the diversity panel. Many Phe-derived metabolites that accumulated in mutants also accumulated in non-Col-0 accessions and was associated to genes with known or suspected functions in the phenylpropanoid pathway as well as genes with no known functions. Overall, we show that cataloguing a biochemical pathway's products through isotopic labeling across genetic variants can substantially contribute to the identification of metabolites and genes associated with their biosynthesis.

Funding statement. This work was supported by the U.S. Department of Energy, Office of Science (BER), Grant DE-SC0020368.

Precise genome editing in new microbial species using SSAP libraries and broad-host recombineering methods

Gabriel T. Filsinger[1,2] (Filsinger@g.harvard.edu), Timothy M. Wannier[2,3], Felix B. Pedersen[4,b], Isaac D. Lutz[5,6,b], Julie Zhang[7,b], Devon A. Stork[2,8], Anik Debnath[3,9], Kevin Gozzi[10], Helene Kuchwara[3], Verena Volf[2,11], Stan Wang[2,3], Xavier Rios[3], Christopher J. Gregg[3], Marc J. Lajoie[3], Seth L. Shipman[12,13], John Aach[3], Michael T. Laub[10], **George M. Church[2, 3]**

[1]Department of Systems Biology, Harvard Medical School, Boston, MA; [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA;[3]Department of Genetics, Harvard Medical School, Boston, MA; [4]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark; [5]Institute for Protein Design, University of Washington, Seattle, WA; [6]Department of Bioengineering, University of Washington, Seattle, WA; [7]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA; [8]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA; [9]Tenza Inc., Cambridge, MA; [10]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA; [11]Harvard University John A. Paulson School of Engineering and Applied Sciences, Cambridge, MA; [12]Gladstone Institutes, San Francisco, CA; [13]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA;

http://arep.med.harvard.edu

Project Goals: In this project we aim to develop efficient homologous recombination (HR) methods in diverse microbes. MAGE and recombineering are powerful tools that improve HR efficiency over 1000-fold in *E. coli*. These methods traditionally rely on SSAP protein, λ-Red β, which does not function in most in other species. Our work focuses on two distinct strategies of expanding recombineering to new chassis organisms. First, we studied the mechanism of SSAPs, including λ-Red β, in order to characterize the lack of portability of these proteins. Second, we developed a library of over 200 homologs of λ-Red β in order to screen diverse variants for functionality. With a reliable strategy for developing improved recombineering methods in species beyond *E. coli*, we plan to develop a suite of methods that will enable strain engineering in bacteria with unique capabilities for bioproduction.

Efficient genome editing methods are essential for biotechnology and fundamental research. Homologous recombination (HR) is the most versatile method of genome editing, but techniques that rely on host RecA-mediated pathways are inefficient and laborious. Phage-encoded ssDNA annealing proteins (SSAPs) improve HR 1000-fold above endogenous levels; however, they are not broadly functional. Using *Escherichia coli*, *Lactococcus lactis*, *Mycobacterium smegmatis*, *Lactobacillus rhamnosus*, and *Caulobacter crescentus* we first investigated the limited portability of SSAPs[1]. We find that these proteins specifically recognize the C-terminal tail of the host's single-stranded DNA-binding protein (SSB), and are portable between species only if compatibility with this host domain is maintained. Furthermore, we find that co-expressing SSAPs with a paired SSB can significantly improve activity, in some species enabling SSAP functionality even without host-compatibility. We used the improved portability of SSAP-SSB pairs to expand recombineering methods to new microbes.

In parallel, we designed and built a library of over 200 SSAP homologs in order to identify proteins that enable efficient genome editing across different prokaryotes[2]. We've screened and validated this library across *Escherichia coli*, *Lactococcus lactis*, *Mycobacterium smegmatis*, *Staphylococcus aureus*, *Agrobacterium tumefaciens*, *Corynebacterium glutamicum*, and *Caulobacter crescentus*. In each, we've identified and designed constructs with at least 10-fold higher genome editing efficiency than any previously developed recombineering methods. The library consistently enriches for a set of functional proteins, and provides the most reliable strategy yet developed for developing efficient HR methods in electrocompetent bacteria[3].

Finally, we demonstrate that high-efficiency HR using SSAPs far surpasses the mutational capacity of commonly used random mutagenesis methods, generating exceptional phenotypes inaccessible through sequential nucleotide conversions[1]. Error-prone methods of strain improvement, such as evolution and mutagenesis, do not efficiently generate key combination mutants that contribute to these expectational phenotypes. Additionally, HR focuses mutagenesis at a precise genomic locus, rather than across the entire genome. We demonstrate the use of high-efficiency HR to modulate protein function, but similar techniques could be used to diversify promoters or specifically disrupt target genes.

Publications

1) Filsinger, G. T., Wannier, T. M., Pedersen, F. B., Lutz, I. D., Zhang, J., Stork, D. A., ... & Church, G. M. (2021). Characterizing the portability of phage-encoded homologous recombination proteins. *Nature chemical biology*, 1-9.
2) Wannier, T. M., Nyerges, A., Kuchwara, H. M., Czikkely, M., Balogh, D., Filsinger, G. T., ... & Church, G. M. (2020). Improved bacterial recombineering by parallelized protein discovery. *Proceedings of the National Academy of Sciences*, *117*(24), 13689-13698.
3) Wannier, T. M., Ciaccia, P. N., Ellington, A. D., Filsinger, G. T., Isaacs, F. J., Javanmardi, K., ... & Church, G. M. (2021). Recombineering and MAGE. *Nature Reviews Methods Primers*, *1*(1), 1-24.

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.

Programming high-order combinatorial genetics with Cas9-mediated gene drive approach for cellular engineering

Xiaoge Guo,[1,2] (x.guo@wyss.harvard.edu), Alejandro Chavez,[3] Angela Tung,[1] Christian Kaas,[4] Aditya Kunjapur,[5] Yi Yin,[6] Surojit Biswas,[1,2] Christian Kramme,[1,2] Richie Kohman,[1,2] Jay Shendure,[7] James J. Collins,[1,9,10,11,12,] **George M. Church**[1,2,12]

[1]Wyss Institute at Harvard University, Boston, MA; [2]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA; [3] Department of Pathology and Cell Biology, Columbia University College of Physicians and Surgeons, New York, NY; [4]Department of Expression Technologies 2, Novo Nordisk A/S, Maaloev, Denmark; [5]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE; [6]Department of Human Genetics, University of Los Angeles, Los Angeles, CA; [7]Department of Genome Sciences, University of Washington, Seattle, WA; [9]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA; [10]Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA; [11]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA; [12]Broad Institute of MIT and Harvard, Cambridge, MA

http://arep.med.harvard.edu

Project Goals: Deciphering combinatorial genetic modifications required for host chassis optimization is key to creating robust and economical systems to realize the promise of metabolic engineering. Our goal is to develop a high-throughput microbial platform that enables rapid generation of targeted, high-order genetic alterations coupled to single-cell mapping of combinatorial genotypes to identify host genome modifications giving rise to enhanced productions of desired biomolecules.

Current apporaches to cellular engineering rely on introducing a single genetic alteration into a genome one at a time and then studying its effects of metabolic output. This “one gene at a time” approach for discovering which combination of mutations are best able to increase metabolic output is not only time-consuming and labor-intensive but also restricted in the number and order of genetic combinations that can be tested. Here, we introduce a Cas9-based gene drive framework for constructing pools of cells with combinatorial genotypes in which each cell is characterized by *N* combinations of defined genetic alteractions, along with a high-throughput method for genotype-phenotype mapping. We demonstrate the utility of our approach in rapidly swapping promoters of 19 target genes with 7 promoters representing a continuum of gene expression level and identifying all the combination of genetic manipulations that will result in high-level production of the carotenoid lycopene. Our strategy allows high-order combinatorial genetics to be explored in a high-throughput targeted manner, and greatly speed up the rate at which we are able to optimize cellular chassis to produce valuable metabolites for use in consumer, biomedical and industrial applications.

This project has been funded by the U.S. Department of Energy grant under Award Number DE-FG02-02ER63445. CK, AC, XG have filed a patent application (US Patent Application 62/396,395*) related to this work. Professor George M. Church’s financial interests are listed on arep.med.harvard.edu/gmc/tech.hmtl.*

Towards genetic incorporation of an Orthogonal Ribosome-tRNA pair and D-amino-acids in *E.coli*.

Jorge Marchand[1], Kamesh Narasimhan[1*] (Kamesh_Narasimhan@hms.harvard.edu), Dominika Wawrzyniak[1], Dan Wiegand[2], Akos Nyerges[1], Maxwell Kazman[1], Abhishek Chatterjee[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 Chemistry, Boston College, Boston, MA.

http://arep.med.harvard.edu

Project Goals: Interactions between the acceptor arm of tRNA and the active-site of ribosome is characterized by a set of Watson-crick base-pairs, conserved across all three domains of life on Earth. By exploring alternative ribosome-tRNA base-pairing interactions at this conserved loci, we have begun to lay the foundations for operationalizing a fully orthogonal genetic code in *E.coli.* Finally, we have also established a robust and sensitive analytical pipeline for detection of D-amino-acids at various stages of translation such as tRNA amino-acylation and in target proteins.

Functionally mature tRNAs across all domains of life of have a conserved terminal 3'-CCA trinucleotide in their acceptor arm. The 3'-CCA terminus of the tRNA engages in a highly-conserved set of Watson-crick base pair interactions with the Peptidyl-transferase centre (PTC) of the Ribosome, as it moves from the A-site to the P-site during the transpeptidation step [1]. Additionally, the integrity of 3'-CCA terminus is subject to surveillance by a host of tRNA processing machineries[2]. We have genetically engineered *E.coli* strains that obviate 3'-CCA tRNA surveillance and repair and have identified a subset of amino-acyl tRNA synthetases that can amino-acylate mutant tRNA acceptor ends. To further aid our efforts in screening for engineered synthetases that can act on mutant tRNA acceptor ends, we are harnessing T-box riboswitches as potential sensors[3]. Together with orthogonal ribosomes that carry compensatory mutations in their PTC to interact with variant tRNA acceptor arms, we are operationalizing a fully orthogonal genetic code in *E.coli*[4].

Finally, a number of barriers remain in the way of incorporating D-amino acids into proteins[5]. Towards addressing this, we have developed robust and sensitive analytical methods for detection of D-amino-acids at various stages of translation such as tRNA amino-acylation and incorporation into target proteins. By integrating these analytical methods with strain engineering and directed evolution of amino-acyl tRNA synthetases we are establishing strategies towards robust and efficient genetic incorporation of D-amino-acids in *E.coli.*

References

1. 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).
2. 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).
3. 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
4. Orelle, C. *et al.* Protein synthesis by ribosomes with tethered subunits. *Nature* **524,** 119–124 (2015).
5. Kuncha, S. K., Kruparani, S. P. & Sankaranarayanan, R. Chiral checkpoints during protein biosynthesis. *J. Biol. Chem.* (2019). doi:10.1074/jbc.REV119.008166

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.

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[1,3], Maximilien Baas-Thomas[2], Nili Ostrov[1], Shirui Yan[1], Alexandra Rudolph[2], Jenny Ahn[1], and **George M. Church**[1,3]

[1]Department of Genetics, Harvard Medical School, Boston, MA; [2]Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA; [3]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. The development of a specialized integration system and the optimization of 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 computational and experimental methods to identify fitness-decreasing changes and troubleshoot these cases. Leveraging cutting-edge computational tools and accelerated laboratory evolution[3] allowed us to predict target loci accurately and correct fitness within days. We are now extending our computational algorithms to provide an all-in-one solution for the genome-scale recoding of a wide array of prokaryotes.

As we approach the final assembly of a virus-resistant E. coli genome, we are also implementing dependency on non-standard amino acids and encoding modules for stringent biocontainment.

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 64-x. For a complete list of Dr. Church's financial interests, see also arep.med.harvard.edu/gmc/tech.html.

In situ sequencing for deciphering spatial taxonomic structures of plant-associated microbial communities.

Andrew C. Pawlowski[1,2]* (Andrew_Pawlowski@hms.harvard.edu), Jonathan M. Conway[3], Richie Kohman[2], Jeffery L. Dangl[3], and **George M. Church**[1,2].

[1]Harvard Medical School, Boston; [2]Wyss Institute for Biologically Inspired Engineering, Boston, Massachusetts; [3]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

http://arep.med.harvard.edu/

Project Goals:
We aim to develop methods for characterizing microbial communities with single-cell and spatial resolution. Here, we develop a fluorescent _in situ_ sequencing (FISSEQ) technology[1-3] for acquiring the spatial arrangement of bacteria to better understand how plant microbiomes impact host physiology.

Most investigations reduce microbial physiology to monoculture conditions, which does not consider their abundant interactions in natural environments. Indeed, much of our understanding of microbes stems from gene deletions, heterologous expression, and _in vitro_ enzyme characterization. Microbes need to be studied _in situ_ where their spatial organization holds biological importance for responding to their environment. However, _in situ_ characterizations of these complex phenotypes in natural environments are generally not feasible at scale and remain largely unknown.

Here, we describe our recent developments for the taxonomic identification of microbes with single-cell and spatial resolution. We have established _in situ_ hybridization (ISH)-based probes that can distinguish between sequences with as little as one nucleotide difference and with unique barcodes (i.e., unique molecular identifier [UMI]) for highly-multiplexed _in situ_ sequencing readouts. For method development, we improve the nucleotide discrimination of our probes to multiple positions that can be discerned by a high-fidelity ligase, expanding the number taxa that we can assay simultaneously. We then demonstrate our technology by assaying 25 diverse microbes and demonstrate spatial single-cell resolution with Sequencing by Combinatorial Hybridization. Notably, our library generation and Expansion Microscopy methodology is applicable to a wide range of bacteria with different cell walls and cell morphologies. We can differentiate several members of the same genus (e.g., within *Pseudomonas*, *Bacillus*, *Lactobacillus*, *Staphylococcus*, and *Borrelia* genera), demonstrating that our technology can distinguish between bacteria that standard FISH methods cannot.

We are now applying this technology to investigate the spatial distribution of a 185-member synthetic community on Arabidopsis roots in collaboration with the Dangl Lab[4]. While our technology was developed for spatial taxonomics, we are also pursuing an additional simultaneous sequencing modality; spatial transcriptomics of the Arabidopsis host to investigate the change in spatial gene expression in response to microbial community structure. To this end, we have demonstrated Expansion Microscopy of Arabidopsis root tips, assembled full-length ribosomal DNA genes from draft genome assemblies using pathracer, and are working towards assaying bacterial localization using a universal probe and gene expression of control mRNA targets.

This work forms the framework for investigating the taxonomic structure of microbial communities and is the foundation for future *de novo in situ* sequencing of microbes in natural environments.

References

1. Lee, J., Daugharthy, E. *et al.* Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc* 2015. 10(3): 442 58
2. Lee JH, Daugharthy ER, *et al*. Highly multiplexed subcellular RNA sequencing *in situ*. *Science*. 2014 21;343(6177):1360-3.
3. Alon S, Goodwin DR, Sinha A, Wassie AT, Chen F, *et al*. Expansion sequencing: Spatially precise *in situ* transcriptomics in intact biological systems. *Science*. 2021 29;371(6528):eaax2656.
4. Finkel, OM, Salas-González, I, Castrillo, G, Conway JM, et al. A single bacterial genus maintains root growth in a complex microbiome. *Nature* 2020. 587, 103–108.

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.

High-throughput pooled functional screens *via* in-vivo production of Single-stranded DNA, toward applications in photosynthetic microbial hosts

Max G. Schubert[1]* (max_schubert@hms.harvard.edu), Daniel B. Goodman[2], Fahim Farzadfard,[3] Timothy K. Lu,[3] Seth L. Shipman[2], Himadri Pakrasi[4], Pramod Wangikar[5], **George M. Church**[1,3]

[1]Harvard University, Boston, MA; [2]University of California at San Francisco, CA; [3]Massachussetts Institute of Technology, Cambridge, MA; [4]Washington University, St. Louis, MO; [5]IIT-Bombay, Powai, Mumbai, India

http://arep.med.harvard.edu

Project Goals: Enable and explore recombineering using substrates produced *in-vivo*, via specialized bacterial retro-elements, creating a pooled Functional Genomics system, combining efficient editing and NGS-based tracking of mutants. Deploy this tool and other Next-generation genome editing technology toward Energy-related goals, including the study of photosynthesis.

Tremendous genetic variation exists in nature, but our ability to create and characterize individual genetic variants remains far more limited in scale. Likewise, synthetic variants aid our understanding of gene and genome function, but computational design of variants outpaces experimental measurement of their effect. Here, we show *in-vivo* production of single-stranded DNA via the targeted reverse-transcription of Retrons enables efficient and continuous generation of precise genomic edits in *Escherichia coli* at greater than 90% efficiency[1]. Curiously, the newly-identified single-stranded annealing protein (SSAP) CspRecT improves efficiency of this process by as much as 11-fold, a larger effect than observed when CspRecT is used with electroporated DNA. Because barcoded mutations are created, this tool also effectively couples phenotypes to a targeted sequencing output, enabling pooled high-throughput screens of genetic variants, a process we call Retron Library Recombineering (RLR)[1]. We measure antibiotic resistance resulting from synthetic variants using both qualitative and quantitative RLR protocols for pooled phenotypic measurement. RLR can also be performed using natural variants as input, and we demonstrate this by using sheared genomic DNA of an evolved bacterium as an input substrate for RLR. In this way, we identify causal variants leading to antibiotic resistance, and demonstrate a saturating genomic RLR library, in which tens of millions of barcoded experiments are performed within each single pool, and all genetic variants in a strain are exhaustively tested. The capacity to use non-designed DNA for such a screen stands in contrast to CRISPR-based methods, and opens the door to many applications using undefined and/or degenerate variation. Pooled genomic editing using ssDNA produced *in vivo* thus presents new avenues for creating and exploring variation at the whole genome scale.

The future of Bioenergy depends on applying these rapid genetic engineering techniques to organisms using atmospheric CO_2 for carbon, and sunlight or renewables for energy. To this end, we have begin to apply our approaches to fast-growing cyanobacteria, an exciting new area of bioenergy research[2,3]. Growing at speeds previously thought impossible, these photosynthetic bacteria are a possible next-generation renewable synthetic biology host. Because functional genomics approaches depend on complete, accurate genomes, we have begun by completing the genome of PCC11801, the cyanobacterial strain with the fastest growth rates in ambient air (~2.3hr doubling time)[3]. Our draft genome resolves existing assembly gaps, establishes that PCC11801 is a more distant relative to the model strain PCC7942 than was previously

appreciated, and confidently detects 6 novel plasmids stably maintained in this strain. These previously unreported plasmids represent ~12% of the genetic material in the strain, possibly important for the strain's intriguing phenotypes. Each plasmid shares locally collinear blocks with other plasmids reported in this *Synechococcus* clade, suggesting an episomal component to the core genome of *Synechococcus*. Directed evolution in this strain background enabled us to isolate a strain growing to higher density in batch growth, and having less adherence to the growth vessel. We are excited to develop pooled functional genomics approaches for fast-growing cyanobacteria, and our highest priority is to use our existing approaches[4-6] to develop recombineering technologies in this organism. These technologies could aid our investigate these new phenomena, and further enable this new renewable synthetic biology chassis as they have for past hosts[6].

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. 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).

4. 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.

5. Wannier, T. M. et al. Improved bacterial recombineering by parallelized protein discovery. Proc. Natl. Acad. Sci. U. S. A. 117, 13689–13698 (2020).

6. Wannier, T. M. et al. Recombineering and MAGE. Nature Reviews Methods Primers 1, 1–24 (2021).

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.

Prevalence of autotrophy and the characterization of carbon reduction in dissimilatory phosphite oxidizing microbes

Sophia D. Ewens (sdewens@berkeley.edu)* [1, 2], Tyler P. Barnum[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 will test three specific hypotheses:

1. **DPO is an environmentally prevalent metabolism that co-occurs with global phosphite preserves.**
2. **DPO metabolism is universally conferred by the conserved *ptx-ptd* operon.**
3. **DPO is universally associated with CO_2 fixation**

The work here advances hypothesis 3.

Phosphite (P^{3+}) is a highly soluble, reduced P compound that can account for up to 30% of total dissolved P in diverse environments[1]. In 2000, Schink et al. isolated the first microorganism capable of dissimilatory phosphite oxidation (DPO) in which phosphite is used as a chemolithotrophic electron donor in cellular energy metabolism[2]. This organism, *Desulfotignum phosphitoxidans* FiPS-3, is an autotrophic homoacetogen for which DPO activity was attributed to the *ptx-ptd* gene cluster, and genomic analyses suggested that FiPS-3 fixed CO_2 via the Wood-Ljungdahl pathway[2]. Following the isolation of FiPS-3, *Phosphitivorax anaerolimi* Phox-21 was identified in the metagenome of a wastewater digester enrichment, serving as the second known DPO microbe[3]. However, unlike FiPS-3, Phox-21 contained no known pathways for CO_2 fixation, despite being grown autotrophically in the absence of alternative electron acceptors[3]. Genomic analysis revealed that Phox-21 must fix CO_2 via the reductive glycine pathway[3]. At the time, the reductive glycine pathway had been proposed as a synthetic pathway for CO_2 fixation[4], making Phox-21 the first natural representative identified to harbor this novel CO_2 fixation pathway[3]. Recently, *Desulfovibrio desulfuricans* was biochemically proven to use the reductive glycine pathway to fix CO_2, validating its legitimacy as the seventh known carbon fixation pathway[5]. Since the identification of Phox-21, our group has identified 21 novel DPOM through metagenomics of wastewater enrichments (Ewens, *et al.* PNAS, 2021)[6]. Taxonomic analyses revealed that DPOM span six taxonomic classes, but despite this diversity, physiological and analyses of the metagenome assembled genomes (MAGs) suggests that the typical DPOM is a chemolithoautotroph that specializes in phosphite oxidation coupled to CO_2 reduction[6].

CO_2 as an Electron Acceptor: Enrichment biochemistry revealed that DPOM preferentially grew in microcosms in which CO_2 was the only exogenous electron acceptor, and DPO was not definitively coupled to any electron acceptor other than CO_2 [6]. A physiological survey of one of our highly enriched DPO cultures further showed that CO_2 was necessary and sufficient to support phosphite oxidation and growth[6]. The final product of CO_2 reduction in Phox-21 remains enigmatic, as the genes for pyruvate conversion to acetate (phosphotransacetylase and acetate kinase) are missing from the genome[3]. Lactate is a possibility, as the genomes of Phox-21 and several other DPO MAGs contain D-lactate dehydrogenase, which converts pyruvate to lactate at the expense of NADH. This is an energetically favorable reaction that accounts for all reducing equivalents produced via phosphite oxidation:

$6PO_3^{3-} + 3CO_2 + 3H_2O \rightarrow 6PO_4^{3-} + C_3H_6O_3 \ \Delta G^{\circ\prime} = -29$ kJ/mol e^- (-348 kJ/mol lactate)

CO_2 Fixation to Biomass: In addition to serving as the electron acceptor for DPOM, CO_2 is also fixed into biomass as the carbon source. We supplemented our physiological observations with genomic analyses and found that, as observed in Phox-21, comparative genomics of DPO MAGs revealed a notable absence of any canonical CO_2-reduction pathways[6]. While none of the DPO MAGs contained any canonical CO_2 fixation pathways, the majority of DPOM genomes appear capable of CO_2-fixation to pyruvate via the reductive glycine pathway[6]. Even if not a universal carbon fixation pathway in DPOM, our analyses suggest the reductive glycine pathway might be an important autotrophic mechanism across diverse DPO taxa.

Ongoing work is focused on parsing out the detailed mechanisms of the carbon reduction and carbon fixation pathways of DPOM using HPLC and carbon tracing studies. These studies will also be critical to understanding how nutrients are being exchanged in the DPOM communities, as evidenced in parallel work by our group.

References:

1. Pasek, M. A., Sampson, J. M. & Atlas, Z. Redox chemistry in the phosphorus biogeochemical cycle. *PNAS* **111**, 15468–15473 (2014).
2. Schink, B. & Friedrich, M. Phosphite oxidation by sulphate reduction. *Nature* **406**, 37–37 (2000).
3. Figueroa, I. A. *et al.* Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E92–E101 (2018).
4. Yishai, O., Bouzon, M., Döring, V. & Bar-Even, A. In vivo assimilation of one-carbon via a synthetic reductive glycine pathway in Escherichia coli. *ACS Synth. Biol.* (2018) doi:10.1021/acssynbio.8b00131
5. Sánchez-Andrea, I. *et al.* The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans. *Nat. Commun.* 1–12 (2020) doi:10.1038/s41467-020-18906-7.
6. Ewens, S. D. *et al.* The Diversity and Evolution of Microbial Dissimilatory Phosphite Oxidation. *bioRxiv* 2020.12.28.424620 (2020) doi:10.1101/2020.12.28.424620.

Environmental and Biological Constraints on Dissimilatory Phosphite Oxidizing Microorganisms

Kyle S. Metcalfe (kyle.metcalfe@berkeley.edu)*,[1] Sophia D. Ewens,[1,2] Yi Liu,[1,2], Sepideh Sadeghi,[3] W. Andrew Jackson,[3] Alfonso F. Davila,[4] Michi Taga,[1] Kenny Mok,[1] and **John D. Coates**[1,2]

[1] Department of Plant and Microbial Biology, University of California, Berkeley
[2] Energy & Bioscience Institute, University of California, Berkeley
[3] Department of Civil, Environmental, and Construction Engineering, Texas Tech University
[4] Space Science and Astrobiology Division, NASA Ames Research Center

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 will test three specific hypotheses:

1. **DPO is an environmentally prevalent metabolism that co-occurs with global phosphite preserves.**
2. **DPO metabolism is universally conferred by the conserved *ptx-ptd* gene cluster.**
3. **DPO is universally associated with CO_2 fixation**

The work here advances hypothesis 1: Phosphorus (P) is an essential nutrient, but the majority of P is trapped in mineral deposits as oxidized phosphate (P^{5+}). Alternative P redox states are often ignored in P cycle models, despite the fact that reduced species have been identified in diverse environments[1]. Phosphite (P^{3+}) is a highly soluble, reduced P compound accounting for up to 30% of total dissolved P in diverse environments[2]. In 2000, Schink et al. isolated the first microorganism capable of dissimilatory phosphite oxidation (DPO) in which phosphite is used as a chemolithotrophic electron donor[3]. This organism, *Desulfotignum phosphitoxidans* FiPS-3, is an autotrophic acetogen for which DPO activity was attributed to the *ptx-ptd* gene cluster[4]. This gene cluster has since been identified in many metagenome-assembled genomes (MAGs) from wastewater enrichments[5], spanning six phylogenetic classes[6]. A search of global metagenomic databases revealed the presence of the *ptx-ptd* cluster in a numerous uncultured microoganisms from diverse environments[6]. Here, we propose new geochemical and biological constraints on DPO microorganisms (DPOM) in the environment, through a synthesis of insights from 1) geochemical modeling, 2) enrichment cultures of DPOM from wastewater, and 3) enrichment cultures of DPOM from estuarine sediment and groundwater. Assuming DPO was coupled to CO_2 reduction to formate,[5] geochemical modeling constrained energy yields for a range of PO_4^{3-}/PO_3^{3-}, which we compared with measurements from a range of environmental settings to identify environments where DPO was likely to provide energy for microbial metabolism. While PO_3^{3-} concentrations and electron accepting capacity provided geochemical constraints for environmental DPOM activity, further evidence from enrichment cultures of DPOM suggests that the community context of DPOM may also biologically constrain activity. Ewens *et al.*[6] found that most DPOM are related to syntrophs, which depend on methanogens to mediate

thermodynamically unfavorable metabolic reactions.[7] While PO_3^{3-} oxidation is too thermodynamically favorable to require syntrophic exchange, a symbiotic nutrient exchange may explain DPOM resistance to isolation.[5] By introducing a variety of inhibitors to a highly enriched PO_3^{3-} oxidizing culture (HEPO), we found that DPO activity is immediately inhibited by 2-bromoethanesulfonate (BES), a specific inhibitor of methanogens. Since methanogens are prolific corrinoid producers, we hypothesized that methanogens may supply DPOM with essential corrinoids in exchange for PO_4^- and reduced carbon products, constraining DPOM to those environments that host methanogens. Genomic analyses revealed that DPOM are incapable of corrinoid synthesis while supporting a role for corrinoids in DPO metabolism. Extractions coupled to HPLC-MS identified four corrinoids in the HEPO culture with a purported role for DPO activity. Enrichment cultures also provided geochemical constraints to DPOM where abundant PO_4^{3-} minerals such as struvite [$MgNH_4PO_4 \cdot 6H_2O$] neogenesis are observed.[8] Precipitation of highly-insoluble minerals should pose a challenge to DPOM by cell encrustation, constraining DPOM to conditions where mineral precipitation on cell surfaces is precluded unless a mechanism exists for enhancing PO_4^{3-} solubility. To further constrain environmental DPO activity, we established 240 DPOM enrichments from sediment samples and a longitudinal transect of the Sacramento River and the Los Angeles Basin, respectively. Results revealed prevalent PO_3^- oxidation in these environments and revealed a correlation between DPO activity and the *in-situ* potential for DPO in a variety of anaerobic environments. Ongoing work will aim to characterize DPO MAGs from enrichments that span the full diversity of environments sampled. By pairing geochemical analyses with metagenomics, we will map the metabolic and phylogenetic diversity of DPOM to their environmental context.

References

1. Figueroa, I. A. & Coates, J. D. Microbial Phosphite Oxidation and Its Potential Role in the Global Phosphorus and Carbon Cycles. *Adv. Appl. Microbiol.* **98**, 93–117 (2017).
2. Pasek, M. A., Sampson, J. M. & Atlas, Z. Redox chemistry in the phosphorus biogeochemical cycle. *PNAS* **111**, 15468–15473 (2014).
3. Schink, B. & Friedrich, M. Phosphite oxidation by sulphate reduction. *Nature* **406**, 37–37 (2000).
4. Simeonova, D. D., Wilson, M. M., Metcalf, W. W. & Schink, B. Identification and Heterologous Expression of Genes Involved in Anaerobic Dissimilatory Phosphite Oxidation by Desulfotignum phosphitoxidans. *Journal of Bacteriology* **192**, 5237 (2010).
5. Figueroa, I. A. *et al.* Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E92–E101 (2018).
6. Ewens, S. D. *et al.* The Diversity and Evolution of Microbial Dissimilatory Phosphite Oxidation. *bioRxiv* 2020.12.28.424620 (2020) doi:10.1101/2020.12.28.424620.
7. McInerney, M. J. *et al.* Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. *Ann N Y Acad Sci* **1125**, 58–72 (2008).
8. Schink, B., Thiemann, V., Laue, H. & Friedrich, M. W. Desulfotignum phosphitoxidans sp. nov., a new marine sulfate reducer that oxidizes phosphite to phosphate. *Arch Microbiol* **177**, 381–391 (2002).

Determination of Phosphite (HPO_3^{-2}) by a New IC/MS/MS Method Using an ^{18}O-labeled HPO_3^{-2} Internal Standard

Sepideh Sadeghi[*](Sepideh.sadeghi@ttu.edu)[1] , Todd A. Anderson[3], W. Andrew Jackson[1], and **John D. Coates**[2]

[1*]Department of Civil, Environmental, and Construction Engineering, Texas Tech University, Lubbock, Texas

[2]Department of Plant and Microbial Biology, University of California, Berkeley

[2]Energy & Biosciences Institute, University of California, Berkeley

[3]Department of Environmental Toxicology, Texas Tech University, Lubbock, Texas

Project Goals: This project investigates the role of microbial dissimilatory phosphite oxidation (DPO) in the global phosphorus and carbon biogeochemical cycles. As part of this we are investigating the natural occurrence of DPO and phosphite in a broad range of geochemical environments, and examining the fundamental physiological and biochemical aspects of DPO. We are combining systems biology and -omics approaches with physiological and geochemical studies to elucidate the geochemical impact, environmental prevalence and metabolic machinery underlying DPO. To achieve our goals, we are testing 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

The work described here advances hypothesis 1.

A new method has been developed to determine trace amounts of phosphite (HPO_3^{-2}) in environmental samples using ion chromatography with electrospray tandem mass spectrometry (IC-ESI/MS/MS). The method includes the production and use of an ^{18}O-labeled HPO_3^{-2} internal standard (IS). This isotopically labeled IS significantly improved sensitivity and could account

for matrix suppression. The method detection limit (MDL) was determined as 0.017 and 0.034 μg L^{-1} of HPO_3^{-2} (6.5 and 13 ng P L^{-1}) using a 500 and 25 μL injection loop, respectively. Precision (1–10%) and accuracy (recoveries = 96–106%) were established for a range of environmental samples using known (spiked) addition. The impact of ionic interferences was investigated by evaluating the response of the internal standard in the presence of common anions with respect to distilled deionized water. The most significant interference was due to nitrate (100 mg-NO_3^- L^{-1}) with a 99.99% reduction in IS intensity. The method was successfully applied to wastewater effluent, anaerobic digester influent, centrifugation supernatant (centrate), biosolid, surface water, tap water, and soil samples. Relatively low concentrations < 0.25 μg HPO_3^{-2} L^{-1} were measured in tap water, surface water and wastewater effluent, 1.6 μg kg^{-1} HPO_3^{-2} in soil samples, 2 μg HPO_3^{-2} L^{-1} in centrate, 2.8 μg kg^{-1} HPO_3^{-2} in biosolid, and ~ 4 μg kg^{-1} HPO_3^{-2} in anaerobic digester influent. Limited suppression was observed for all matrices. The largest IS peak area suppression (~98%) was observed in WW effluent with 500 μL injection loop; however, this method was able to quantify HPO_3^{-2} with good recoveries and precision despite the mentioned suppression, supporting the ability of the proposed method to quantify HPO_3^{-2} in different environmental matrices.

Keywords: Phosphite, IC/MS/MS, Internal Standard,

Genomic dissection of anthracnose resistance response in the sorghum nested association mapping populations

Hugo Edgardo Cuevas, USDA-ARS, Mayaguez, PR, Louis K Prom, USDA, College Station, TX, Joseph E. Knoll, Crop Genetics and Breeding Research Unit, USDA-ARS, Tifton, GA, Jason Wallace, University of Georgia, Athens, GA and Wilfred Vermerris, Department of Microbiology & Cell Science, University of Florida, Gainesville, FL

Abstract Text:

Sorghum [*Sorghum bicolor* L. (Moench)] is C4 tropical grass used for food, animal feed, forage, and bioenergy that is drought-tolerant and has lower nutrient requirements compared to many other grasses. The advantages of sorghum as biofuel feedstock are expected to further increase the area of sorghum cultivation. The establishment of sorghum production outside of its natural dry environment presents a challenge due to the presence of multiple abiotic and biotic constraints that can reduce biomass and seed yield and quality. Anthracnose, caused by the fungal pathogen *Colletotrichum sublineolum* in Kabat and Bubák (syn. *Colletotrichum graminicola* [Ces.] G.W. Wilson), is a prevalent disease in warm and humid sorghum cultivation regions. In highly susceptible lines, anthracnose can cause substantial yield losses (up to 50%) of both grain and biomass.

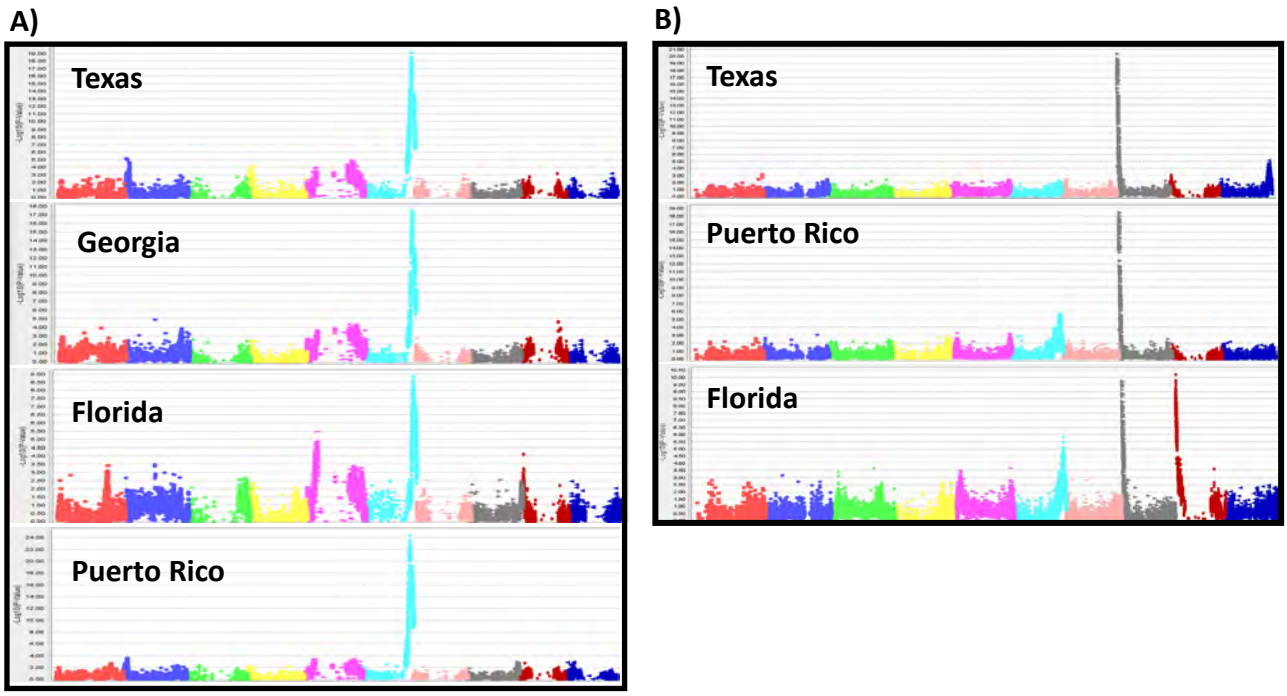
Several recent studies have identified loci responsible for broad-spectrum resistance to anthracnose in tropical and temperate adapted sorghum germplasm. The evaluation of the sorghum association panel (SAP) identified 40 accessions resistant across multiple locations. Genome-wide association analysis identified three loci at the distal region of chromosome 5, which explains 56% of the observed phenotypic variation. Therefore, other resistant sources present in the SAP were not detected due to their low frequency (<0.05) or because they were masked by overcorrection for population structure. Among the 40 resistance accessions identified in the SAP, we determined that three (SC1103, SC265 and SC1345) were used as founder lines in the sorghum nested association mapping (NAM) populations. **Therefore, to uncover resistance sources present in the SAP, we evaluated these three sets of recombinant inbred lines (RILs) populations for anthracnose resistance.**

The NAM population was developed to provide a germplasm and genomic resource that increases the power for the genetic dissection of economically important traits. Ten diverse lines representing the genetic variation of the SAP were crossed with a common parental line (RTx430) for the development of ten sets of >200 RILs each. A two-year replicated field trial of SC1103 and SC265 RILs populations in Texas, Georgia, Florida and Puerto Rico found segregation for anthracnose resistance response. High-density recombination linkage maps previously constructed based on genotyping by sequencing (GBS) of the RILs populations was used to delimit genomic regions associated with resistance response based on linkage analysis. Genome scan for SC265 RILs populations associated the distal regions of chromosome 6 with resistant response across locations (Figure 1A). This genomic region explains up to 46% of the observed phenotypic variation. Genome scan for SC1103 RILs populations associated the top region of chromosome 8 with resistance response in Texas and Puerto Rico (Figure 1B). This region and the top region of chromosome 9 were associated with resistance response in Florida, while no association was detected in Georgia. These genomic regions explain up to 40% of the

observed phenotypic variation. Based on a one year replicated trial of the SC1345 RILs population in Puerto Rico, the distal region of chromosome 5 was associated with resistance response. Since this region include the locus Sb005G172300 (i.e. genetic control similar to previously characterized in line SC112-14) and SC1345 line has the resistant allele this RIL population was not evaluated across location.

The results of this study indicate that SC1103 and SC265 contains new anthracnose resistance sources that could not be detected by the GWAS of the SAP. The development of other RILs populations is necessary to uncover novel resistance sources in the SAP. These new resistance sources present in temperate adapted germplasm are immediately available for sorghum breeding programs.

Figure 1. Anthracnose resistance response in sorghum nested association mapping populations evaluated at Texas, Georgia, Florida and Puerto Rico in 2019 and 2020. A) Genome scan for recombinant inbred lines derived from SC265; B) Genome scan for recombinant inbred lines derived from SC1103.



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

Elucidating Aromatic Utilization Mechanisms in Engineered *Rhodococcus opacus* Strains for Lignin Valorization

Winston E. Anthony[1]* (winstoneanthony@wustl.edu), JinJin Diao[2], Garrett Roell[2], Yifeng Hu[2], Rhiannon Carr[2], Kirsten Davis[2], Drew DeLorenzo[2], Bin Wang[1], Jie Ning[1], **Marcus Foston**[2], **Fuzhong Zhang**[2], **Hector Garcia Martin**[6,7,8,9], **Yinjie Tang**[2], **Tae Seok Moon**[2], and **Gautam Dantas**[1,3,4,5]

[1] The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63110, USA; [2] Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, St. Louis, MO, 63130, USA; [3] Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108, USA; [4] Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, 63130, USA; [5] Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108, USA, [6] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA, [7] Joint BioEnergy Institute, Emeryville, CA, 94608, USA, [8] Agile BioFoundry, Emeryville, CA, 94608, USA, [9] Basque Center for Applied Mathematics, Bilbao, Spain, [9] Pacifc Northwest National Laboratory, Richland, WA, USA

Project Goals: We aim to combine adaptive evolution, gene deletion analysis, and multi–omics approaches to identify aromatic tolerance and utilization mechanisms in the promising biofuel production strain *Rhodococcus opacus* PD630 (*R. opacus*). Our systems biology approach provides insights into the catabolic potential of *R. opacus* as a chassis for the conversion of lignocellulose, specifically thermochemically depolymerized lignin (i.e., aromatics), into valuable products.

R. opacus is naturally tolerant to aromatic compounds found in lignin-derived mixtures. We have demonstrated the potential of *R. opacus* for increased survivability in high concentrations of aromatics through adaptive evolution. Through genomic and functional characterization of wild type and adapted strains, pathways for aromatic degradation and funneling into central metabolism have been elucidated. Expression profiles have only been generated for select carbon sources, however, limiting our understanding of aromatic utilization and tolerance [1, 2].

To increase our knowledge of aromatic utilization and tolerance, we grew wild type *R. opacus* PD630 and mutant strains in minimal media supplemented with model lignin breakdown products at a total aromatic concentration permissive to WT growth. Additionally, we grew the mutant strains at higher concentrations of the relevant aromatics to examine the transcriptional changes which supported the increased-tolerance phenotype. Additionally, 13C metabolic flux analysis and targeted metabolomics were completed for WT/mutants growth on aromatics to rigorously measure and compare how aromatic substrates were consumed [3, 4].

We have been performing multi–omics analyses and gene deletion experiments to determine mechanisms of aromatic tolerance and utilization. Specifically, we have utilized transcriptomics, machine learning-based transcript-to-flux prediction models, and recently developed synthetic biology tools to elucidate the intriguing mechanisms of aromatic utilization [5-9]. This study will deepen our understanding of aromatic tolerance and utilization mechanisms in diverse *R. opacus* mutants by expanding the list of aromatic compound mixtures [10-12]. In

addition, this work will enable us to provide a genome-scale model of *R. opacus* to facilitate the development of the promising biofuel production organism.

Publications (+ Corresponding author)

1. WR Henson, T Campbell, D DeLorenzo, Y Gao, B Berla, SJ Kim, M Foston, TS Moon+ and G Dantas+. Multi-omic elucidation of aromatic catabolism in adaptively evolved *Rhodococcus opacus*. *Metab. Eng.* 49, 69–83 (2018)

2. A Yoneda, WR Henson, NK Goldner, KJ Park, KJ Forsberg, SJ Kim, MW Pesesky, M Foston, G Dantas+ and TS Moon+. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating *Rhodococcus opacus* PD630. *Nucleic Acids Res.* 44, 2240–2254 (2016)

3. GW Roell, R Carr, T Campbell, Z Shang, WR Henson, J Czajka, H Martin, F Zhang, M Foston, G Dantas, TS Moon+ and YJ Tang+. A concerted systems biology analysis of phenol metabolism in *Rhodococcus opacus* PD630. *Metab. Eng.* 55, 120-130 (2019)

4. WD Hollinshead, WR Henson, M Abernathy, TS Moon+ and YJ Tang+. Rapid Metabolic Analysis of *Rhodococcus opacus* PD630 via parallel ^{13}C-Metabolite Fingerprinting, *Biotechnol. Bioeng.* 113, 91-100 (2016)

5. DM DeLorenzo, WR Henson and TS Moon+. Development of Chemical and Metabolite Sensors for *Rhodococcus opacus* PD630. *ACS Synth. Biol.* 6, 1973–1978 (2017)

6. DM DeLorenzo, AG Rottinghaus, WR Henson and TS Moon+. Molecular toolkit for gene expression control and genome modification in *Rhodococcus opacus* PD630. *ACS Synth. Biol.* 7, 727–738 (2018)

7. DM DeLorenzo and TS Moon+. Selection of stable reference genes for RT-qPCR in *Rhodococcus opacus* PD630. *Sci. Rep.* 8:6019 (2018)

8. DM DeLorenzo and TS Moon+. Construction of genetic logic gates based on the T7 RNA polymerase expression system in *Rhodococcus opacus* PD630. *ACS Synth. Biol.* 8, 1921−1930 (2019)

9. WR Henson, F Hsu, G Dantas, TS Moon+ and M Foston+. Lipid metabolism of phenol tolerant *Rhodococcus opacus* strains for lignin bioconversion. *Biotechnol Biofuels*. 11:339 (2018)

10. W Anthony, RR Carr, DM DeLorenzo, T Campbell, Z Shang, M Foston, TS Moon+ and G Dantas+. Development of *Rhodococcus opacus* as a chassis for lignin valorization and bioproduction of high-value compounds. *Biotechnol Biofuels*. 12:192. (2019)

11. A Chatterjee, DM DeLorenzo, R Carr and TS Moon+. Bioconversion of renewable feedstocks by *Rhodococcus opacus*. *Curr. Opin. Biotechnol.* 64, 10-16. (2020)

12. K Davis and TS Moon+. Tailoring microbes to upgrade lignin. *Curr. Opin. Chem. Biol.* 59, 23-29. (2020)

This work is funded by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research, Biological System Sciences Division, award # DE-SC0018324.

Chemical and Morphological Structure of Solubilized Lignin Extracted via Ethanol, Tetrahydrofuran, and γ-Valerolactone Pretreatments from Wild Type and Transgenic Switchgrass

Luna Liang[1] *(lliang3@vols.utk.edu), Yun-Yan Wang[1], Samarthya Bhagia[2], Zhi Yang[2], Xianzhi Meng[1], Sai Venkatesh Pingali[2], Nathan Bryant[1], Vaidya Mathamangalath Sethuraman[2], Loukas Petridis[2], Yunqiao Pu[2], Barbara Evans[2], Hugh M. O'Neill[2], Arthur Ragauskas[1,2] and **Brian H. Davison[2]**

[1]University of Tennessee, 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.

The recalcitrance of lignocellulosic biomass remains a challenge in the biofuels and bioenergy process due to its complex physical and chemical structures of plant cell walls. To overcome the biomass recalcitrance, pretreatment and genetic modification are two techniques in the biological conversion of biomass to change or modify structures of biomass components in the plant cell wall. In this study, three organosolv pretreatments using ethanol (EtOH), tetrahydrofuran (THF), and γ-valerolactone (GVL) were applied on wild type and transgenic switchgrass including down-regulation of caffeic acid/5-hydroxyconiferyl aldehyde *O*-methyltransferase gene (COMT) and over-expression of MYB4 gene (MYB) (see Figure). The physicochemical properties of fractionated lignin precipitated from EtOH, THF, and GVL pretreatments were analyzed by gel permeation chromatography (GPC), small angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR) techniques including ^{31}P and two-dimensional ^{13}C-^{1}H heteronuclear single quantum coherence (HSQC). In addition, molecular dynamics simulation was used to provide the molecular modes of lignin molecules in the pretreatment solvents of EtOH, THF and GVL. The pretreated solubilized lignins revealed a significant decrease in molecular weight compared to the untreated native lignin, especially in EtOH pretreatment. A near complete removal of β-O-4 interunit linkages was also observed in EtOH pretreated lignin. Furthermore, THF pretreated lignin presented the highest molecular weight, $\beta - O - 4$ linkages and aliphatic hydroxyl contents among the three organosolv lignin streams.

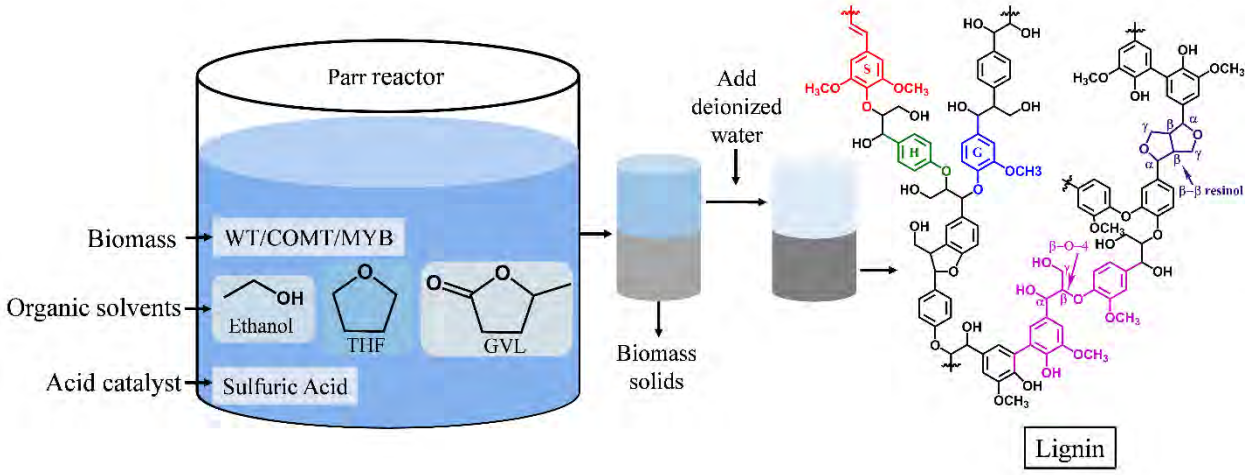


Figure: Pretreatment processes of WT, COMT, and MYB switchgrass.

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 Office of Biological and Environmental Research in the DOE Office of Science.

Developing next-generation modeling techniques to analyze biomembrane neutron scattering data

Louxi Tan[1], Elizabeth G. Kelley[1], Micholas D. Smith[2], Loukas Petridis[3], Sai Venkatesh Pingali[3], John Katsaras[3], Jeremy C. Smith[2,3], Hugh M. O'Neill[3], James G. Elkins[3], Jonathan D. Nickels[1*] (jonathan.nickels@uc.edu), and **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 biomolecules that comprise plant cell walls and microbial biomembranes.

The cellular membranes of fermentative microorganisms are a major target for the toxic effects of advanced biofuels, including *n*- and *iso*-butanol as well as solvents used in the pretreatment of lignocellulosic biomass such as tetrahydrofuran (THF). These amphiphilic molecules partition into the lamellar structure of the membrane bilayer affecting its viscosity, stability, and structure, both in the transverse and lateral directions. Small-angle neutron scattering (SANS) is ideally suited to measure structural properties of membranes, due to its probe-free nature that enables measurements with minimal perturbation to the membrane and its broad spatial resolution (i.e., ~1-100 nm). However, improved models are needed to analyze and extract the maximum information from the SANS data measured from the heterogeneous lamellar structures of microbial membranes.

Our team has studied model microbial membranes in the presence of co-solvents, Figure 1 (Smith et al. 2020), and is currently looking at studies of bacterial lipid extracts and the membranes of living bacteria. These studies have already revealed differences in the partitioning and localization of co-solvents, along with clear effects on membrane structure. As the membrane compositions become more complex and more biologically relevant, there is a clear need for reliable and robust structural models that can credibly extract structural information, despite the complex composition of biological membranes and the variable and broad spatial distribution of the co-solvent molecules partitioned within the membrane.

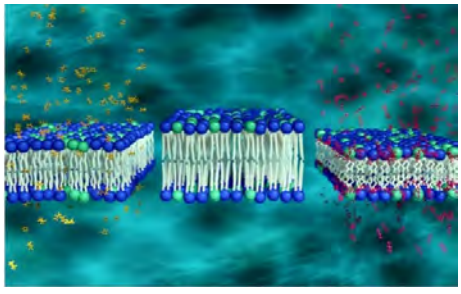


Figure 1. SANS and MD simulations reveal different partitioning and structural effects of THF (left, yellow) and *n*-butanol (right, purple) in model microbial membranes (bilayer of white fatty acids and blee or green headgroups).

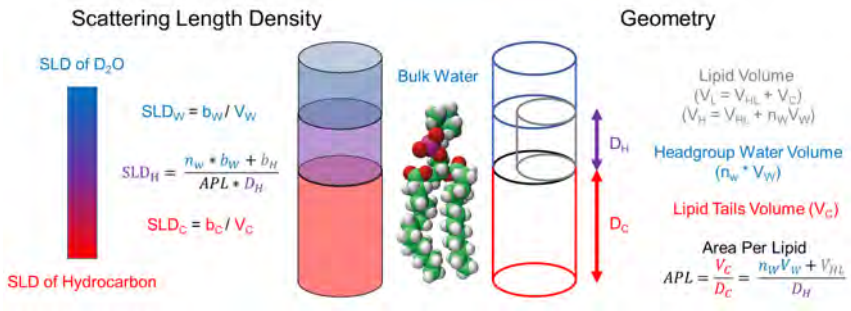


Figure 2. Implementation of the slab model for symmetric lamellar structures, shown here on a per lipid basis in one leaflet. Our implementation in Tan et al. (2021) provides a self-consistent fitting algorithm to extract the key parameters of area per lipid (APL) and water content in the head group region (n_H).

A robust solution to this question is a slab model, a relatively simple structural description that can be used to describe lamellar phases including lipid bilayers. To a first approximation, membrane structure can be described by three slabs (or sheets) namely, a central solvent-free core and two symmetric outer layers composing the solvated shell. This model well-describes the distribution of neutron scattering length density within a membrane sample and therefore is applied to interpreting SANS spectra derived from membrane lipid bilayers. Prior implementation of this model in common scattering software packages was prone to generating unreliable results due to the covariance of scattering length density and bilayer thickness. Here, we report on an improvement to the existing models within the publicly available software suite, *SasView*, which enforces physical consistency through the area per amphiphile molecule and number of solvent molecules included within the solvent-exposed outer layer. The model was applied to fit standard lipid bilayer scattering data sets, determine structural parameters consistent with prior literature values, and illustrate the typical and ideal cases of fitting for neutron scattering data obtained using single or multiple contrast matching conditions. The model has been submitted for inclusion in subsequent releases of *SasView* and will aid the broader scattering community studying lamellar structures.

Adding self-consistency to this model is an important first step to the development of lamellar models capable of describing the partitioning of co-solvents between the membrane and the bulk solvent, and the resulting changes in membrane structure. The improved slab model provides the opportunity to compare experimental membrane thickness and solvent positioning data to those quantities predicted by molecular dynamics simulations of biomembranes under solvent stress. Further, this comparison will permit future, experimentally guided, improvements to molecular mechanics membrane force-field accuracy, ultimately leading to more predictive computational tools for the study of membrane behavior under environmental stress.

References

1. L. Tan et al. “Implementation of a self-consistent slab model of bilayer structure in the SasView suite.” 2021, *J Appl Crystal* 54, DOI: 10.1107/S1600576720015526.
2. Smith, M. D., et al. “Solvent-induced membrane stress in biofuel production: molecular insights from small-angle scattering and all-atom molecular dynamics simulations.” 2020, *Green Chem* 22(23), 8278-8288. DOI: 10.1039/d0gc01865a

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This SFA program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Visualization of plant cell wall polymers using neutron scattering and deuterium labelling in *planta*

Sai Venkatesh Pingali[1]* (pingalis@ornl.gov), Zhi Yang, Samarthya Bhagia, Barbara Evans, Hugh O'Neill, Arthur Ragauskas, **Brian H. Davison**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]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.

The plant cell wall is a complex, multi-polymeric system that consists of primarily carbohydrates and lignin. Cellulose strands come together to form microfibrils while the amorphous polymers, hemicellulose and lignin form a network structure and fill in the interstitial space. Small-angle neutron scattering is ideally suited to study the complex hierarchical structure of biomass. However, neutron sensitivities of the different plant biopolymers are similar making structural association non-trivial and ambiguous. A promising approach is partial deuteration of the plants to increase the difference in the neutron sensitivity between the plant cell wall polymers. Here, we will present the results from three different partially deuterated plants: switchgrass (*Panicum virgatum*), kale (*Brassica oleracea*), and eucalyptus (*Eucalyptus camaldulensis*).

Partially deuterated switchgrass plants were obtained by hydroponic cultivation in media containing 50% D_2O to produce tiller biomass with 34% deuterium incorporation determined by NMR and contrast variation small-angle neutron scattering (CV-SANS).[1] The cell wall composition was found to be similar to the same cultivar grown hydroponically in H_2O, but with slightly increased lignin content and different lignin deposition pattern as determined by TEM and confirmed by SANS (see Figure). Surprisingly, these plants also had a lower recalcitrance to enzymatic hydrolysis. Partially deuterated kale was obtained by hydroponic cultivation in 31% deuterated media. Fourier-transform infrared spectroscopy (FTIR) results indicated that D/H substitution for carbohydrate is higher than for lignin. By combining CV-SANS and FTIR, it was determined that 50% of covalently bonded hydrogens were replaced with deuterium atoms in cellulose while only 10% for lignin. Similar results were observed for partially deuterated eucalyptus plants. These results open new avenues to visualize structural features of amorphous plant polymers in the plant cell wall from those of the well-organized cellulose microfibrils.

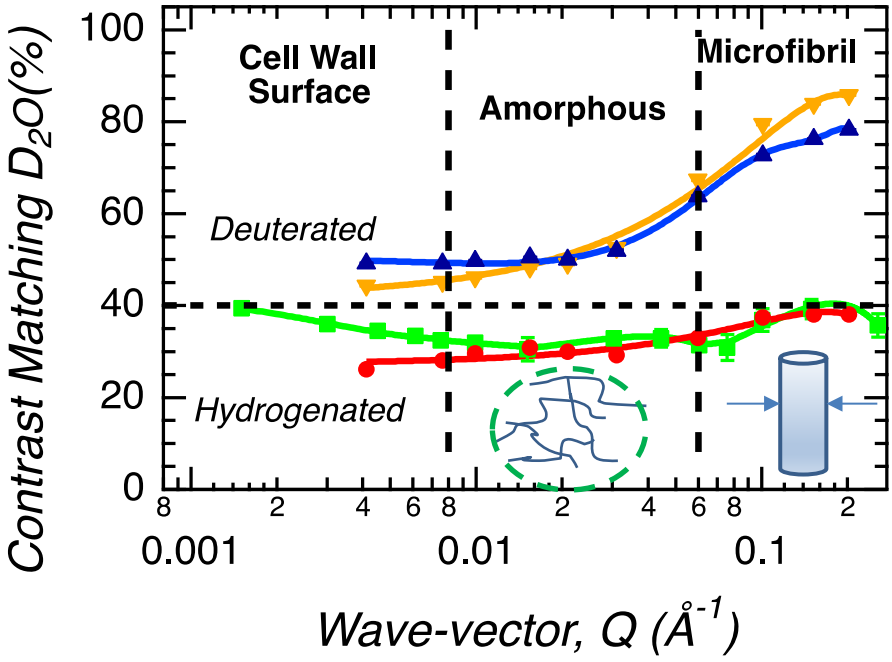


Figure: The variation in contrast matching D_2O solvent mixture as a function of wave-vector, Q. Three samples were hydroponically grown switchgrass in 100% H_2O (red dots; red line), 50% D_2O (blue filled up-triangle; blue line), and 40% D_2O at 30 °C (orange filled down-triangle; orange line) and field grown switchgrass in 100% H_2O (green filled-squares; green line).

References:

1. Evans BR, Pingali SV, Bhagia S, O'Neill HM, Ragauskas AJ, "Structural Studies of Deuterium-Labeled Switchgrass Biomass", In Understanding Lignocellulose: Synergistic Computational and Analytic Methods, *ACS Symposium Series* 1338, ed. Micholas Dean Smith, Chapter 2, 17-32 (2019) DOI: 10.1021/bk-2019-1338.ch002

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 Office of Biological and Environmental Research in the DOE Office of Science.

Title: Divergence in stress tolerance mechanisms across the Brassicaceae family highlight strategies for maintaining growth and physiological balance under extreme environments

Authors: Ying Sun,[1] Kieu-Nga Tran[2], Lina Duan[1], Dong-ha Oh[2], Andrea Ramirez[1], Guannan Wang[2], Maheshi Dassanayake[2], **José R Dinneny (dinneny@stanford.edu)** [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:

How plants achieve salt tolerance has been a focal question in our quest to develop bioenergy crops that can be grown on marginal lands for sustainable energy production. One bottleneck to overcome this challenge has been our lack of understanding in how naturally stress adapted plants, extremophytes, respond and thrive under salt stress. To address this need, we used the extremophyte models, *Schrenkiella parvula* and *Eutrema salsugineum* with the closely related premier model, *Arabidopsis thaliana* in a multi-omics experimental design. In this study, we used a comparative transcriptomic, ionomic, and metabolomic profiles to gain insight on how these extremophytes respond to salt stress differently from the stress sensitive model, *A. thaliana.*

We found that *S. parvula* and *E. salsugineum* accumulated less Na^+ and maintained higher K^+ in the shoots compared to *A. thaliana* under salt stress. However, Na^+ accumulation in *E. salsugineum* was comparable to that of *A. thaliana* in the roots. Among the two extremophytes, only *S. parvula* maintained a low Na^+ concentration in the roots upon salt exposure. Salt stress led to a reduction in the abundance of macro and micro-nutrients in *A. thaliana* roots, while both halophytes could maintain their overall nutrient balance similar to control levels at comparable salinities. Concordant with the ionomes, the shoot metabolomes of the extremophytes showed minimal changes compared to *A. thaliana*. Among the notable changes to the root metabolome, *S. parvula* increased in abundance of its primary sugars, amino acids, and other intermediates in the TCA cycle while the abundance of these metabolites decreased in *E. salsugineum* under salt stress. However, this reduction did not lead to a significant net reduction of these pools, as *E. salsugineum* had higher basal levels of these metabolites in roots compared to *S. parvula* and *A. thaliana*. Interestingly, *S. parvula* metabolomes converged with *A. thaliana* more than with *E. salsugineum* prior to salt stress and transitioned to a metabolome that was more representative of

E. salsugineum with increased salt concentrations or treatment durations. The transcriptomic responses to salt stress further supported the metabolic adjustments of each species. Up-regulated genes in all three species were enriched in stress-related pathways including ROS scavenging and osmolyte biosynthesis. In addition, differently regulated genes in *S. parvula* were indicative of uninterrupted root growth during salt stress governed by genes associated with auxin and ABA pathways, quite contrasting to the other two species. In the shoots, the transcriptomic effect on photosynthesis was significantly less in the extremophytes compared to *A. thaliana*. Overall, the extremophytes showed stress preparedness, both at the transcriptome and metabolome levels to allow sufficient nutrient uptake to promote growth and development under salt stress levels that impaired growth in Arabidopsis. Both extremophytes indicated successful yet different strategies for stress tolerance at the metabolome and transcriptome levels during salt stress.

While the hormonal networks controlling growth have been extensively characterized in stress-sensitive plants, it is unclear how these pathways are rewired in plants that maintain growth in extreme environments. We have compared the developmental and molecular responses of four closely related members of the Brassicaceae family including two salt-tolerant species (*S. parvula* and *E. salsugineum*) and two salt-sensitive species (*Sisymbrium irio* and *A. thaliana*) to the salt stress-induced hormone, abscisic acid (ABA). While ABA inhibits root growth in most species, we uncovered substantial growth-promoting effects in *S. parvula*, due to an enhancement in cell elongation. Comparative transcriptomics informed by phylogenetic relationships uncovered lineage and extremophile-specific differences in ABA response. DNA Affinity Purification followed by sequencing (DAP-Seq) was utilized to establish gene regulatory networks (GRNs) in each species for the entire ABA-RESPONSIVE ELEMENT BINDING FACTORS (AREB/ABF) clade. Comparative GRN analysis identified relative conservation in the core ABA signaling GRN, while the auxin growth-hormone GRN was highly divergent, revealing how patterns of gain and loss of cis-regulatory elements mediate novel physiological outcomes. Our findings demonstrate that the targets of hormone signaling pathways are highly divergent between species and that diametric inversion of growth regulation is possible, even between closely related species of the same plant family (Sun et al., 2020).

References/Publications

1. Sun Y, Oh DH, Duan L, Ramachandran P, Ramirez A, Bartlett A, Dassanayake M, **Dinneny JR** (2020) Divergence in a stress-associated gene regulatory network underlies differential growth control in the Brassicaceae family. bioRxiv preprint doi: 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: Identification of Cell-type Marker Genes from Plant Single-cell RNA-seq Data Using Machine Learning

Authors: Haidong Yan[1](yanhd@vt.edu), Qi Song[1,2,4], Jiyoung Lee[1,2], **John Schiefelbein**[3], **Song Li**[1,2*]

Institutions: [1]School of Plant and Environmental Sciences (SPES). [2]Graduate program in Genetics, Bioinformatics and Computational Biology (GBCB), [3]Department of Molecular, Cellular, and Developmental Biology, University of Michigan. Ann Arbor, MI 48109. *Corresponding author (songli@vt.edu).

Website: https://github.com/LiLabAtVT/SingleCellClassification

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: An essential step of single-cell RNA sequencing analysis is to classify specific cell types with marker genes in order to dissect the biological functions of each individual cell. In this study, we integrated five published scRNA-seq datasets from the Arabidopsis root containing over 25,000 cells and 17 cell clusters. We have compared the performance of seven machine learning methods in classifying these cell types, and determined that the random forest and support vector machine methods performed best. Using feature selection with these two methods and a correlation method, we have identified 600 new marker genes for 10 root cell types, and more than 70% of these machine learning-derived marker genes were not identified before. We found that these new markers not only can assign cell types consistently as the previously known cell markers, but also performed better than existing markers in several evaluation metrics including accuracy and sensitivity. Markers derived by the random forest method, in particular, were expressed in 89-98% of cells in endodermis, trichoblast, and cortex clusters, which is a 29-67% improvement over

known markers. Finally, we have found 111 new orthologous marker genes for the trichoblast in five plant species, which expands the number of marker genes by 58-170% in non-Arabidopsis plants. Our results represent a new approach to identify cell-type marker genes from scRNA-seq data and pave the way for cross-species mapping of scRNA-seq data in plants.

References/Publications

1. Identification of cell-type marker genes from plant single-cell RNA-seq data using machine learning. Haidong Yan, Qi Song, Jiyoung Lee, John Schiefelbein, Song Li doi: https://doi.org/10.1101/2020.11.22.393165

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0020358.

Title: Identification of cell-type marker genes from plant single-cell RNA-seq data using machine learning

Haidong Yan[1], Qi Song[1,2,4], Jiyoung Lee[1,2], John Schiefelbein[3], Song Li[1,2*]

[1]School of Plant and Environmental Sciences (SPES). [2]Graduate program in Genetics, Bioinformatics and Computational Biology (GBCB), [3]Department of Molecular, Cellular, and Developmental Biology, University of Michigan. Ann Arbor, MI 48109. *Corresponding author.

Abstract

An essential step of single-cell RNA sequencing analysis is to classify specific cell types with marker genes in order to dissect the biological functions of each individual cell. In this study, we integrated five published scRNA-seq datasets from the *Arabidopsis* root containing over 25,000 cells and 17 cell clusters. We have compared the performance of seven machine learning methods in classifying these cell types, and determined that the random forest and support vector machine methods performed best. Using feature selection with these two methods and a correlation method, we have identified 600 new marker genes for 10 root cell types, and more than 70% of these machine learning-derived marker genes were not identified before. We found that these new markers not only can assign cell types consistently as the previously known cell markers, but also performed better than existing markers in several evaluation metrics including accuracy and sensitivity. Markers derived by the random forest method, in particular, were expressed in 89-98% of cells in endodermis, trichoblast, and cortex clusters, which is a 29-67% improvement over known markers. Finally, we have found 111 new orthologous marker genes for the trichoblast in five plant species, which expands the number of marker genes by 58-170% in non-Arabidopsis plants. Our results represent a new approach to identify cell-type marker genes from scRNA-seq data and pave the way for cross-species mapping of scRNA-seq data in plants.

Plant-Microbe Interfaces: Metaproteomics reveals shifts in microbial activity and regulation in artificially assembling communities as they respond to environmental variation

Paul E. Abraham[1]* (abrahampe@ornl.gov), Him Shrestha,[1,2] Manasa Appidi,[1,2] Manuel I. Villalobos Solis,[1] Jia Wang,[1] Dana Carper,[1] Leah Burdick,[1] Dale Pelletier,[1] Robert Hettich,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; and [2]The University of Tennessee-Knoxville, 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.

Populus interacts with numerous microbes that have been shown to enhance its health, growth, and development. While PMI researchers have observed distinct microbiome compositions that associate with the *Populus* rhizosphere and root endosphere, little is known about the dynamic molecular mechanisms that influence microbial community assembly across distinct environments. As environmental conditions fluctuate (e.g., nutrients, pH), individual microbes will alter their cellular behavior and this will ultimately impact how they interact with other community members. Unfortunately, studying microbe-microbe interactions and behaviors in a natural setting is challenging, because of the complexity of natural ecosystems and the limited control over environmental factors. A strategy to approach this major challenge relies on addressing community complexity in a progressive manner, by first building a detailed understanding of relatively simple subsets of the community and then achieving high predictive power through combining different building blocks (e.g., hosts, community members, and environmental factors). Emergent properties observed in simpler communities can then be re-evaluated as more complex systems are studied and, when a particular property becomes less relevant, then higher-order interactions can be identified.

In this study, ten taxonomically diverse bacterial members of the *Populus deltoides* rhizosphere were co-cultured either in a complex (R2A) or minimal glucose (MOPs+glucose) medium. Metaproteomics were utilized to understand community assembly, structure, adaptation, and

functionality. Microbial structure was assessed, and results showed that microbial relative abundance stabilizes after few passages and the community is dominated by few microbial strains. In general, we observed two different stable microbial community compositions between the two media types tested. Proteome-wide analyses revealed that *Pseudomonas* was the dominant member in both media, and expressed proteins involved in the synthesis of antibiotics, toxins, and siderophores. Pairwise-microbe interaction data reinforced these antagonistic properties of *Pseudomonas* by showing a presence of zone inhibition for most members of the community.

As members of the community are challenged with unfavorable conditions created either intentionally by other members of the community (e.g., production of antibiotics from *Pseudomonas*) or unintentionally through the collective activity of community metabolism (e.g., limited nutrients, lower oxygen levels, etc.), microbial persistence requires the expression of necessary coping mechanisms. The proteome-wide analysis of *Pantoea* and *Bacillus,* which maintained relatively high protein biomass across the rich media passages, revealed two distinct behavioral changes used to cope with the changing environments. In brief, *Pantoea* was observed to effectively adapt to the changes in the communities of R2A media by increasing the abundance of proteins related to antibiotic resistance and motility as well as shifting metabolism from aerobic to anaerobic processes. *Bacillus*, on the other hand, sporulated as an adaptive response to nutritional deprivation and environmental stress.

Overall, metaproteomics provided an accurate assessment of community assembly across two distinct growth media and improved our understanding of the molecular mechanisms associated with dynamic microbe-microbe interactions and behaviors across changing communities and environments. For this 10-member artificial community, *Pseudomonas* acted as a detrimental member of the community through the production of proteins responsible for the synthesis of antibiotics, toxins, and siderophores. In response to unfavorable conditions, other members maintained relatively high abundances by adapting or becoming dormant. Moving forward, these findings will be leveraged to predict the behavior of more complex systems.

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: Simplified community approach to investigate multiple levels of selection in a host-microbiome relationship

Dana L. Carper[1] (carperdl@ornl.gov), 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.

The long-lived woody perennial *Populus* harbors a diverse consortium of microbial associates. To gain insight into the complex *Populus* host and microbial interactions, we developed a synthetic community system that employs subsets of bacteria isolated from *Populus*. Selection of bacterial communities is controlled by multiple interacting factors including environment, host genetics, and bacterial genetics. To explore some of these factors, we constructed two types of synthetic communities, (1) a highly diverse 150-member community to examine the interaction of host genetics and environment, and (2) two communities (*Variovorax* and *Rhizobium*) of closely related bacterial strains to examine how bacterial genetics determine colonization of host organ specificity. These communities were used to inoculate double autoclaved soil containing host plants. For the 150-community we used *Populus trichocarpa, Populus tremula x alba,* and *Arabidopsis thaliana*, to test the effect of host species and exposed them to three environmental perturbations (control ambient temperature, warming, and cold). We used 16S rRNA sequencing to determine which strains colonized the hosts and how the environment affects community structure. We found strong selection of the bacterial community based on host genus (*Populus* and *Arabidopsis*, 28.4%), host species (10.8%), and environmental perturbations (8.9%). *Arabidopsis* samples were mainly colonized by a *Pantoea* species, while in *Populus* a *Rhodanobacter* and a *Mycobacterium* species were the main constituents of the community. To examine what genetic traits lead to colonization on the bacterial side, we inoculated *Populus trichocarpa* and *Populus deltoides* plants with a community of *Rhizobium* strains and *Variovorax* strains. We used metagenomic sequencing to determine which strains colonized either the

rhizosphere (region around the roots) or the endosphere (interior region of the roots). We then used a comparative genomics approach to determine what genetic traits could lead to colonization of a specific compartment. We found 4 strains of *Variovorax* (out of 28), 32 strains of *Rhizobium* (out of 82) enriched in the endosphere, 9 strains of *Variovorax,* and 32 strains of *Rhizobium* enriched in the rhizosphere. In general genes involved in fatty acid biosynthesis, lipid transport and metabolism, and amino acid transport were enriched in the rhizosphere regardless of bacterial species. These experiments show the utility of synthetic communities to answer questions on multiple environmental scales.

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: Beneficial mycorrhization mediated by a lectin receptor-like kinase

Jay Chen[1*] (chenj@ornl.gov), Wellington Muchero,[1] Zhenzhen Qiao,[1] Timothy B. Yates,[1,2] Him K. Shrestha,[1,3,4] Nancy L. Engle,[1] Amy Flanagan,[5] Jennifer L. Morrell-Falvey,[1] Yali Sun,[1] Timothy J. Tschaplinski,[1] Paul E. Abraham,[1,4] Jessy Labbé,[1] Zeng-Yu Wang,[5] Robert L. Hettich,[1,4] Gerald A. Tuskan,[1] and **Mitchel J. Doktycz[1]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN; [3]Genome Science and Technology, University of Tennessee, Knoxville, TN; [4]Chemical Science Division, Oak Ridge National Laboratory, Oak Ridge, TN; and [5]Noble Research Institute, Ardmore, OK

http://PMIweb.ornl.gov

Project Goals: The goal of the Plant-Microbe Interfaces (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/fungi can establish mutualistic relationships with host plants, providing a large variety of nutritive and protective compounds in exchange for photosynthesized sugars. However, the molecular signals mediating the establishment of these beneficial relationships remain unclear. Our previous genetic mapping and whole-genome resequencing studies identified a gene deletion event of a lectin receptor-like kinase gene *PtLecRLK1* in *Populus* that was associated with poor root colonization by the ectomycorrhizal fungus *Laccaria bicolor.* By introducing *PtLecRLK1* into a perennial grass known to be a non-host of *L. bicolor*, switchgrass (*Panicum virgatum L.),* we found that the *L. bicolor* colonizes the *PtLecRLK1* transgenic switchgrass roots, which illustrates that introduction of *PtLecRLK1* has the potential to convert a non-host to a host of *L. bicolor*. Further transcriptomic and proteomic analyses on inoculated transgenic switchgrass root samples revealed genes/proteins overrepresented in the mutualistic interaction and underrepresented in the pathogenic defense pathway, consistent with the view that pathogenic defense response is downregulated during mutualistic interaction. Metabolomic profiling revealed that root colonization in the transgenic switchgrass was associated with an increase in N-containing metabolites and a decrease in organic acids, sugars, and phenolics, including hydroxycinnamate conjugates, which are often seen in the early steps of establishing mutualistic interactions in compatible partners. This work illustrates that *PtLecRLK1* is able to render a plant susceptible to colonization by the ectomycorrhizal fungus *L. bicolor* and sheds

light on engineering mycorrhizal symbiosis into a non-host to enhance plant productivity and fitness on marginal lands.

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]* (christels@ornl.gov), Julie E. Chavez,[1] Paul E. Abraham,[1] Alyssa A. Carrell,[1] Sara S. Jawdy,[1] Leah Burdick,[1] Dawn M. Klingeman,[1] Melissa A. Cregger,[1] Joshua K. Michener,[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.

Horizontal Gene Transfer (HGT) is one of the main drivers of prokaryotic evolution. HGT is particularly common in the rhizosphere, due to the abundance and diversity of microbes and niches. Plants and microbes are continuously engaged in an evolutionary arms race. While the slower-growing plants benefit from nurturing useful microbes and deterring antagonistic ones, shorter microbial generation times provide an opportunity for microbes to evade host controls. Horizontal transfer of pathways that regulate plant-microbe interactions are one of the fastest evolutionary paths. However, many details about the ecological and evolutionary impact of this transfer remain unclear. In this study, we assess the breadth, rates, and consequences of HGT in a synthetic soil bacterial community of low complexity, both in vitro and in situ using *Populus trichocarpa* as a model host.

In a first set of experiments, we tracked the abundance and association of a conjugative plasmid in a constructed bacterial community in vitro, using the Hi-C technique to crosslink plasmid DNA to the genome of its bacterial host. When employing capture probes that enrich for genome/plasmid interactions, preliminary data suggest that the target plasmids can be tracked even when harbored only transiently and by a very low percentage of the total microbial population (<0.01%). Analysis of plasmid association time courses will elucidate the dynamics and network effects of its transfer. By varying the constituents of the microbial community, the environmental conditions, and the identity of the plasmids being tracked, we hope to increase our understanding of the mechanics and impact of this process in nature.

In a second approach, we seek to investigate the consequences of successful HGT events in the rhizosphere. Plants such as *Populus* are hypothesized to exude salicin and related compounds to

modulate its microbiome. Several *Pseudomonas* spp. isolated from *Populus* roots were engineered to obtain energy and carbon from salicyl alcohol, a degradation product of salicin. To determine how the acquisition of such a pathway could alter a strain's niche and/or behavior, both engineered and wild-type *Pseudomonas* strains were labeled with DNA barcodes and inoculated onto otherwise sterile plant roots. DNA barcode sequencing was very sensitive, even for rhizosphere samples, and so localization and abundance of the barcoded strains could be tracked with high spatial resolution. These and future experiments with different substrate pathways will provide insights into the evolutionary and ecological consequences of gene acquisitions by HGT on plant roots.

Combining our two approaches, we aim to determine the holistic effects of MGEs in complex, diverse environments, and investigate microbial evolutionary responses to mechanisms by which plants shape the rhizosphere 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: Temporal dynamics of the *Populus* microbiome across scales

Nicholas C. Dove[1]* (dovenc@ornl.gov), Alyssa A. Carrell,[1] Dawn M. Klingeman,[1] Melissa A. Cregger,[1,2] Christopher W. Schadt,[1,3] and **Mitchel J. Doktycz**[1]

[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 of 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 on the length and time scales of natural systems, is a challenge that requires long-term fundamental research. Using amplicon and metagenomic sequencing of the plant microbiome in combination with ecological assembly and source tracking models, we work to characterize alterations in the *Populus* microbiome 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.

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. Over the course of a year after planting in the spring of 2017, we collected soils, roots, and leaves from these genotypes to assess microbial community assembly patterns and processes. We found that the initial assembly of the *Populus* microbiome is time-, genotype-, and habitat-dependent, and is moderated by both selective and stochastic factors. We hypothesize that the initial assembly of the plant microbiome may establish the trajectory for forthcoming microbiome states (via priority effects) and could

determine the overall future health of the plant. We have continued to collect and analyze samples from this site, now in its 4th year, to address this hypothesis and determine intra- and inter-annual changes in the *Populus* microbiome and how they vary with the health of the plant.

The second project further characterizes the initial *Populus* microbiome assembly as moderated by an ecologically significant disturbance—wildfire. For this project we took advantage of a high-intensity prescribed fire in Central Utah and collected soils, roots, and leaves of < 1 y old regenerating *P. tremuloide*s saplings in control, moderate, and severely burned areas. We found that fire severity influences the relative dominance of microbial inoculum and the vertical inheritance of the sapling microbiome from the parent tree resulting in changes to the leaf microbiome, notably a three-fold increase in fungal pathogens with increasing severity. Overall, this work demonstrates, for the first time to our knowledge, that fire impacts the plant microbiome, outside of the bulk soil and rhizosphere and highlights potential for further research towards increasing plant fitness and ecosystem recovery after fire events.

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 ha) 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 these clonal stands, we will be 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. In the initial year of study, we characterized the age, genotype, and soils of >200 trees in multiple clonal stands, and our first microbiome sampling campaign is planned for summer 2021.

Integrating these three projects we work to understand the temporal changes in the *Populus* microbiome from seasons to centuries (Figure 1).

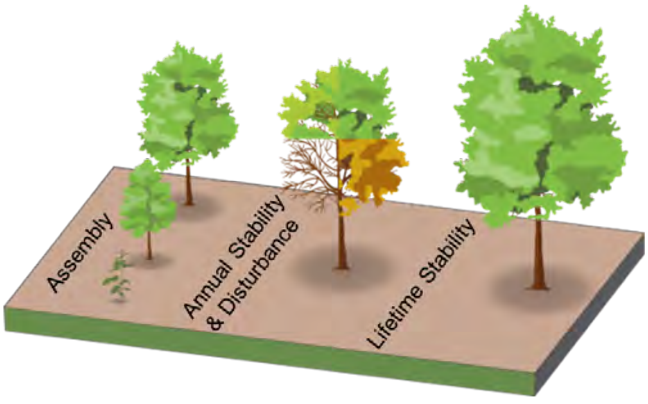


Figure 1: Temporal scales of the *Populus* microbiome.

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: Comparative genome evolution of *Populus* root endophytes

Yi Hong Ke[1]* (yi.hon.ke@duke.edu), Brian Looney,[1] Alejandro Rojas-Flechas,[2] Jake Nash,[1] Khalid Hameed,[1] Christopher Schadt,[3] Jose-Eduardo Marques Galvez,[4] Francis Martin,[4] Jessy Labbé,[3] Daniel Jacobson,[3] Claire Veneault-Fourrey[4], Igor Grigoriev,[5] Kerrie Barry,[5] Rytas Vilgalys,[1] and **Mitchel J. Doktycz**[3]

[1]Duke University, Durham, NC; [2]University of Arkansas, Fayetteville, AR; [3]Bioscience Division, Oak Ridge National Laboratory, Oak Ridge, TN; [4]INRAE Centre de Nancy, Champenoux, France; and [5]Joint Genome Institute, Berkley, CA

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.

Endophytic fungi have a strong influence on ecosystems by benefiting the health and survival of their host plants against biological and non-biological stresses. Fungal endophytes are polyphyletic and hyperdiverse, yet their common adaptations that enable an endophytic lifestyle are largely unknown. We sequenced genomes of over 40 fungal taxa representing 17 unique lineages of endophytes associated with *Populus*. We compared the 17 endophyte lineages with their closest-related free-living saprobe species to identify common genomic features associated with endophytic symbiosis lifestyle. Among the genomic features we compared, endophytic fungi were observed to consistently have significantly greater genome size and gene count than their sibling taxa, including more CAZymes and more small secreted proteins. To correct bias from the phylogenetic structure, we used phylogenetic linear regression models and phylogenetically independent contrasts to compare genome evolution of each lineage with its nearest non-symbiotic relative. Results of multiple phylogenetically-corrected analyses suggest several mechanisms by which fungi have adapted to an endophytic lifestyle. Most fungal endophytes generally have slower growth rates, lower competitive ability (as free-living fungi), while still maintaining a certain amount of saprobic ability as well as ability to communicate with their plant host. To further differentiate the exact genes contributing to endophytic lifestyle, we used an AIC model selection to identify a core set of genes discriminating endophytes from their closely-related saprobes. Six CAZymes, CE12, CBM67, GH39, CBM32, CBM38, GH1, were consistently identified as key

genes regardless of phylogenetic correction method used. These genes are likely to be the core genes that enable the endophytic lifestyle, and further investigations in other plant systems is required to verify their universality.

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-Microbes Interfaces: *In situ* chemical monitoring and imaging of *Populus* and its root microbiome using engineered devices with a porous membrane

Muneeba Khalid[1*] (khalidm@ornl.gov), Jennifer L. Morrell-Falvey,[1] Amber B. Webb,[1] Sara Jawdy,[1] John F. Cahill,[1] Scott T. Retterer,[1,2] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, and [2]Center for Nanophase Materials Sciences, 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 deeper understanding of complex biological systems demands spatial and temporal profiling of the chemical signals and metabolites that drive organization and function. Conventional chemical imaging technologies often rely on destructive techniques for sampling, preventing ongoing tracking of biological systems over time. The ability to observe chemical processes *in situ*—chemotaxis or signaling, for example—using nondestructive chemical imaging offers a new direction to understand the dynamics of developing biological systems. Therefore, we propose a promising pathway for non-destructively imaging chemical information within the microenvironment of complex living systems through space and time. A microfluidic device featuring a microchannel design, microporous membrane and a supporting gasket presents a viable tool for online chemical imaging of a biological setup. As part of the Plant-Microbe Interface project, we are interested in studying chemical interactions between *Populus* and its associated root microbiome. Currently, we are mapping the distribution, growth and motility of select microbes as well as growing *Populus trichocarpa* cuttings in our unique membrane-based microfluidic platforms. Our ultimate goal is to trace microbial distribution along the root and correlate this with the chemical environment in order to better understand the impact of the microbiome on plant growth promotion, nutrient uptake and stress tolerance.

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: Functional characterization of *Populus* proteins involved in plant-microbe interactions

Haiwei Lu[1]*(luh2@ornl.gov), Guoliang Yuan,[1] Tao Yao,[1] Paul Abraham,[1] Changtian Pan,[2] Yiping Qi,[2] Claire Veneault-Fourrey,[3] Jin-Gui Chen,[1] Wellington Muchero,[1] Xiaohan Yang,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]University of Maryland, College Park, MD; and [3]Université de Lorraine, 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.

Plants, including woody species, live in a microbe-rich environment under natural conditions, and interact with a diverse range of microbes. The outcome of plant-microbe interactions spans the full spectrum from beneficial to pathogenic. With a focus on *Populus* – an ecologically and industrially important tree species whose genome encodes hundreds of small proteins with unknown functions (Yang et al., 2011), we characterized plant proteins involved in interactions with beneficial and pathogenic microbes, respectively, with the goal of pushing the outcomes toward enhanced beneficial impact on the plant host. One aspect of our project focuses on examining the role of *P. trichocarpa* small secreted proteins (PtSSPs) in mutualistic symbiosis. By screening of 417 putative PtSSP-encoding genes that have been shown to be responsive to the mutualistic fungal symbiont *Laccaria bicolor* (Plett et al., 2017), we identified a total of 14 PtSSPs with DNA-binding ability. Focusing on two of the 14 PtSSPs, we performed yeast two hybrid (Y2H) experiments to identify their protein targets in *L. bicolor*. Meanwhile, we performed *Agrobacterium rhizogenes* mediated hairy root transformation and *Agrobacterium tumefaciens* mediated stable transformation, respectively, to overexpress individual PtSSP encoding genes fused to variants of the green fluorescent protein (GFP) gene in the hybrid poplar clone INRA 717-1B4 (*P. tremula* × *P. alba*). We are currently screening for transgenic poplar events and/or poplar roots showing high GFP expression to set up plant- *L. bicolor* co-culture experiments to examine the secretion and movement of the PtSSPs.

In addition, extensive transcriptome changes are often observed in response to inoculation with fungi and bacteria, and efficient manipulation of plant-microbe interactions likely requires the ability to modify the expression of multiple plant genes simultaneously. Therefore, another focus of our project is to develop a multiplex gene regulation system by employing CRISPR activation (CRISPRa), CRISPR interference (CRISPRi), RNA interference (RNAi), and their combinations. Using *P. deltoides* WV94 – a species naturally resistant to the invasive fungal pathogen *Sphaerulina musiva*, we are testing the efficiency of CRISPRa by targeting the orthologs of the G-type lectin receptor-like protein kinase encoding gene *Potri.005G018000*, which confers susceptibility to *S. musiva* (Muchero et al., 2018). We are also developing a protoplast-based system to facilitate the screening of efficient single guide RNAs (sgRNAs) and their combinations.

1. Yang X, Tschaplinski TJ, Hurst GB, Jawdy S, Abraham PE, Lankford PK, et al. (2011). Discovery and annotation of small proteins using genomics, proteomics, and computational approaches. *Genome Research*, 21(4), 634-641.
2. Plett JM, Yin H, Mewalal R, Hu R, Li T, Ranjan P, et al. (2017). *Populus trichocarpa* encodes small, effector-like secreted proteins that are highly induced during mutualistic symbiosis. *Scientific Reports*, 7(1), 1-13.
3. Muchero W, Sondreli KL, Chen JG, Urbanowicz BR, Zhang J, Singan V, et al. (2018). Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant–pathogen interactions in a tree. *Proceedings of the National Academy of Sciences*, 115(45), 11573-11578.

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: Exploring plant-microbe associations using Random Walk with Restart on multiplex networks

Piet Jones,[1,2] J. Izaak Miller[2]* (millerji@ornl.gov), David Kainer,[2] Anna Furches,[1,2] Manesh Shaw,[2] Wellington Muchero,[2] Jay Chen,[2] Doug Hyatt,[2] Mirko Pavicic,[2] Hari B. Chhetri,[2] Dan Jacobson,[1,2] and **Mitchel J. Doktycz**[1,2]

[1]University of Tennessee Bredesen Center for Interdisciplinary Research and Graduate Education, 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.

'Omics data provide insights into specific aspects of an organism or population, while combining multi-omics data enables the holistic study of complex biological systems including plant-microbe symbioses. Here we use a multi-omics approach to identify genes that affect the relationship between host (*Populus trichocarpa*) and microbial taxa from the host microbiome. Samples of xylem tissue were taken from a *P. trichocarpa* GWAS population, arrayed in common gardens. First, RNA-Seq was performed on the samples. Then, reads that did not map to *P. trichocarpa* reference (V.3.0) were classified using Parakraken against all publicly available genomes. Reads that were classified as genera in the PMI culture collection and host genes that were significantly associated to these genera via a genome wide association study were selected for further investigation. We used a multi-omics approach, building eight gene-gene networks from experimental datasets and publicly available networks. Each network models a different type of biological relationship between genes or gene-products and forms a single layer in a multiplex network. We then used a Random Walk with Restart Lines of Evidence algorithm (RWR-LOE) to explore the multiplex network, starting from the set of genes of interest. In the RWR-LOE approach, we use the principle of network-topology-based-association to identify genes that fall below the initial significance threshold, but are nonetheless associated with the genes-of-interest. We then used a series of Monte Carlo simulations to identify the subset of genes from the initial genes of interest that are highly functionally related.

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: Characterizing the perception of lipochitooligosaccharides signaling in fungi

Tomás Allen Rush[1*] (rushta@ornl.gov), Erica Teixeira Prates.[1] Omar N. Demerdash,[1] Chris Ellis,[1] Manesh Shah,[1] Joanna Tannous,[1] Daniel Jacobson,[1] Jessy Labbé,[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.

Lipochitooligosaccharides (LCOs) are signaling molecules with various moieties that were thought to be uniquely produced by symbiont microbes (rhizobia bacteria and mycorrhizal fungi) to communicate and colonize their hosts (Liang et al., 2014). A paradigm shift in this understanding was recently initiated with the discovery that LCOs are produced across the whole Fungi kingdom and are responsible for several changes in fungal growth patterns (Rush et al., 2020). However, the alternate roles of LCOs and how fungi perceive them remain largely unknown. Focusing on the *Populus* ectomycorrhizal associate *Laccaria bicolor* and the soilborne and opportunistic human fungal pathogen, *Aspergillus fumigatus,* as pilot fungal systems, we use a known mechanism of LCO recognition in plants to identify similarities in LysM domain receptors. We apply a computational workflow of extensive molecular dynamics simulations and machine-learning methods for binding affinity prediction of LysM-LCO complexes to investigate the molecular basis of the specificity towards different LCO molecules. Candidate binding sites were identified based on common structural features assembled for the optimum binding of the hydrophobic acyl chain and the carbohydrate moieties (and substituents) that constitute the LCOs. In addition, we provide a bioinformatic network map to highlight the commonality of LysM receptors across multiple kingdoms. The selected LysM candidates are being validated by genetic editing of mutants deficient in those LysM receptors. This pioneering work contributes to the understanding of the cross-talk and their signals between microbes, their host, and how this structures communities.

References/Publications

1. Liang, Y., Tóth, K., Cao, Y., Tanaka, K., Espinoza, C., & Stacey, G. (2014). Lipochitooligosaccharide recognition: an ancient story. *New Phytologist*, *204*(2), 289-296.
2. Rush TA, Puech-Pagés V, Bascaules A, Jargeat P, Maillet F, Haouy A, QuyManh-Maës A, Carriel CC, Khokhani D, Keller-Pearson M, Tannous J, Cope KR, Garcia K, Maeda J, Johnson C, Kleven B, Choudhury QC, Labbé J, Swift C, O'Malley MA, Bok JW, Cottaz S, Fort S, Poinsot V, Sussman MR, Lefort C, Nett J, Keller NP, Bécard G, & Ané JM (2020) Lipo-chitooligosaccharides as regulatory signals of fungal growth and development. *Nature Communications* 11, 3897

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 Pilot Project: Defining gene function and its connection to ecosystem processes

Christopher W. Schadt[1*] (schadtcw@ornl.gov), Melanie A. Mayes,[2] 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

Project Goals: The Bio-Scales Pilot Project focuses on understanding how genes influence traits and ecosystem-level processes. Initial efforts examine 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 will sample a subset of *Populus* genotypes with differential phenotypes related to N transformation and biomass production, sequence their associated microbiomes (in collaboration with the Joint Genome Institute), and collect environmental metadata. Our second objective is to prepare data and metadata for integration and analysis using the National Microbiome Data Collaborative and other tools and approaches (e.g., GWAS, network analysis) to predict how plant traits influence soil microbiome functions and how together the plant and the microbiome modify ecosystem processes.

The new Bio-Scales pilot project aims to rapidly determine gene functions and traits and how they scale to influence ecosystem-level processes. Our current work hypothesizes that specific combinations of host 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 with the rhizosphere. Key plant chemical traits that influence the plant microbiome, rhizosphere biogeochemical processes, and subsequent rhizophere 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. 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. The resulting data and metadata in our study will be produced 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 data set 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.

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.

Plant-Microbe Interfaces: Structural basis for a bacterial Pip system plant effector recognition protein

Shukun Luo,[1] Bruna G. Coutinho,[2] Prikshat Dadhwal,[1] Yasuhiro Oda,[2] Jiahong Ren,[3] Amy L. Schaefer[2*] (amyschae@uw.edu), Dale A. Pelletier,[4] **E. Peter Greenberg,[2] Caroline S. Harwood,[2] Liang Tong,**[1] and **Mitchel J. Doktycz**[4]

[1]Department of Biological Sciences, Columbia University, New York, NY; [2]Department of Microbiology, University of Washington, Seattle, WA; [3]Faculty of Biological Science and Technology, Changzhi University, Shanxi, China; and [4]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 of plants. We identified an ethanolamine derivative, *N*-(2-hydroxyethyl)-2-(2-hydroxyethylamino) acetamide (HEHEAA), as a potent effector of PipR-mediated gene regulation in the plant endophyte *Pseudomonas* GM79. HEHEAA-dependent PipR activity requires an ATP-binding cassette-type active transport system, and the periplasmic substrate-binding protein (SBP) of that system binds HEHEAA. To begin to understand the molecular basis of PipR system responses to plant factors, we crystallized a HEHEAA-responsive SBP in the free- and HEHEAA-bound forms. The SBP, which is similar to peptide-binding SBPs, was in a closed conformation. A narrow cavity at the interface of its two lobes is wide enough to bind HEHEAA, but cannot accommodate peptides with side chains. The polar atoms of HEHEAA are recognized by hydrogen-bonding interactions, and additional SBP residues contribute to the binding site. This binding mode was confirmed by a structure-based mutational analysis. We also show that a closely related SBP from the plant pathogen *Pseudomonas syringae* pv *tomato* DC3000 does not recognize

HEHEAA. However, a single amino acid substitution in the presumed effector-binding pocket of the *P. syringae* SBP converted it to an HEHEAA-binding protein. The *P. syringae* PipR depends on a plant effector for activity and our findings imply that different PipR-associated SBPs bind different effectors.

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: Protein post-translational modifications in a 10-member microbial community from the *Populus* rhizosphere offer critical insights into how microbes adapt to changing environments

Manuel I. Villalobos Solis[1*] (villalobosmi@ornl.gov), Him Shrestha,[1,2] Manasa Appidi,[1,2] Robert L. Hettich,[1] Paul E. Abraham,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; and [2]The University of Tennessee-Knoxville, 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.

Bacterial members within microbial communities that associate with the *Populus* rhizosphere adapt to their environmental conditions not only by modulating the abundance of the proteins they produce, but also by regulating the functions of these molecules. Post-translational modifications (PTMs) of proteins are one of the most important, yet understudied mechanisms that microbes use to rapidly activate, alter or suppress protein functions. Optimized bioinformatic pipelines for metaproteomics exploiting *de novo* peptide sequencing now allow researchers to broadly characterize PTMs in proteins extracted from bacterial isolates and communities. Here, we used this metaproteomics capability to interrogate the PTM landscape from a simplified artificial community consisting of ten predominant bacterial members in *Populus deltoides* rhizosphere co-cultured in complex and minimal glucose media.[1] The advantage of this synthetic community is its low complexity, high controllability, and reproducibility in a laboratory setting from which molecular insights that are closely representative of how bacteria behave in their natural environments can be obtained.

Briefly, equal volumes of *Bacillus* sp. Bc15, *Caulobacter* sp. AP07, *Duganella* sp. CF402, *Pantoea* sp. YR343, *Paraburkholderia* sp. BT03, *Pseudomonas* sp. GM17, *Rhizobium* sp. CF142, *Sphingobium* sp. AP49, *Streptomyces* sp. YR139, and *Variovorax* sp. CF313 with the same normalized cell densities were mixed in either a complex medium or minimal medium supplemented with 0.2% glucose, harvested after 48-hrs, then transferred into fresh media by

diluting 1:10 for a total of 15 passages in biological triplicates. From these 15 passages, a metaproteomics workflow was applied to study PTMs.

Our results showed that approximately 18% of proteins identified for all organisms at both rich and minimal media conditions had PTMs with biological significance. The most frequently occurring types of PTMs were methylations, dehydrations, and oxidations/hydroxylations. Such modifications were part of proteins impacting prominent cellular processes, including translation, ribosomal structure and biogenesis, energy production and conversion, protein turnover and chaperon functions, as well as amino acid metabolism and transport. Protein-centric analyses highlighted two types of PTM regulation: static modifications that are essential for proper protein function; and dynamic, reversible modifications that alter the functional or structural state of proteins. Two representative examples of these types of PTM-based regulations were a static β-methylthiolation modification localized on a universally conserved aspartic acid residue in bacterial ribosomal proteins S12; and a dynamic lysine methylation only present in the sequences of elongation factor thermo unstable proteins for some bacterial members being impacted by nutrient depravation.

Overall, this study demonstrates that high-resolution mass spectrometry not only affords the ability to broadly characterize PTMs in a biological system, but that it also provides a level of sensitivity capable of revealing regulatory mechanisms influencing the activity of single proteins.

References/Publications

1 Brown, S. D. *et al.* Twenty-one genome sequences from Pseudomonas species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of Populus deltoides. *J Bacteriol* **194**, 5991-5993, doi:10.1128/JB.01243-12 (2012).

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: Formation and characterization of emergent microbial communities

Jia Wang[1]* (wangj@ornl.gov), Dana L. Carper,[1] Leah H. Burdick,[1] Him K. Shrestha,[1,2] Manasa R. Appidi,[1,2] Paul E. Abraham,[1] Collin M. Timm,[3] Robert L. Hettich,[1] Dale A. Pelletier,[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]Research and Exploratory Development Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD

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.

Microbial communities colonize throughout plant tissues and contribute to host function. How these communities form and how individual members contribute to shaping the community structure are not well understood. Synthetic microbial communities composed of well-studied individual isolates can be valuable model systems for elucidating the organizational principles of communities. Using genome-defined strains, systematic analysis by computational modeling can lead to mechanistic insights and metabolic interactions among species. In this study, 10 bacterial strains isolated from the *Populus deltoides* rhizosphere were co-cultured and passaged in two different media environments to form stable microbial communities. The membership and relative abundances of the individual strains in the resultant community stabilized after around 5 passages and resulted in only a few dominant strains that depended on the medium. To unravel the underlying metabolic interactions within the community, flux balance analysis was used to model microbial growth and predict metabolic interaction involved in organizing the microbial communities. These analyses were complemented by measuring growth curves of the individual strains, performing metaproteomics of the community and carrying out pairwise interaction screens between species. A fast growth rate can be advantageous for maintaining survival in the microbial community, and the final presence of a member also depends on selective antagonistic relationships and metabolic exchanges within the community. Revealing the mechanisms of interaction among plant-associated microorganisms provides insights into strategies for engineering microbial communities that can potentially increase plant growth and disease

resistance. Deciphering the membership and metabolic potentials of microbial communities will also enable the design of synthetic consortia with desired biological functions.

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: Microbial responses to the chemical environment of the rhizosphere

Amber B. Webb[1]* (biblean@ornl.gov), Manasa R. Appidi,[2] Paul E. Abraham,[1,2] Sara Jawdy,[1] Jennifer Morrell-Falvey,[1,2] and **Mitchel J. Doktycz**[1,2]

[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.

Understanding the interactions among plants and bacteria within the rhizosphere and the role of the chemical environment in shaping these interactions is a central component of the PMI SFA. The ability to visualize and manipulate these interactions in real-time is an invaluable tool in developing our understanding of the mechanisms that regulate these interactions. While prior work has focused on endpoint analysis or imaging individual plants in real time at high resolution using confocal microscopy, we need to develop tools for high throughput, non-destructive imaging of plant-microbe interactions in real time. Here, we describe current efforts to develop such strategies using plants grown in Petri dishes with agar and imaging using a Cytation 5 high content imaging system. With this system, we are optimizing protocols that allow us to assess root colonization by microbes starting with a well-studied isolate, *Pantoea* YR343 labeled with GFP. We are developing these tools as a means of studying the chemical environment of the rhizosphere in the context of a microbial community. These microbial communities are made up of diverse bacterial species that exhibit a variety of sensory capabilities that allow them to respond in different ways to the chemical environment of the rhizosphere. In order to better understand how bacteria respond to the chemical environment of the rhizosphere, we utilized a proteomics approach that resulted in the identification of many different proteins that are upregulated in *Pantoea* YR343 in the presence of a plant. Currently, we are applying high-throughput imaging approaches, combined with genetic tools to develop a robust platform for analyzing the role of select proteins in plant-microbe interactions in the context of the rhizosphere and its local chemical environment.

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.

Characterizing fungal inhibitors in drought-stressed switchgrass

Sarvada Chipkar[1,2], Jailynn Johnson[1], Katie Smith[1], Marissa Gallmeyer[1], Katherine Overmyer[4,5], Arthur Daniel Jones[3,6], Joshua Coon[4,5], 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 been an important research theme in recent years. Lignocellulosic biomass is a major resource that can be used for 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. Previously it was reported that drought experienced during the growth of switchgrass led to complete inhibition of yeast growth during fermentation [1]. In this project, we characterized specific compounds that led to this inhibition. Switchgrass harvested in drought and non-drought years were pretreated using Ammonia Fiber expansion (AFEX). Untreated and AFEX processed samples were extracted using solvents (i.e. water, ethanol, and ethyl acetate) to selectively remove potential inhibitory compounds and determine whether pretreatment affects the inhibition. A key goal of the project was to determine whether the microbial-inhibitors are plant-generated compounds, by-products of the pretreatment process, or a combination of both. High solids loading enzymatic hydrolysis was performed on all samples followed by fermentation using genetically modified, xylose consuming yeast strain *Saccharomyces cerevisiae* Y330. 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.

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). "Inhibition of microbial biofuel production in drought-stressed switchgrass hydrolysate." Biotechnol Biofuels 9: 237.

A High-Efficacy CRISPR Interference System for Gene Function Discovery in *Zymomonas mobilis*

Amy B. Banta[1,2], Amy L. Enright[1,2*] (alenright@wisc.edu), Cheta Siletti[1†], **Jason M. Peters**[1,2,3,4]

[1]DOE Great Lakes Bioenergy Research Center, [2]School of Pharmacy, [3]Department of Bacteriology, and [4]Department of Medical Microbiology & Immunology, University of Wisconsin-Madison; [†]Present address: Department of Food Science, University of Wisconsin-Madison

Project Goals: Enable rational engineering of the promising biofuel producer *Zymomonas mobilis* by establishing CRISPR interference (CRISPRi) for gene function discovery in this non-model organism.

The Alphaproteobacterium *Zymomonas mobilis* is a promising biofuel producer due to its streamlined glycolytic conversion of sugar to ethanol. However, rational engineering of *Z. mobilis* for production of advanced biofuels, such as isobutanol, is hindered by an incomplete understanding of *Z. mobilis* gene functions, particularly those involved in metabolism and alcohol stress. Furthermore, essential genes are linked to many of the industrially desirable attributes of *Z. mobilis*, requiring the use of advanced genetic techniques to preserve cell viability upon manipulation of such genes. CRISPR interference (CRISPRi) is a method which allows for assessment of all genes, including essential genes. In CRISPRi, precise gene knockdown is achieved by a programmable guide RNA complexed with a catalytically inactive Cas9 protein to block transcription of a complementary DNA target. CRISPRi is both inducible and titratable, enabling assessment of essential genes via temporal regulation and precise control over knockdown strength to reveal phenotypes while maintaining viability. Here, we report a high-efficacy *Z. mobilis* CRISPRi system and demonstrate its utility through knockdown of genes that are essential for growth, required for the uniquely efficient ethanologenic metabolism of *Z. mobilis*, or involved in isobutanol tolerance. Our *Z. mobilis* CRISPRi system paves a straightforward path to gene function discovery which can be used to improve rational engineering efforts for increased biofuel production by this non-model organism.

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

This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Great Lakes Bioenergy Research Center under Award Number DE-SC0018409.

Disruption of *Brachypodium distachyon* Lichenase Alters Metabolism of Mixed-linkage Glucan and Starch

Mingzhu Fan[1]*(mzfan2012@gmail.com), Jacob K Jensen[1,2], Sang-Jin Kim[1], Jia-Yi Chan[1], Starla Zemelis-Durfee[1], Claudia Beaudry[1], Federica Brandizzi[1], **Curtis G Wilkerson**[1]

[1]Michigan State University, East Lansing; [2]University of Copenhagen, Copenhagen.

https://www.glbrc.org/

Project goals: Engineering bioenergy crops with increased accumulation of an easily fermentable cell wall polysaccharide MLG, without affecting plant growth.

Biomass represents a significant energy resource and as such, it is important to improve the efficiency of its conversion to fuel by altering its chemical makeup. Mixed-linkage glucan (MLG), which is widely distributed in grasses, is an easily fermentable cell wall polysaccharide and ideal compound for the production of biofuels. Previous studies have shown that transgenic plants overexpressing MLG synthases accumulate more MLG but with detrimental growth defects and reduced biomass yields. In this study, we identify a gene encoding a lichenase we have named *Brachypodium distachyon LICHENASE 1* (*BdLCH1*), which is highly expressed in the endosperm of germinating seeds and elongating internodes. RNA in situ hybridization showed that it is also expressed in chlorenchyma cells of mature leaves and stems. *lch1* mutants generated using CRISPR-Cas 9 technology accumulated more MLG in the vegetative tissues examined and did not show significant growth defects. Compared with wild-type plants, disruption of *BdLCH1* resulted in an 8-fold increase in MLG content in senesced leaves. Immunolabeling with MLG monoclonal antibody showed that MLG in parenchyma cells of mature leaves and stems was not removed as was the case in wild-type plants. MLG stability during development was improved in *lch1* mutants.

Transcription of *BdLCH1* was induced by darkness in both wild-type and *lch1* mutants. While MLG content in wild-type plants decreased significantly after dark-incubation, *lch1* plants contained a similar amount of MLG as those grown under a regular light/dark cycle. Dark-incubated 8-week-old *lch1* plants also had a faster rate of starch breakdown than that in wild-type plants. Disruption of *BdLCH1* not only changed MLG accumulation but also altered starch degradation in darkness. These results indicate that engineering bioenergy crops by modifying the expression of lichenases is a promising strategy for biofuel production.

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.

Multiomics analysis of mitochondrial versus cytosolic compartmentalization of the isobutanol pathway in *Saccharomyces cerevisiae*

Francesca V. Gambacorta[1,2]* (fgambacorta@wisc.edu), **Brian F. Pfleger**[1,2], Mary Tremaine[1], Trey K. Sato[1], Ellen R. Wagner[1,3], Audrey P. Gasch[1,3], Tyler B. Jacobson[1,4], Daniel Amador-Noguez[1,4], Laura K. Muehlbauer[1,5], Joshua J. Coon[1,5], Mick A. McGee[1], and Justin J. Baerwald[2]

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison,Wi; [2]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wi; [3]Laboratory of Genetics, University of Wisconsin-Madison, Madison, Wi; [4]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wi; [5]Department of Chemistry, University of Wisconsin-Madison, Madison, Wi

http://glbrc.org

Project Goals: We aim to establish which subcellular compartment in *S. cerevisiae* the isobutanol pathway should be localized to in order to produce isobutanol at high yields.

Isobutanol, a branched-chain higher alcohol, is considered a promising alternative biofuel to ethanol. As the dominant bioethanol producer, *Saccharomyces cerevisiae* has a demonstrated high alcohol tolerance and thus is considered a suitable host for isobutanol production. Our overarching goal is to redirect *S. cerevisiae*'s native metabolism from ethanol to isobutanol production. Efforts in eliminating ethanol production have proven challenging because ethanol synthesis plays an essential role in recycling reducing equivalents necessary for glycolysis and cell growth[1]. In order for us to establish isobutanol as the predominant NADH-oxidizing pathway in yeast, we need to substantially increase *S. cerevisiae*'s ability to turnover this essential cofactor via isobutanol fermentation. A number of different metabolic engineering techniques have been implemented to balance this requirement and thus increase isobutanol flux. One strategy entails localizing the five enzymatic steps involved in isobutanol biosynthesis into a single compartment: either the mitochondria or the cytosol. However, there is debate as to what subcellular compartment is optimal for maintaining flux through the pathway. In previous reports, the mitochondrial localized isobutanol pathway outperformes the cytosolic version, but the cytosolic version is likely to be more robust under industrially relevant anaerobic conditions where the mitochondria enter a minimal energy-requirement mode. Thus, we seek to understand what barriers in the cytosolic strain can be addressed in order for us to develop a platfrom isobutanol producing *S. cerevisiae* strain.

In our work, we have equipped yeast with either the cytosolic or mitochondrial localized isobutanol pathway and performed a multiomics analysis to elucidate why the mitochondrial compartmentalization strategy favors isobutanol production. During the course of the fermentation, we collected samples for metabolomics, proteomics, and transcriptomics analyses at early-exponential, mid-exponential, and early-stationary phase. In agreement with previous reports, the metabolite data showed that the strain harboring the mitochondrial-localized isobutanol pathway outperformed the cytosolic version by 5.5-fold and produced 2.32 mM isobutanol after the 48-hr fermentation. An increase in the upstream isobutanol metabolite levels in the strain with the cytosolic localized pathway suggests that the difference in titer is due to a

bottleneck at the third enzyme in the pathway, dihydroxy-acid dehydratase (DHAD). DHAD requires a [4Fe-4S] cluster which is synthesized by both mitochondrial ISC (iron-sulfur cluster) and cytosolic CIA (cytosolic iron-sulfur cluster assembly) proteins. Thus, we hypothesized that the DHAD activity was impaired when expressed in the cytosol due to this complex, cross-compartmental assembly. Furthermore, the protein and transcript data of the strain showed increased ISC/CIA activity and altered sulfur metabolism; functional enrichment analysis showed proteins involved with cysteine biosynthesis/homocysteine degradation, siroheme biosynthesis, and sulfate assimilation were positively enriched compared to the strain with the mitochondrial localized isobutanol pathway. Taken together, we hypothesize that in the strain with the cytosolically localized isobutanol pathway, DHAD activity is impaired because the CIA machinery insufficiently makes, targets, and/or loads the Fe-S cluster into the cytosolically localized DHAD resulting in non-functional protein. Since the cytosolic strain is likely to be more robust in industrial conditions, our current objective is to address this observed DHAD bottleneck. The knowledge gained from this multiomics study will inform the design of the next isobutanol producing strain.

Publications

1. Gambacorta, F. V., Dietrich, J. J., Yan, Q. & Pfleger, B. F. Rewiring yeast metabolism to synthesize products beyond ethanol. *Current Opinion in Chemical Biology* **59**, 182–192 (2020).

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.

Analysis of Carbon Capture in Lignocellulosic Biorefineries

Caleb H. Geissler[1,2*] (geissler@princeton.edu) and **Christos T. Maravelias**[1,2,3]

[1]Princeton University, Princeton; [2]DOE Great Lakes Bioenergy Research Center, Princeton University, New Jersey; and [3]Princeton Andlinger Center for Energy and the Environment, Princeton, New Jersey

Project Goals: This project aims to use modeling and optimization methods to evaluate the effect of pretreatment method and feedstock selection on the performance of an ethanol biorefinery with carbon capture. Performance is evaluated using economic, environmental, and energetic metrics.

Bioenergy with carbon capture and sequestration (BECCS) is a widely studied technology for global warming mitigation. While industrial power plants that use coal or natural gas emit CO_2 only in flue gas, ethanol biorefineries have multiple point sources of CO_2 emissions: nearly pure CO_2 from fermentation, biogas from anaerobic digestion, and flue gas from solid residue combustion. The relative amounts of CO_2 emitted from each of these point sources depend on what feedstocks and pretreatment methods are used. The solid residues and biogas are typically burnt in biorefineries to meet their energy demand, and excess electricity is sold to the grid. However, at high rates of carbon capture, electricity would have to be purchased, and thus the biorefineries would no longer be energetically self-sufficient.

Industrial power plants have the advantage of economies of scale, but large biorefineries would require transport of feedstocks from farther away, which increases both feedstock cost and emissions. At these large distances, biomass can be transported a short distance to a processing depot where it is densified into pellets prior to transportation to the biorefinery by either truck or rail. This densification at depots decreases the transportation cost increase associated with greater distances. Even with depots, there is significant variation in reported cost-optimal biorefinery capacities, and even more variation for biorefineries with carbon capture. The supply chain is also important for economic and energetic considerations, as emissions from growth, harvesting, and transportation affect the greenhouse gas balance and energy consumption of the entire system.

A superstructure-based optimization model is used that embeds all possible selections of technologies, and solved for the lowest-cost solution. Sensitivity analyses are performed on carbon sequestration credits, carbon capture rates, and biorefinery capacity to evaluate the economic, energetic, and environmental performance of biorefineries using different combinations of feedstocks and pretreatments.

Pretreatment methods and feedstocks that lead to high carbon emissions from fermentation and anaerobic digestion have lower average carbon capture costs. Biorefineries with high biomass to ethanol yields can capture high percentages of the carbon in the feedstock if energy is purchased,

but less excess energy available from residue combustion limits carbon capture when energetic self-sufficiency is required. Biorefineries with low energy requirements and that use feedstocks with high energy content can capture the highest rates of carbon. At high capacities that use processing depots and rail transportation, increasing capacity has minimal impact on the greenhouse gas balance and energy consumption of the system. Depending on the specific configuration of the biorefinery, capture from fermentation only or capture from fermentation and biogas is required to achieve net-negative carbon emissions. In all cases studied, the energy consumption of the supply chain and biorefinery does exceed the energy in the ethanol produced.

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.

Evolution of Cellulose Structure throughout gamma-Valerolactone-Assisted and Enzymatic Biomass Deconstruction

Elise B. Gilcher[1,2]* (gilcher@wisc.edu), Nathaniel Kuch[1,2], Joshua Del Mundo[3], Samantha F. Ausman[1], Catherine F. M. Clewett[1], Esther W. Gomez,[3] Enrique D. Gomez[3], Brian G. Fox[1,2], Thatcher W. Root[1] and **James A. Dumesic[1,2]**

[1]University of Wisconsin, Madison, [2]DOE Great Lakes Bioenergy Research Center, Madison, WI, [3]Pennsylvania State University, University Park

Project Goals: Our goals are to elucidate how cellulosic material is degraded by gamma-valerolacton assisted mild acidolysis and enzymatic hydrolysis with the engineered cellulase, CelR. Overall, we aim to understand how to increase yields of usable sugars from the degredation of different biomass sources.

Abstract text:

Biomass recalcitrance during deconstruction remains a key inhibitor to successful implementation of affordable biomass processing technologies. A clear connection between the cell wall structure and biomass deconstruction is necessary to understand how lignocellulosic material is broken down, and to identify defining features of residual cellulose. In this work, we have employed solid-state ^{13}C cross-polarization magic angle spinning (CP/MAS) nuclear magnetic resonance (NMR) spectroscopy to track domain changes in cellulose microfibrils throughout various chemical and enzymatic hydrolysis treatments. This work builds on previously demonstrated capabilities using NMR to predict enzymatic hydrolysis sugar yields of pretreated cellulose,[1] and moves toward a molecular understanding of how cellulose is degraded by GVL mild acidolysis. Raw biomass was pretreated in GVL-water co-solvents with mild sulfuric acid concentrations at increasing temperatures and subsequently hydrolyzed with an engineered cellulase. Residual solids from each step were characterized using ^{13}C CP/MAS NMR and spectra were fit with subpeaks corresponding to different cellulose microenvironments at the C-4 carbon center. The changes of the C-4 carbon subpeaks were tracked throughout each treatment to understand the physical changes occurring within different cellulose microfibril domains. We show that the xylan-cellulose and inaccessible fibril surface resonances decrease significantly upon hydrolysis with mild acidic GVL-water co-solvent pretreatment. Wide-angle X-ray scattering (WAXS) results support increasing Segal Crystallinity and tighetening of the lattice structure of residual cellulose samples with increasing GVL pretreatment. Enzymatic

hydrolysis leads to further depletion of inaccessible and paracrystalline peaks in NMR, but does not degrade I_β crystalline regions. This behavior can be interpreted as an opening of bound microfibril surfaces previously inaccessible to the solvent through GVL acidolysis and end-on degradation of the residual cellulose by enzymatic hydrolysis. The cleaving of xylan-cellulose linkages and opening of inaccessible fibril surfaces by acid co-solvent pretreatment primes the cellulose for enzymatic attack. This technique can monitor the evolution of structural changes to the cellulosic material and allows for comparison between different cellulose deconstruction methods. It guides larger questions of recalcitrance - mainly the need to hydrolyse unreacted I_β crystalline cellulose - and gives insight to needed improvement of subsequent deconstruction methods based on residual solids structures.

References/Publications

1. Walker, Theodore W., et al. "Solid-State NMR Studies of Solvent-Mediated, Acid-Catalyzed Woody Biomass Pretreatment for Enzymatic Conversion of Residual Cellulose." ACS Sustainable Chemistry & Engineering 8.16 (2020): 6551-6563.

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.

Engineering DXS for Improved Flux into the MEP Pathway

Jonathan Greenhalgh[1,2*] (jgreenhalgh2@wisc.edu), Aaron Lin,[1,2] Haiyang Zheng,[1] and **Philip Romero**[1,2]

[1]University of Wisconsin-Madison, Madison; [2]Great Lakes Bioenergy Research Center, Madison, Wisconsin

https://www.glbrc.org

Project Goals: Using machine learning trained on data from a high-throughput growth selection to design and characterize new DXS enzymes with enhanced activity for improving the overall flux into the MEP pathway.

The MEP pathway synthesizes isoprenoids, which are valuable next generation biofuels and bioproducts. The rate-limiting step of the pathway is the condensation of pyruvate and glyceraldehyde 3-phosphate by 1-deoxy-D-xylulose-5-phosphate synthase (DXS). In this work we engineered new DXS variants to increase flux into the MEP pathway. We designed, constructed, and screened a large library of chimeric DXS enzymes. We used nanopore sequencing to map how DXS sequence affects activity, and trained machine learning models to predict highly active DXS variants. We designed, synthesized, and tested four new DXS variants and confirmed their ability to complement Δ*dxs E. coli* strains. We also measured the designed enzymes activities in vitro and found the best DXS design was comparable to that of the best parental DXS. We will utilize insights from these experiments, combined with directed evolution, to further optimize the DXS variants. Future work will include biochemical and kinetic characterization of the designed enzymes, metabolomics to profile how they alter metabolite pools, and transferring the designed enzymes to other bioenergy organisms such as *Zymomonas mobilis*.

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.

Reinforcement Learning to Optimize Medium Chain Fatty Acid Production from Lignocellulosic Stillage using Anaerobic Microbial Communities

Madeline Hayes[1,2] * (mmhayes4@wisc.edu), Jaron Thompson,[1] Jun Feng,[1,2] **Ophelia S. Venturelli**[1,2]

[1]University of Wisconsin-Madison, Madison, WI; [2]DOE Great Lakes Bioenergy Research Center, Madison, WI

Project Goals: The goal of this research is to apply a data-driven approach to guide the design of microbial communities that maximize production of target metabolites.

Microbes exist in complex communities in every environment on earth, leveraging their flexible metabolic capabilities to perform a wide variety of chemical transformations. Anaerobic fermentative communities perform many functions of interest with collective metabolism, but complex natural communities have proven very difficult to explore due to increasing dimensionality. Understanding the ecological dynamics of these communities has proven challenging; even with the advent of advanced sequencing techniques that allow unprecedented characterization of community composition, frameworks for understanding competition and cooperation as they relate to metabolic flux are still relatively undeveloped.

To this end, the integration of machine learning into biological systems to help inform complex and high-dimensional problems is an emerging technique with particularly useful applications in microbiome and microbial community ecology research. Exploration of complex communities using machine learning to build computational models that can handle high-dimensional problems and predict target metabolic functions holds much promise for engineering productive microbiomes in a wide variety of environments, and for a wide variety of production targets. Medium chain fatty acids (MCFA's, fatty acid chains of length C6-C10) are high value compounds used in industrial processes, including as fuels and specialty pharmaceutical precursors. Currently, these compounds are produced or harvested from non-renewable or unsustainable sources, such as petroleum products or palm oil, and therefore represent a desirable target for sustainable production. Leveraging microbial community metabolism to convert plant-based carbon to these valuable products has proven promising but remains a complex problem as community size and dimensionality increases.

We seek a reduced-complexity synthetic consortium of anaerobic bacteria that optimizes the conversion of carbon from lignocellulosic stillage waste into high-value MCFA products. We propose a design-test-learn approach that leverages principles from reinforcement learning to seek optimal experiments that systematically explore and exploit microbial community functions. This iterative approach involves collecting data, building computational models from the data that predict target production, generating rational hypotheses to narrow future experimental spaces, and testing these hypotheses by collecting more data. The modelling approaches employed here include the generalized Lotka-Volterra model, which predicts community dynamics based on monospecies growth parameters and pairwise interactions, and an ensemble of machine learning methods including linear regression, Gaussian processes, random forest regression, and feed-forward neural networks. These models provide an ensemble of

predictions that serve as features for a final Bayesian linear regression model that provides probabilistic estimates of MCFA production. Previously untested microbial communities that are predicted to maximize MCFA production become candidates for future experiments.

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.

Plant-soil-microbial interactions in detritusphere and its impact on N_2O emission

Alyssa Kim[12,*] (kimkyu46@msu.edu), Jenie Gil[12], Nathaniel Ostrom[12], Hasand Gandhi[12], Maxwell Oerther[12], Yakov Kuzyakov[3], Andrey Guber[12], and **Alexandra Kravchenko[12]**

[1]Michigan State University, East Lansing, MI; DOE Great Lakes Bioenergy Research Center, East Lansing, MI; [3]Georg-August-Universität Göttingen, Göttingen, Germany

Project Goals: Evaluation of the environmental performance in bioenergy cropping systems is an important part for sustainable bioenergy production. The goal of this project is to examine the contribution of decomposing switchgrass roots to nitroux oxide (N_2O) emission. We aimed to track the fate of switchgrass-driven nitrogen (N) in the microbial biomass, soil (organic and inorganic N), and atmosphere (N_2O) under different soil physical conditions. We also aimed to elucidate the temporal and spatial dynamics of microbial activity and N_2O emission and find the key controls of them.

Abstract text: Switchgrass (*Panicum Virgatum L.*) is a promising bioenergy crop that can preserve land from environmental disturbances and play a role in carbon (C) sequestration. For a comprehensive assessment of the ecological and environmental functions of the switchgrass production system, N_2O emissions, especially at the decomposition stage of switchgrass roots (that often occurs after harvest), has to be thoroughly investigated.

We conducted a greenhouse study of switchgrass root decomposition, under various controlled soil conditions. We treated the soils of the same origin (thus expected to have the same microbial community) to have different moisture contents (40% and 70% water-filled pore space) and pore size distributions (dominant pores of >30μm Ø and < 10μm Ø, referred to as large and small pores). For a more realistic assessment, we used *in-situ* grown switchgrass roots instead of incorporating root fragments. It enabled us to include an indirect effect of the microenvironments and microbial community structures formed during the plant growth (rhizosphere). The switchgrass roots were subjected to dual-isotope labeling (^{15}N and ^{13}C) to track the fate of N and C. Isotopic levels were measured during the decomposition to estimate the size of the pools of microbial biomass, soil, and atmosphere. Microbial activity was measured using the zymography approach, which is a 2-dimensional mapping technique of extracellular enzymes. We mapped the distribution of N-acetyl-glucosaminide (chitinase) and separated the chitinase activity on the roots and soils to assess spatial dynamic.

In our results, up to 0.4 % of the switchgrass root-driven N was emitted as N_2O gas, only within 21 days of the decomposition, suggesting the necessity of management practice to mitigate the formation of strong N_2O hotspots after harvest in the bioenergy system. Approximately 21 ~35% of root N was transformed to dissolved organic N, while less than 1 % of the root N remained as ammonium (NH_4^+) and nitrate (NO_3^-) during the incubation. Decreasing NH_4^+ and increasing

NO_3^- suggested nitrification. Surprisingly, inorganic and organic N, enzyme activity, and N_2O emission were greater in the soil with the prevalence of large pores. However, there was no difference in microbial biomass between the soil pore size treatments. Higher chitinase activity in the soils with the prevalence of large pores suggests that the fungal activity was higher in those soils compared to the soils dominated with small pores. Root chitinase activity was positively correlated with the root driven N_2O emission rate ($p< 0.01$, $R^2=0.22$), supporting that the microbial hotspot formed near the root is the hotspots of N_2O emission. The ample supply of labile substrates degraded by extracellular enzymes might be a key control to the magnitude of the N_2O hotspots in detritusphere.

Tracking the fate of N during the plant root decomposition provides a new perspective on the strategies to minimize N_2O emissions in switchgrass bioenergy cropping systems. Our study also indicates that the intensity of root-driven N_2O hotspots can highly depend on the soil's physical characteristics, being mediated by decomposed substances and enzyme activity.

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.

Enzymatic Deconstruction of Cellulosic Biomass

Nathaniel J. Kuch[1]* (nkuch@wisc.edu), Elise Gilcher,[1] Craig A. Bingman,[1] Theodore W. Walker,[1] Kirk A. Vander Meulen,[1] Alex Parker,[1] Mark Kutschke,[1] James A. Dumesic,[1] **Brian G. Fox**[1]

[1]University of Wisconsin-Madison, Madison

Project Goals: The goal of this project is to expand the understanding of enzymatic deconstruction of cellulosic plant biomass by carbohydrate active enzymes on both the atomic and macromolecular level. This project is using solid-state nuclear magnetic resonance (ssNMR), crystallography, and enzyme catalysis to contextualize enzyme hydrolysis and synergistic deconstruction experiments at the molecular level.

Abstract text:

Enzymatic deconstruction of plant biomass has been the focus of much research, especially with increasing interest in using plant biomass as a renewable resource to build various value-added products such as biofuels, bioplastics, and other platform chemicals. Research in this area has not yet produced a model that incorporates detailed, atomic level resolution and understanding across the scope of possible options. For example, the molecular basis for synergy between glycoside hydrolases that target different polysaccharides (e.g., an endo-xylanase and an exo-cellulase) is still not fully understood across the breadth of potential substrates and enzymes. This represents a significant knowledge gap, which this project is working to fill.

We established an analytical index of the interactions between biomass and a model cellulase (CelR) through the strong correlation between the ratio of the two peaks in the split C4 resonance observed by ssNMR (called X_{NMR}) and the final yield of enzymatic hydrolysis [1]. The correlation between X_{NMR} and enzymatic hydrolysis is diagnostic across biomass from different plant species and also with different co-solvents used in biomass pretreatment.

Our current work is focused on understanding the reactivity of CelR on pure cellulose and pretreated biomass treated with γ-valerolactone (GVL). With both crystalline and amorphous cellulose, CelR shows similar k_{cat} while K_M changes; however, with biomass, CelR exhibits bimodal kinetic behavior (rapid initial velocity followed by slower phase of hydrolysis over an extended time period). This potentially indicates that CelR is operating on two distinct populations of cellulose in the GVL-treated material. Amplitudes of the two kinetic phases and X_{NMR} are being analyzed with a panel of biomasses and pretreatments to identify how different populations of cellulose might respond to hydrolysis.

To better understand the interaction of CelR with cellulose, we solved its crystal structure. Similar to an earlier structure of a related enzyme from *Thermomonospora fusca* [2],

CelR has an open binding cleft and active site that would allow access of amorphous cellulose strands, but this single domain of the enzyme shows only weak reactivity with polysaccharides. When the natural carbohydrate binding module (CBM) is present, the crystal structure shows a tightly packed, extensive surface lined with aromatic residues that has a roughly linear orientation toward the active site. This form of the enzyme is ~10x more reactive than the catalytic domain alone and appears to behave as a processive endocellulase. When the enzyme is engineered to contain a second non-native CBM attached by flexible linker, the reactivity is doubled again. A combination of structural and catalytic contributions to this progressive improvement in reactivity will be reported.

References/Publications

1. Walker T.W., Kuch N., Vander Meulen K.A., Clewett C.F.M., Huber G.W., Fox B.G., Dumesic J.A. 2020. Solid-State NMR Studies of Solvent-Mediated, Acid-Catalyzed Woody Biomass Pretreatment for Enzymatic Conversion of Residual Cellulose. *ACS Sustainable Chemistry & Engineering* **2020** *8* (16), 6551-6563.

2. Sakon J., Irwin D., Wilson D.B., Andrew Karplus P. 1997. Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca. *Nature Structural Biology* **1997** *4*(10), 810-818.

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.

Engineering naringenin into the lignins of transgenic poplar

Elizabeth L. Mahon[1,3*] (elizabeth.mahon@botany.ubc.ca), Lisanne de Vries,[1,3] Soo-Kyeong Jang,[1] Sandeep Middar,[1] Hoon Kim,[3] John Ralph,[2,3] and **Shawn D. Mansfield**[1,3]

[1]University of British Columbia, Vancouver, Canada; [2]Department of Biochemistry and [3]DOE, Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison USA

Project Goals: Modifying the content and composition of lignin in poplar by diverting carbon away from the phenylpropanoid pathway and towards the production of naringenin, a valuable flavonoid, in lignifying tissues.

Lignin, a major chemical constituent of lignocellulosic biomass, is an important strategic target for genetic modification as this polyphenolic polymer poses a significant barrier to efficient industrial processing. Gain of function approaches that improve processing efficiency as well as add value to biomass can play a key role in development of cost-effective biomass processing methods. Lignin is typically composed of three canonical monolignols that undergo oxidative radical coupling to form lignin in the cell walls of developing xylem tissue. Efforts to genetically engineer the monolignol biosynthetic pathway have led to significant changes in both content and composition of lignin, highlighting the remarkable metabolic plasticity of lignin biosynthesis. Moreover, compounds found beyond the monolignol pathway, such as flavonoids and stilbenes, have been found naturally incorporated into lignins of various plant species.[1,2] Chalcone synthase (CHS) catalyzes the first committed reaction in the production of flavonoid compounds by combining *p*-coumaroyl-CoA, a precursor in the monolignol biosynthetic pathway, with three malonyl-CoA units to produce naringenin chalcone which is then cyclized to naringenin. Using a xylem-specific promoter, we have genetically engineered hybrid poplar (*Populus alba* × *grandidentata*) to express a previously characterized chalcone synthase (*CHS*) gene (*MdCHS3*) derived from apple (*Malus* x *domestica*).[3] *MdCHS3*-poplar displayed an accumulation of naringenin in xylem methanolic extracts, not inherently observed in wild-type trees. NMR analysis revealed the presence of naringenin in the extractive-free enzyme lignin (EL), residue of cellulase-treated xylem tissue, of *MdCHS3-poplar*, indicating the novel incorporation of this flavonoid into poplar secondary cell wall lignins. The transgenic trees also displayed lower total lignin content, increased cell wall carbohydrate content, and performed significantly better in saccharification assays than their wild-type counterparts. Moving forward, *MdCHS3*-poplars represent a useful genetic background into which new flavonoid biosynthetic enzymes may be introduced in order to produce other valuable lignin-compatible flavonoids.

References

1 del Río, J. C., Rencoret, J., Gutiérrez, A., Kim, H. & Ralph, J. Hydroxystilbenes are monomers in palm fruit endocarp lignins. *Plant Physiology* **174**, 2072-2082 (2017).

2 Lan, W. *et al.* Tricin, a flavonoid monomer in monocot lignification. *Plant Physiology* **167**, 1284-1295 (2015).

3 Yahyaa, M. *et al.* Characterization of three chalcone synthase-like genes from apple (*Malus* x *domestica* Borkh.). *Phytochemistry* **140**, 125-133 (2017).

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 DE-SC0018409.

Title: Metabolomic and Proteomic Analysis of *Zymomonas mobilis* During Nitrogen Fixation Reveals Metabolic Remodeling of Biofuel Producing Pathways

Authors: Julia I. Martien,[1,2]* (martien@wisc.edu), Edna A. Trujillo,[3,4] David M. Stevenson,[1,2] Joshua J. Coon,[1,3,4,5], and **Daniel Amador-Noguez**,[1,2]

Institutions: [1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison; [2]Department of Bacteriology, University of Wisconsin-Madison; [3]Department of Chemistry, University of Wisconsin-Madison; [4]National Center for Quantitative Biology of Complex Systems, University of Wisconsin-Madison; and [5]Morgridge Institute for Research, Madison, WI

Project Goals: Characterize native metabolic regulation of biofuel producing pathways (such as ethanol production and isoprenoid synthesis) in order to inform genetic engineering of *Z. mobilis* for improved biofuel production.

Abstract text: *Zymommnas mobilis* is a promising biofuel producer capable of rapid glucose consumption and ethanol production. Recently, it was demonstrated that *Z. mobilis* can fix N_2 as a sole nitrogen source (1). Under N_2 fixing conditions, *Z. mobilis* exhibited a higher specific rate of ethanol production than when NH_4^+ was supplied in the media (1, 2). In order to better understand the metabolic remodeling that occurs during N_2 fixation, we performed metabolomics, proteomics, and thermodynamic analysis of *Z. mobilis* under conditions of N_2 fixation compared to replete NH_4^+ availability. We also performed metabolomic and proteomic analysis during the dynamic shift to N_2 fixing conditions (NH_4^+ downshift) and during the shift to NH_4^+ replete conditions (NH_4^+ upshift). We found that intracellular concentrations of intermediates of the Entner-Doudoroff (ED) glycolytic pathway were depleted during N_2 fixation. Protein levels of zinc-dependent alcohol dehydrogenase (encoded by adhA, ZMO1236) increased by 10-fold during the shift to N_2 fixation, helping to explain the previously observed increase in specific ethanol production. Positional stable isotope labeling revealed that labeled forms indicative of reverse flux were more abundant under NH_4^+ replete conditions for all five labeled schemes tested, implying increased thermodynamic favorability of the ED pathway during N_2 fixation. We also observed severe depletion of intermediates of the methylerythritol 4-phosphate (MEP) pathway during N_2 fixation, which was accompanied by decreased protein abundance of deoxyxylulose 5-phosphate synthase (DXS), the first enzyme of the MEP pathway. Unexpectedly, we found that intracellular arginine levels were over 3-fold higher during N_2 fixation and decreased by over 3-fold within 10 minutes of NH_4^+ addition. Based on an overall depletion in intermediates of arginine biosynthesis during N_2 fixation and dynamic changes in protein abundance of a group IV pyridoxal-dependent decarboxylase, encoded in an operon with a deoxyhypusine synthase-like gene, we hypothesize that polyamine synthesis from arginine plays an important role in *Z. mobilis* physiology during changes in NH_4^+ availability. This study has expanded our fundamental understanding of nitrogen metabolism in *Z. mobilis*, identified DXS protein abundance as a native control-point for MEP pathway activity, and demonstrated that metabolic remodeling during N_2 fixation results in increased thermodynamic favorability of

the ED pathway *in vivo*. These results will help to inform future efforts for metabolic engineering in *Z. mobilis* to increase biofuel production.

References/Publications

1. Kremer TA, LaSarre B, Posto AL, McKinlay JB. 2015. N2 gas is an effective fertilizer for bioethanol production by Zymomonas mobilis. Proc Natl Acad Sci U S A 112:2222–6.
2. Palamae S, Choorit W, Chatsungnoen T, Chisti Y. 2020. Simultaneous nitrogen fixation and ethanol production by Zymomonas mobilis. J Biotechnol.

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

Investigating the Interplay between the Phyllosphere Microbiome, Epicuticular Wax, and Root Mucilage on Sorghum Resilience to Water and Nitrogen Limitation.

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:

Sorghum (*Sorghum bicolor)* produces several exudates on its external aerial surfaces (epiphytic phyllosphere). In particular, sorghum accumulates high levels of epicuticular wax on stems over its development, and it also grows aerial roots that produce protective sugar-rich mucilage. The epicuticular wax supports tolerance to sorghum water limitation. However, the functional role of the sorghum aerial root mucilage is still unclear, though it has been shown that diazotrophs colonize the mucilage of maize aerial roots and can provide substantial fixed nitrogen for the host. Here, we hypothesize that the wax and mucilage are suitable environments for specialized microbiomes that support sorghum resilience to stress from water limitation and nitrogen limitation, respectively. We applied 16S rRNA gene amplicon sequencing to study the bacterial communities associated with the aerial root mucilage from N-fertilized and non-fertilized sorghum plants at two-time points, 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 mucilage compartment regardless of the fertilization treatment. However, differential abundance analysis revealed specific bacterial taxa significantly more abundant on the mucilage of N-fertilized plants compared with non-fertilized treatment and vice versa. Similarly, we found that Proteobacteria, and Firmicutes are the dominant bacterial group in the epicuticular wax compartment. In parallel, we are building a large bacterial collection by targeting a wide range of traits, including nitrogen-fixation, phosphate-solubilization, methanol-utilization, and others to evaluate the functional diversity of the bacterial communities in sorghum aerial root mucilage and epicuticular wax. 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 drought.

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.

Deletion of the *ntrYX* two Component System in *Rhodobacter sphaeroides* Causes the Generation of Diverse Extracellular Membrane Structures

Daniel Parrell[1,2,3]* (dparrell@wisc.edu), Kimberly Lemmer,[2,3] Timothy 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: The goals of this project are to understand the process by which bacterial cells produce extracellular membranes in order to engineer strains that overproduce these membranes. Bacterial membranes are a good source of lipids which serve as an important source of primary material for a number of pharmaceutical, industrial, and biofuel applications. In this work, we use the photoheterotrophic bacterium *Rhodobacter sphaeroides* to study membrane synthesis and use cryo-electron microscopy (cryo-EM) to examine how extracellular membranes and other cellular structures are produced. Addressing these goals will provide a significant benefit to the study of renewable biofuels and bioproducts.

Rhodobacter sphaeroides is a facultative photoheterotrophic bacterium that serves as an important host for research into the production of primary materials for industrial purposes and biofuels. Previously, a Tn5 transposon insertion screen, paired with a Nile red assay for lipid production, was used to isolate *R. sphaeroides* strains that overproduce extracellular lipids. By identifying and characterizing the processes that lead to increased lipid secretion in these isolates, this work aimed to broadly achieve production of renewable chemicals and fuels from a biological source. One of the strains, which produces the most extracellular lipids, was disrupted at the *ntrYX* gene locus. NtrY and NtrX comprise elements of a two-component regulatory system known to control exopolysaccharide production, as well as processes induced by respiration and anaerobic growth conditions of other organisms. Deletion of *ntrYX* in *R. sphaeroides* compromises envelope stability and cell division. Cryo-electron tomography (cryo-ET) data supporting those observations will be presented. In addition to the phenotypes caused by *ntrYX* deletion, a diverse array of extracellular membrane structures and chains of vesicles were observed by cryo-EM and cryo-ET, which indicated that there was an increase in the production of extracellular lipids. Our observations demonstrate that these extracellular membraneous structures are closely associated with cells and that their production occurs at the cell surface, consistent with observations that the *ntrYX* disruption causes instability to the bacterial envelope. Tomogram segmentation using EMAN2 neural network training on these membraneous structures will be presented to dissect the physical properties of the extracellular vesicles. Data will also be presented on the observation that cold shock appears to have a synergistic effect on the production of these extracellular vesicles and more complex membrane-derived structures. Future work will determine the molecular mechanisms resulting in cell division defects and instability of the envelope in this and other lipid secreting mutants. By

further understanding the mechanisms by which these *R. shaeroides* strains overproduce lipids, great strides can be made 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 Numbers DE-SC0018409.

Redundancy in Aromaric O-demethylation and Ring Opening Reactions in *Novosphingobium aromaticivorans* and their Impact in Biological Conversion of Lignin

Jose M. Perez[1,2,3]* (perez8@wisc.edu), Wayne S. Kontur,[2,3] Carson Gehl,[2,3,4] Derek M. Gille,[2,3] Yanjun Ma,[2,3] Alyssa V. Niles,[2,3] German Umana,[2,3] Timothy J. Donohue,[2,3,5] and **Daniel R. Noguera[2,3]**

[1]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI; [2]DOE Great Lakes Bioenergy Research Center, Madison, WI; [3]Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI; [4]Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI; [5]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA

Website: https://www.glbrc.org/

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, an heterogenous and recalcitrant aromatic heteropolymer. One strategy to make value from lignin is to use chemical techniques to deconstruct it into mixtures of phenolic compounds and to funnel these mixtures into a single product using engineered bacteria. *Novosphingobium aromaticivorans* DSM12444 can naturally degrade multiple lignin-derived phenolic compounds and it has been previously engineered to produce 2-pyrone-4,6-dicarboxylic acid (PDC) from a variety of compounds that are naturally catabolized via 3-methoxygallic (3-MGA) acid or protocatechuic acid (PCA). Two critical reactions involved in aromatic compounds catabolism by *N. aromaticivorans* that can affect their conversion into PDC are O-demethylation and oxidative aromatic ring opening.

In this work, we investigated enzymes predicted to be responsible for O-demethylation of syringic acid, vanillic acid, and 3-MGA, and enzymes predicted to be responsible for the ring opening of 3-MGA, gallic acid, and PCA. Our results show that DesA is involved in syringic acid and vanillic acid O-demethylation, whereas LigM is involved in vanillic acid and 3-MGA O-demethylation. We also found evidence of a potential new O-demethylase involved in the O-demethylation of 3-MGA into gallic acid. In addition, our results support that LigAB is the main enzyme responsible for the aromatic ring opening of 3-MGA, gallic acid, and PCA. However, we also found a previously uncharacterized aromatic ring opening dioxygenase, LigAB2, that has high activity with gallic acid and plays a minor role in the degradation of 3-MGA and PCA.

The data obtained in this study revealed a previously uncharacterized route for aromatic compound degradation 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* the carbon flux from syringic acid is channeled ~85% via aromatic ring opening of 3-MGA and ~15% via its O-demethylation to gallic acid. Finally, by inactivating O-demethylation of 3-MGA in the originally constructed PDC-producing strain of *N. aromaticivorans*, the resulting strain (PDC2) increases the PDC yield from syringic acid to nearly stoichiometric.

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.

Investigating the Effects of Environmental Stress on Feedstock Quality and Biofuel Production by Field-to-Fuel Optimization Research Pipeline

Trey K. Sato[1]* (tksato@glbrc.wisc.edu), Derek Debrauske[1], Pavani Tumbalan[2], Karleigh Kreig[3], Leela Joshi[3], Clifford Foster[2], Audrey P. Gasch[1], Phil Robertson[2], Robert Landick[1], Dirk Norman[1], Millicent Sanciangco[2], **Rebecca G. Ong[3]** and **Kurt D. Thelen[2]**

[1] DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI; [2]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI; and [3]Department of Chemical Engineering, Michigan Technological University, Houghton, MI

https://www.glbrc.org/

Project Goals

Our previous work suggests that feedstock composition can change in response to drought stress, resulting in the production of inhibitory compounds that impede conversion to biofuels by yeast. We established the Field-to-Fuel Optimization research group to better understand the impacts of drought and other environmental stresses on feedstock yield and composition to their downstream effects on microbial biofuel production.

Abstract

In order to establish an economically productive and sustainable bioenergy industry, biorefineries will need to produce biofuels and bioproducts from renewable plant feedstocks consistently from year to year. However, the quality and quantity of biomass used in such pipeline can be profoundly impacted by numerous environmental factors. Previously, we determined that yeast fermentation of lignocellulosic hydrolysates from switchgrass grown during a drought year (2012) in Wisconsin was severely impaired compared to switchgrass from a non-drought year (2010) (Ong *et al.*, 2016). Based on this, we hypothesized that drought and other environmental stresses could alter feedstock composition, which may subsequently impact downstream biomass deconstruction and conversion into biofuels. To directly test this hypothesis, we formed a multi-disciplinary, integrative research pipeline whose goal is to provide crucial knowledge that can enable lignocellulosic biofuel industries to identify biomass growth conditions that negatively impact microbial conversion into biofuels, and determine agronomic practices and microbial engineering strategies that mitigate this inhibition. Together, we designed and planted experimental field sites at five different locations in Michigan and Wisconsin with a variety of bioenergy feedstocks under different treatment conditions (drought, nitrogen depletion, fungicide treatment, etc.) for three growing seasons (2018-2020). These sites included subplots of switchgrass fitted with rainout shelters that siphon off natural rainwater, allowing us to mimic drought conditions and directly compare with control switchgrass samples.

As a first test of our research pipeline, we processed and analyzed 2018 switchgrass samples that were either grown under a rainout shelter ("rainout" treatment) or outside the shelter ("ambient" treatment). We developed and utilized customized pretreatment and enzymatic hydrolysis methods to process small amounts of harvested rainout and ambient switchgrass from a total of 19 separate plots across locations. Next, we performed small-scale anaerobic fermentation experiments and found that yeast fermentation and biofuel production was impaired in rainout switchgrass hydrolysates compared to paired ambient hydrolysates from multiple sites in WI and MI. However, inhibition of biofuel production was not consistently found across all plots at some sites. Additional investigation determined that the dry biomass yields of rainout switchgrass were not consistently lower than their paired ambient controls, suggesting variable effectiveness of individual rainout shelters. By modifying the rainout shelters, we were able to achieve significantly reduced biomass yields for rainout switchgrass across all locations in 2020, indicating significant drought-stress. Future studies are planned to evaluate the rainout and ambient samples from 2020 and compare to the results obtained from 2018. Finally, as part of this project, we developed web applications to track the large number of biomass samples and share their associated data from planting through harvest, deconstruction, and fermentation between groups. In the future, we will apply this Field-to-Fuel Optimization research pipeline to determine whether other environmental conditions that impact biomass and biofuel yields, biochemically identify inhibitory plant molecules induced by the conditions, and uncover genetic engineering strategies that enable biofuel-producing microbes to overcome the inhibition.

References

Ong, R.G., *et al*. (2016). Inhibition of microbial biofuel production in drought-stressed switchgrass hydrolysate. *Biotechnol Biofuels*. 2016 Nov 8;9:237. doi: 10.1186/s13068-016-0657-0

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.

Lignin Valorization by Integrating Chemical Depolymerization and Microbial Funneling

Canan Sener (csener@wisc.edu),[1,2*] Jose M. Perez,[1,2] German E. Umana,[1,2] Shamik Misra,[1,2] Jason Coplien,[1,2] Dennis Haak,[1,2] Steven D. Karlen,[1,2] Christos Maravelias,[1,2,3] John Ralph,[1,2] Daniel R. Noguera,[1,2] **Timothy J. Donohue**[1,2]

[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 lignin-to-bioproduct process chain for the production of 2-pyrone-4,6-dicarboxylic acid (PDC) through microbial funneling of the phenolic monomers from catalytic depolymerization of lignin.

Abstract: Lignocellulosic biomass is a bountiful source of renewable carbon for the sustainable production of fuels and chemicals.

Lignocellulosic biomass is composed of 70-85 wt% polysaccharides (cellulose and hemicellulose) and 15-30 wt% lignin, a heteropolymer of aromatic units. Lignin is the largest source of renewable aromatics on the planet; however, it is quite recalcitrant to removal from the polysaccharides and to its own chemical and/or biological upgrading. Maximizing the value obtained from lignocellulosic biomass feedstocks requires production of liquid fuels and commodity chemicals from both the polysaccharide and lignin fractions.

In this work, we have used chemical and biological upgrading in tandem to extract greater value from the lignin fraction by converting it into 2-pyrone-4,6-dicarboxylic acid (PDC). We will show that high-quality lignin can be isolated from lignocellulosic biomass under mild reaction conditions using γ-valerolactone (GVL) and water as the solvent system and dilute sulfuric acid as a catalyst for polysaccharide depolymerization. We demonstrate that the resulting GVL-lignin can be successfully depolymerized by hydrogenolysis over a Pd/C catalyst into a mixture of monomeric and oligomeric phenolic compounds. The product mixture contains compounds with similar chemical structures that are difficult to separate. Indeed, this has been a major bottleneck in obtaining value-added products from lignin. A compelling solution to this problem is the biological funneling of the mixture of aromatic compounds to a single compound. For biological funneling, we use an engineered strain of *Novosphingobium aromaticivorans* DSM12444 to transform the mixture of aromatic compounds containing syringyl, guaiacyl, and *p*-hydroxyphenyl substructures to a single product, PDC. Thus, we show that a complex mixture of lignin hydrogenolysis products can be reduced to a single product that can be extracted from the culture broth with a simple separation and purification step (*e.g.*, precipitation with sodium chloride). Additionally, we show that this strategy is agnostic to the biomass type by successfully applying this strategy to GVL-lignin extracted from hardwoods (poplar and maple) and grasses (switchgrass and sorghum).

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.

Drought Caused a 33% Reduction in Switchgrass (*Panicum Virgatum L.*) Biomass with Minimum Effects on Net CO_2 Assimilation

Mauricio Tejera[1]*(mauri@msu.edu), **Berkley Walker**[1]

[1] DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI

https://www.glbrc.org/

Project Goals

Characterize physiological switchgrass response to drought and identify how that response impacts fermentation and ethanol production.

Abstract

Switchgrass (*Panicum virgatum* L.) is a perennial C_4 grass, widely considered as a bioenergy crop given its low-input requirements, cold tolerance, and high yields even in marginal lands. Marginal lands are prone to frequent drought events which could impact switchgrass biomass yield and quality. Changes in plant chemical composition under drought stress have been shown to inhibit downstream fermentation processes and ethanol production, however, the physiological base of these changes remains elusive. During the 2020 growing season we installed ~20m^2 rainout shelters to impose drought treatments and study how switchgrass diel carbon assimilation and allocation changed during growing season. Shelters successfully excluded rainfall and limited soil water content to only 10% of rainfed treatments. In addition, they caused a 33% biomass yield reduction ($P = 0.084$). Surprisingly, the drought treatment had smaller effects on switchgrass leaf water potential (LWP), diurnal course of CO_2 assimilation and stomatal conductance. Over the five sampling dates throughout the season, switchgrass CO_2 assimilation was ~15% higher ($p < 0.05$) in rainfed treatments only in mid and late summer. This discrepancy between whole-plant (biomass yield) and leaf-level (CO_2 assimilation) changes could suggest that i) switchgrass plants adjust other aspect of their physiology under drought (e.g., tillering, leaf area) or ii) the small changes at the leaf-level have larger impacts when accumulated over the growing season. At the end of the growing season, we conducted a leaf-level ^{13}C-CO_2 labeling event to compare carbon allocation to different metabolic pathways in drought treatment and rainfed plants. We found no differences in the labeled molecule profile between treatments ($P < 0.1$) and recently assimilated carbohydrates were mainly directed to central metabolism. Altogether, these results help characterize switchgrass response to drought at multiple levels, and identify pathways that are upregulated under drought conditions, with implication on ethanol fermentation.

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: Chemical genomic profiling of hydrolysates and toxins: implications for yeast strain engineering

Authors: Elena Vanacloig Pedros[1]* (vanacloigped@wisc.edu), Kaitlin Fischer[1,2], Lisa Liu[1], Derek Debrauske[1], Michael Place[1], Chris Todd Hittinger[1,2,3], Trey K. Sato[1] and **Audrey P. Gasch**[1,2,3]

Institutions: [1] DOE Great Lakes Bioenergy Research Center, Univ. of Wisconsin-Madison, Madison, WI; [2] Laboratory of Genetics, [3] Center for Genomic Science Innovation, Univ. of Wisconsin-Madison, Madison, WI

Project Goals: We used chemical genomic profiling of yeast gene deletion strains to better understand the mechanisms of toxicity of hydrolysates and their inhibitors, as well as to identify engineering targets to generate hydrolysate tolerant strains more efficient in production of biofuels.

Abstract text: Budding yeast *Saccharomyces cerevisiae* has been extensively used in fermentative industrial processes, including biofuel production from sustainable plant-based hydrolysates. A myriad of toxic compounds and stressors are generated during biomass deconstruction, inhibiting biofuel and biochemical production by microbes. Here, we studied how these compounds affect yeast cells, both to understand the mechanisms of toxicity of hydrolysate inhibitors and to improve efficiency in conversion by engineering more tolerant yeast cells. To do so, we used chemical genomics by exposing a gene deletion library to each of 34 inhibitory chemicals, including solvents used in pre-treatment, toxins generated during hydrolysis of plant material, and biofuel products that induce stress at high levels. The results identified classes of toxins based on similarities and differences in their chemical genomic profiles and revealed surprising insights into the mechanisms of cellular defense. Our results also revealed widespread antagonistic effects of gene deletion strains across specific classes of inhibitors, pointing to conflicting strategies of cellular defense that may pose difficulties for engineering universally tolerant yeast strains. We further propose strategies designed to overcome these engineering challenges.

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: Impact of genome assemblies, genotyping methods, variant types, and ploidy levels on genomic prediction in switchgrass

Authors: Peipei Wang[1,2]* (peipeiw@msu.edu), Fanrui Meng[1,2], Christina B Azodi[4], Michael D. Casler[5], and **Shin-Han Shiu**[1,2,3]

Institutions: [1]Department of Plant Biology; [2]DOE Great Lakes Bioenergy Research Center; [3]Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824, USA; [4]St.Vincent's Institute of Medical Research, Fitzroy 3065, Victoria, Australia; [5]USDA, U.S. Dairy Forage Research Center, Madison, WI and DOE Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53706, USA

Project Goals: Assess factors influencing genomic prediction accuracy in switchgrass; optimize predictions of 20 traits; probe the genetic basis underlying multiple target traits.

Abstract text: Genomic prediction where genotype information is used to predict phenotypes has accelerated breeding processes[1-3] and can provide mechanistic insights into phenotypes of interest. Switchgrass (*Panicum virgatum* L.) is a perennial biofuel feedstock with multiple traits targeted for accelerated breeding using genomic prediction approaches. To optimize switchgrass genomic prediction, we assessed the impact of genome assembly versions, marker sequencing strategies, marker types, marker allelic complexities, and polyploidy levels to predict 20 traits in a switchgrass association panel with 510 individuals[4-6]. We found that genomic prediction models performed similarly regardless of the genome assembly: v.1 or the recently available v.5 assembly. This occurred because the majority of variants—e.g., 70.7% of the single nucleotide polymorphism (SNP) markers from Genotyping-By-Sequencing (GBS)—are shared between these assemblies. Models using markers generated with exome capture outperformed those based on GBS markers for five traits. But in four traits, GBS marker-based models had higher prediction accuracy because the variants underlying the polymorphisms of these four traits tend to be located in intergenic regions. We also found that SNP-based models performed better than Insertion/Deletion (indel) based models for 12 traits, and biallelic marker-based models outperformed models using multiallelic markers for 17 traits. This was due to more SNPs and biallelic markers than indels and multiallelic markers, respectively, as models built with the same numbers of markers had similar accuracy. The most significant model improvement was observed when tetraploids were separated from octoploids, which can be partially explained by the higher trait variances in tetraploid populations. By considering the population structures and factors mentioned above, we present improved genomic prediction models for each of the 20 traits. Finally, we identify candidate genes that are the genetic basis underlying multiple target traits by studying the markers that have the greatest impact on model performance. Our study provides insights into the best practices for performing genomic prediction, which can be used for improving switchgrass agronomic traits through selective breeding.

References/Publications

1 VanRaden, P. M. *et al.* (2009) Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92(1), 16-24.

2 Lorenzana, R. E. & Bernardo, R. (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* 120(1), 151-161.
3 Crossa, J. *et al.* (2017) Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22(11), 961-975.
4 Lu, F. *et al.* (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9(1), e1003215.
5 Lipka, A. E. *et al.* (2014) Accelerating the Switchgrass (*Panicum virgatum* L.) breeding cycle using genomic selection approaches. *PLoS One* 9(11), e112227.
6 Evans, J. *et al.* (2015) Diversity and population structure of northern switchgrass as revealed through exome capture sequencing. *Plant J.* 84(4), 800-815.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018409.

Perenniality Drives Soil Microarthropod Community Differences Across Three Potential Bioenergy Cropping Systems

Authors: Allison Zahorec[1,2]* (zahoreca@msu.edu), Lisa Tiemann[2,3], and **Douglas Landis**[1,2]

Institutions: [1]Dept. of Entomology and [2]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI; [3]Dept. of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI

Project Goals: The goal of this project is to understand how bioenergy crop identity and management influence microarthropod community structure. This work is part of a larger project investigating microarthropod-microbe interactions and their effects on soil carbon accrual.

Abstract text: As Bioenergy with Carbon Capture and Storage (BECCS) continues to show promise as an effective carbon mitigation strategy, bioenergy cropping systems are predicted to become increasingly prevalent in coming decades (Rogelj et al., 2018). Despite its potential as a sustainable alternative to fossil fuels, the widespread implementation of BECCS remains hindered by uncertainties about the ability of bioenergy cropping systems to accrue soil organic carbon (SOC). Specifically, there are serious concerns that the conversion of marginal, non-agricultural lands to managed agroecosystems for bioenergy crop production may result in substantial SOC loss. Before BECCS can be utilized to help offset global emissions and prevent further climate warming, there must be greater understanding of the processes driving SOC accrual and long-term storage in selected bioenergy cropping systems.

While SOC dynamics were traditionally thought to depend primarily upon plant chemistry, it is increasingly understood that soil biota, especially microbes, play a critical role in SOC formation and stabilization (Kallenbach et al., 2016). However, the potentially significant effects of soil fauna on SOC accrual remains unclear. Microarthropods (mites and collembola) are the most abundant and diverse arthropods in soils across ecosystems. Primarily detritivorous and/or microbivorous, microarthropods affect SOC by physically and chemically altering plant litter as well as influencing microbial activity, abundance, and community composition, with potentially important consequences for SOC accrual. When microarthropods, especially mites, were experimentally suppressed, slowed early-stage litter decomposition resulted in 11% less SOC accrual in tallgrass prairie soils based on DayCent modeling (Soong et al., 2016). However, understanding the role of these interactions in the context of bioenergy cropping systems is complicated by the strong impacts of land use and management on microarthropod communities. Therefore, it will be necessary to first understand how crop type and management affect microarthropod communities before attempting to investigate their potential role in SOC accrual in bioenergy cropping systems.

We conducted microarthropod surveys from three bioenergy cropping systems differing in crop life cycle, diversity, and management: an annual monoculture (energy sorghum), reduced-input

perennial monoculture (switchgrass), and no-input perennial polyculture (restored prairie). Microarthropods were collected from soil cores (2018 and 2019) and litter quadrat samples (2019) at three sampling stations within each of five replicate plots (n=15 per treatment) and extracted using Tullgren funnels. Major microarthropod group abundances were counted, with oribatid mites and collembola further categorized to morphospecies to evaluate differences in community composition across cropping systems. We find strong evidence that perennial bioenergy cropping systems support the highest microarthropod abundances. Mites were consistently more abundant in both perennial systems while collembola were as or most abundant in the annual monoculture. Ordination of microarthropod community structure showed that communities in the perennial systems were more similar to each other than the annual monoculture, which was distinct. By supporting greater microarthropod abundances, particularly of mites, this study adds to the growing body of evidence indicating the increased sustainability of perennial bioenergy cropping systems relative to annual systems and is the foundation of ongoing studies evaluating the impact of microarthropods on microbial function and physiology likely to strongly influence SOC dynamics.

References/Publications

1. Rogelj, J., D. Shindell, K. Jiang, S. Fifita, P. Forster, V. Ginzburg, C. Handa, H. Kheshgi, et al. (2018). Chapter 2: Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development. In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Intergovernmental Panel on Climate Change.

2. Kallenbach, C. M., Frey, S. D., & Grandy, A. S. (2016). Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Nature communications*, *7*, 13630.

3. Soong, J. L., Vandegehuchte, M. L., Horton, A. J., Nielsen, U. N., Denef, K., Shaw, E. A., ... & Cotrufo, M. F. (2016). Soil microarthropods support ecosystem productivity and soil C accrual: evidence from a litter decomposition study in the tallgrass prairie. *Soil Biology and Biochemistry*, *92*, 230-238.

Funding statement: This material is based upon work supported in part 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 this research was provided by the National Science Foundation Long-term Ecological Research Program (DEB 1832042) at the Kellogg Biological Station, and by Michigan State University AgBioResearch.

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

Authors: Andrew W. Sher[1]* (awsher@uw.edu), Amir H. Ahkami[2], Soo-Hyung Kim[1], Adam Deutschbauer[3], and **Sharon L. Doty[1]**

Institutions: [1]University of Washington, Seattle; [2]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 the 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. In their native habitat of rocky riverbanks, poplar trees host a diverse assembly of micro-organisms that help them to survive in this harsh environment. Some of the micro-organisms which make up the poplar microbiome can help poplar grow by providing the required nutrients of nitrogen and available phosphorus, which are lacking in the rock and sand dominated riversides. In addition to increasing nutrient acquisition, the micro-organisms can also promote plant tolerance to other environmental stresses including drought. Previously we demonstrated that adding micro-organisms from the wild poplar microbiome to a wide variety of other plants increased the health and growth of these plants under a variety of conditions. Since the start of this newly funded project in autumn 2020, we have isolated additional nitrogen-fixing endophyte strains and are optimizing the microbial consortia that will serve as the bio-inoculants. A suitable field site in a nutrient- and water-limited is being prepared for the spring planting of inoculated poplar plants. Using systems biology approaches at both lab and field scales, we will identify the metabolic and physiological impacts of the bio-inoculants on the host plant under nutrient stress and water limitation. We will then integrate the plant physiology data with the molecular plant-microbe interactions data to develop a systems-level understanding of the genetic and molecular basis for diazotrophic (nitrogen-fixing) endophytic mutualisms. This deeper level of understanding of the plant responses will guide construction of microbial communities that best prime plant pathways for enduring abiotic stresses to optimize the impacts of bioinoculants for environmental and economic sustainability of bioenergy crops.

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 Control of Mineral-Bound Carbon

Joany Babilonia[1*] (jbabilonia@lanl.gov), La Verne Gallegos-Graves[1], Kyana Montoya[1], John Dunbar[1], **Michaeline Albright**[1]

[1]Los Alamos National Laboratory, Bioscience Division, Group B-11

https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-microbial-carbon.php

Project Goals:

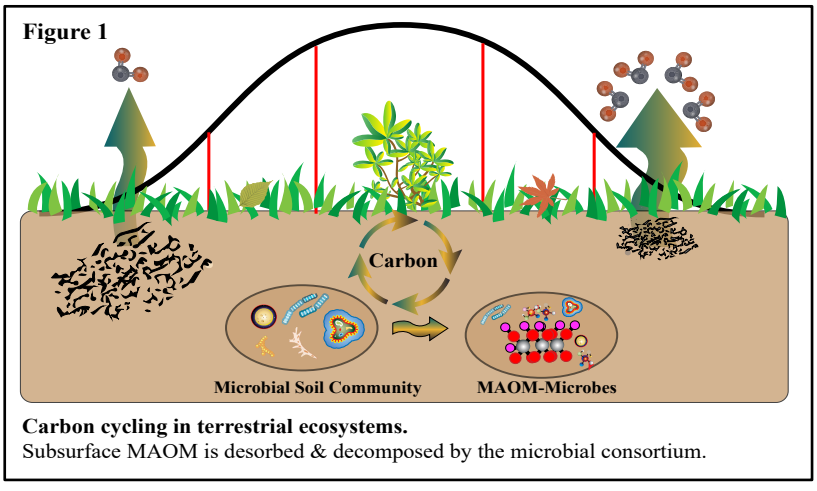
1) Assess the magnitude of microbial-driven variation in mineral associated organic matter (MAOM) subsurface decomposition;
2) Identify specific community features linked to variation in MAOM decomposition

Abstract

The formation of organo-mineral complexes is a key process in long-term stabilization of carbon in soils[1]. Because microbes contribute to desorption and transformation of MAOM, microbial community composition is expected to shape MAOM abundance over millennial scales[2]. To inform soil carbon management strategies, the current project examines the range of variation in carbon flow that may occur during MAOM decomposition dependent on community composition. Further, the project aims to identify community features driving this variation, as has been done with surface litter decomposition[3, 4].

To investigate microbial control of MAOM persistence, we applied a common garden approach by inoculating 103 distinct subsurface soil communities into microcosms. Each microcosm contained a mixture of sterile sand and kaolinite that was pre-loaded with dissolved organic carbon (DOC) obtained from ground Ponderosa pine needles. A total of 206 microcosms were generated and the subsurface soil communities were allowed to desorb and decompose the MAOM, with the loss of MAOM measured as CO_2 efflux. Measurements were taken over a 105-day dark incubation period. 'High' and 'low' respiration (CO_2) phenotypes were down-selected for community (bacteria and fungi) profiling.

Preliminary results of the ongoing study indicate that microbial-driven variation in desorption and decomposition of (MAOM) is large. Respiration profiles show greater than 3.5-fold variation across communities, with the cumulative carbon respiration ranging from 0.5 - 1.92% CO_2 (1.2 – 4.2 mg CO_2) over a 105-day incubation. This amounts to between 16.7 and 43.8% of the available mineral bound DOC in the microcosms. Community profiling of the microbial cohorts representing



Carbon cycling in terrestrial ecosystems.
Subsurface MAOM is desorbed & decomposed by the microbial consortium.

the ends of this range (the 'high' and 'low' CO_2 phenotypes) show that some distinctive communities contribute to the observed range of CO_2 efflux. To gain further insight into specific features linked to 'high' and 'low' respiration, we are applying a machine learning pipeline described previously[5]. Significant features will be used to infer traits that may be driving differences in carbon flow, providing insight into MAOM dynamics.

References

1. Mikutta, R. and K. Kaiser, *Organic matter bound to mineral surfaces: Resistance to chemical and biological oxidation.* Soil Biology and Biochemistry, 2011. **43**(8): p. 1738-1741.
2. Woolf, D. and J. Lehmann, *Microbial models with minimal mineral protection can explain long-term soil organic carbon persistence.* Sci Rep, 2019. **9**(1): p. 6522.
3. Albright, M.B.N., et al., *Soil Bacterial and Fungal Richness Forecast Patterns of Early Pine Litter Decomposition.* Front Microbiol, 2020. **11**: p. 542220.
4. Albright, M.B.N., et al., *Differences in substrate use linked to divergent carbon flow during litter decomposition.* FEMS Microbiol Ecol, 2020. **96**(8).
5. Thompson, J., et al., *Machine learning to predict microbial community functions: An analysis of dissolved organic carbon from litter decomposition.* PLoS One, 2019. **14**(7): p. e0215502.

Funding statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (F255LANL2018).

Cycles of Wetting and Drying Reduce Carbon Efflux and Litter Decomposition in Soil Microcosms

Tayte Campbell[1*] (tayte.campbell@pnnl.gov), Sophia McKever[1], John Dunbar[2], and **Vanessa Bailey**[1]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA
[2]Bioscience (B-11), Los Alamos National Laboratory, Los Alamos, NM

https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-microbial-carbon.php

Project goals: The aim of this project is to combine metagenomic and metabolomic analyses to understand the microbial influence on carbon cycling in simplified microcosms that mimic soil conditions.

The objective of this research is to understand the different mechanisms through which microbial communities respond to changing moisture levels, and the consequences of these differences to overall soil carbon cycling. Climate change models predict changes in precipitation, with some areas receiving more water inputs and others receiving less. To investigate how moisture variation affects CO_2 efflux and soil carbon stabilization, we used simplified microbial communities in sealed microcosms. Simplified microcosm communities developed from soil inocula allow for the control of environmental factors and provide a tractable level of complexity for mimicking microbial communities in soils.

We collected soils from 95 sites across southwestern grassland and forest soils in the United States and used them to inoculate sand microcosms containing Ponderosa pine needle litter as a carbon source. The soil microcosms were incubated over a three-month period with two treatments: one with constantly wet moisture conditions and one treated with three wetting cycles interspersed with long periods of desiccation. We measured CO_2 efflux throughout the incubation and destructively sampled at the end of the incubation to measure dissolved organic carbon (DOC), pine needle litter decomposition, and to characterize the microbial community using 16S sequencing. From these measurements we aim to use 16S sequencing analyses of the microbial community and chemical analyses on the metabolite profiles to link C fluxes with the microbial communities that drive these processes. We found that soil microbial communities subjected to wetting and drying cycles had 23% less pine needle consumption, 45% less cumulative CO_2 efflux, and 19% less DOC than communities maintained at a constantly wet moisture level. Overall, these findings indicate that wetting events followed by long periods of drought result in lower C degradation, sequestration, and efflux in the soil environment Further, microcosm samples inoculated from the same soil showed similar intra-treatment trends of C efflux in both treatments, indicating that the microbial community is in fact driving the intra-treatment differences in C cycling.

Microbial driven variation in carbon flow & stabilization during root litter decomposition

Rae DeVan[1]* (raedevan@lanl.gov), Rachel Hestrin[2] (hestrin1@llnl.gov), Michaeline Nelson-Albright[1] (malbright@lanl.gov), John Dunbar[1] (dunbar@lanl.gov)

1. Los Alamos National Laboratory, Los Alamos, NM, 2. Lawrence Livermore National Laboratory, CA

https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-microbial-carbon.php

Project Goals:

1. Identify predictive links between carbon flow and microbial traits.
2. Understand the interaction of microbial triats and environmental fluctuations.
3. Determine ecosystem level consequences of microbial driven variances in carbon flow.

INTRODUCTION:

This Science Focus Area (SFA) aims to inform climate modeling and enable carbon management in terrestrial ecosystems. Litter decomposition substantially contributes to the global carbon cycle, and recent evidence suggests microbial composition plays a role determining amounts and type of carbon flow during decomposition. Thus far, the majority of research examining the effects of microbial communities on litter decomposition have focused on plant litter and surface microbial communities. However, subsurface plant litter accounts for a large portion of annual litter production and turns over faster than surface material. Furthermore, belowground carbon inputs are more efficiently stabilized and the ultimate fate of soil carbon is dependent on subsurface dynamics. In the current study, we have three main questions:

1. What is the magnitude of microbial driven variation of carbon flow in the shallow subsurface during root litter decomposition, as measured by CO_2 and DOC?
2. Is there a relationship of DOC production and C and N stabilization on mineral surfaces?
3. Does N addition alter the magnitude of carbon flow or C stabilization on mineral surfaces?

To answer the above questions, dual-labeled blue grama root litter was incubated in a common garden microcosm experiment. Litter was inoculated with rhizosphere microbial communities derived from subsurface soils collected from ten elevational transects across the southwest. We found subsurface microbial communities produced DOC distributions similar to those produced by surface microbial communities during leaf litter decomposition, but subsurface communities produced CO_2 distributions that were an order of magnitude larger than surface microbial communities. These differences in carbon flow during root litter decomposition appear to be driven by bacterial rather than fungal communities. The addition of nitrogen had a significant effect on the diversity of both bacterial and fungal communities as well

as carbon flows, but the effects varied by transect which originating microbes were isolated from. This suggests microbial communities present on litter alter the effect of nitrogen deposition on community composition and carbon flow. These results indicate that microbial communities on root litter play an important role in determining the magnitude and direction of carbon flow during decomposition. Future work will examine how these results related to carbon stabilization on mineral surfaces.

Funding statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant

Influence of Microbial Surface Litter Decomposer Communities on CO_2 Emissions from Natural Soils

Rae DeVan[1], Sanna Sevanto[2*] (sanna@lanl.gov), Rose Harris[2], John P. Heneghan[2], Dea Musa[2], George Perkins[2], M. Francesca Cotrufo[3], Michaeline B. N. Albright[1] and **John Dunbar**[1]

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM
[2]Earth and Environmental Science Division, Los Alamos National Laboratory, Los Alamos, NM
[3]Department if Soil and Crop Science, Colorado State University, Fort Collins, CO

https://www.lanl.gov/science-innovation/science-programs/office-of-science-programs/biological-environmental-research/sfa-microbial-carbon.php

Project Goals: Test the influence of surface litter decomposer communities that differ in dissolved organic carbon (DOC) production on carbon cycling, balance and emissions in natural soil, and establish predictive links between carbon dioxide emissions, DOC production, and carbon transport to deeper soil layers during surface litter decomposition.

Carbon from terrestrial plant litter decomposition may be incorporated into microbial biomass, respired to the atmosphere as CO_2, or leached into the soil as dissolved organic carbon (DOC). Due to the large amounts of litterfall each year, the fate of this carbon has important implications for global carbon cycling. While most research has focused on climatic and edaphic controls over this carbon, there is strong evidence that microbial community composition can alter C flow during decomposition. The main goals of the LANL Microbial Carbon Cycling SFA are to inform climate modeling and enable carbon management using soil microbial communities. Previous common garden experiments using decomposing litter in microcosms during this project identified soil communities with divergent carbon flows measured as CO_2 and DOC. To understand the effects of microbially driven carbon flow from surface litter decomposition on net ecosystem C flow, we inoculated dual-labeled blue grama plant litter with microbial communities previously identified as either 'high' or 'low' DOC and incubated the litter on 30 cm, intact soil cores for 8 weeks, with half of the cores containing established blue grama plants. We expected cores inoculated with 'high' DOC communities to produce more DOC and less CO_2 compared to cores inoculated with 'low' DOC communities.

Our results show that CO_2 flux and DOC concentration peaked in the first 3 weeks of decomposition in all cores, but 'high' and 'low' DOC microbial communities behaved differently depending on the presence of plants. All cores with live plants produced similar total CO_2, but the 'low' DOC cores were more enriched in $^{13}CO_2$ indicating higher CO2 release from litter decomposition consistent with expectations. However, in cores without plants, 'low' DOC cores produced more total CO_2, but were less enriched than 'high' DOC cores. DOC results were similarly flipped depending on the presence of plants. In cores with live plants, opposite to what was expected, inoculation with 'high' DOC producing microbial communities resulted in a greater reduction in DOC than inoculation with 'low' DOC communities compared to DOC baseline. In cores with no plants, the treatments behaved as expected with the 'high' DOC cores

producing more DOC than the ‘low’ DOC cores compared baseline. In summary, the presence of plants altered the outcome from microbial treatment, and neither set of cores behaved as expected for both carbon flows. ‘Low’ DOC cores with plants resulted in greater C flow both from CO_2 and DOC compared to ‘high’ DOC cores, while the opposite trend occurred in cores without plants.

*This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (*F255SFA2018).

Title: Comparing In-Situ, Individual Bacterial Growth Rates in Cropped and Successional Soils Using a 16S rRNA Internal Standard

Authors: Cassandra J. Wattenburger[1]* (cjw259@cornell.edu), **Daniel H. Buckley**[1]

Institutions: [1]Cornell University, Soil and Crop Science Section, Ithaca, NY

Project goals:
We hypothesize that carbon fate in soil is governed by microbial growth dynamics, which determine consumption and production of soil organic matter, and that these dynamics are constrained and determined by soil properties. Project goals include:

- **Develop a method for determining in-situ bacterial growth rates using 16S rRNA amplicon sequencing.**
- **Determine how carbon inputs to soil alter growth rate frequency distributions of bacterial communities.**
- **Investigate the impacts of soil management history and resource availability on bacterial community growth dynamics.**

Abstract text: Soil bacteria drive biogeochemical cycles via metabolic activities that break down, assimilate, and transform compounds containing carbon (C), nitrogen, and other elements. Bacterial growth dynamics are inherently connected to bacterial metabolism and C transformation. Only recently have individual in-situ growth rates been measured, and little is known about bacterial community growth dynamics in nature. Here we used an internal standard (16S V4 rRNA modified oligo of *Aquifex aolicus*) with 16S rRNA amplicon sequencing to estimate in-situ growth rates from cropped and successional soils, with water or C amendment (3.6 mg/g dry soil, various soluble and insoluble compounds). We hypothesized that soil habitat and C availability would impact growth dynamics within the bacterial community, specifically that cropped and C-amended soils would harbor more fast growing taxa compared to successional or water-amended soils, reflecting differences in soil disturbance and resource availability. We also hypothesized that in-situ growth rates would correlate positively with 16S rRNA copy number. The internal standard comprised an average of 1.6% of the sequenced reads per sample and displayed a strong negative correlation with DNA yield (Spearman, rho = -0.66, $p < 0.001$). Overall, we were able to estimate in-situ growth rates for 453 taxa across all soils after filtering for sparsity and controlling for false positives. We observed a weak, positive correlation between in-situ growth rates and 16S rRNA copy number (Spearman, rho = 0.166, $p < 0.001$). There was a significant difference in the growth rate frequency distributions between the cropped and successional soils in the water-amended treatment but not in the C-amended treatment (Fisher's exact test, $p = 0.004$ and $p > 0.05$). Although overall growth rate frequency distributions differed between water-amended soils, the number of "fast" and "slow" taxa were not significantly different in either the cropped or successional, water-amended soils (Welch's paired t-tests, $p > 0.05$). Overall, these results demonstrate (i) the utility of using internal standards for estimating individual bacterial growth rates in soils, (ii) that 16S rRNA copy number explains significant variation in observed bacterial growth rates in soils, and (iii) soil habitat and resource availability have large and complex impacts on growth within bacterial communities. Going forward, we will use this method to investigate how differences in bacterial community growth dynamics impact C cycling in soils.

Funding statement: This research was supported by the DOE Office of Science, Biological and Environmental Research Division, under award number F255LANL2018.

Live-cell imaging of *E. coli* biofuel synthesis by spectroscopic stimulated Raman scattering microscopy

Haonan Lin[1,3]*(hnlin@bu.edu), Nathan Tague[1], **Wilson Wong[1,4], Ji-Xin Cheng[1,2,3] and Mary J. Dunlop[1,4]**

[1]Department of Biomedical Engineering, [2]Department of Electrical & Computer Engineering, [3]Photonics Center, [4]Biological Design Center, Boston University, Boston, MA 02215, USA

Project Goal: Fatty acids are essential biofuel precursors that can be synthesized in microbes such as *E. coli.* Currently, quantitation of fatty acid production levels mainly relies on ensemble measurements such as gas chromatography-mass spectrometry (GC-MS), which ignores the potential for genetic or phenotypic variation among cells. Stimulated Raman scattering (SRS) is a high-speed imaging modality that produces label-free chemical maps of molecules, allowing tracking of the production process at the single-cell level. To perform live-cell imaging, we built a microscope incubator and recorded both wide-field transmission images and SRS. The biologically safe SRS laser power enables multiple SRS acquisitions without perturbing cell growth. We performed spectral unmixing on spectroscopic SRS images to determine the concentration of fatty acids within each cell. With the platform, we can obtain information on the potential genetic or phenotypic variation among cells to improve the biofuel production yield.

Abstract

Stimulated Raman scattering (SRS) is a label-free imaging modality that offers chemical contrast based on intrinsic molecular vibrations. To improve chemical specificity in a complex biological system, spectroscopic SRS has been developed to provide a Raman spectrum at each pixel, which can be further decomposed into maps of chemical species such as protein and fatty acid. SRS has seen a wide range of biomedical applications, including cell metabolism, drug delivery and neuron voltage tracking. Fatty acid synthesis by *E. coli* is an ideal application for SRS given the strong signal of fatty acids in the CH region. Due to its high imaging speed, SRS enables imaging the dynamics of cell division and biofuel synthesis, which provides insights into genetic or phenotypic variations among cells. Here, we report our work towards the live tracking of single-cell biofuel synthesis.

To achieve live-cell imaging, we installed a microscope incubator to control the temperature. An eyepiece camera is installed to capture the wide-field of view with white light lamp illumination. We selected 15 mW for 800 nm pump and 25 mW for 1040 nm Stokes on the sample to avoid laser damage on cell growth. As shown in Fig. 1, we prepared a sample slide with an agarose gel pad to trap the cells and provide nutrition for cell growth. We performed a time-series imaging of the same field of view for a period of 19 hours. Two example time points (7 and 19 hours) are shown in **Fig. 1**, each of which consists of

a wide-field image, the raw spectroscopic SRS image and chemical maps of protein and fatty acid after spectral unmixing using the references shown in **Fig. 1c**. We observed that in the 7-hour image, cells were predominately focused on growing instead of fatty acid production, as no signal is detected in the fatty acid channel. In the 19-hour image, fatty acid signals appear in the form of aggregated droplets. Heterogeneity of fatty acid concentrations among different cell colonies can be observed in the fatty acid map. In the future, we will perform the study on *E. coli* strains with various fatty acid production levels, with a focus on the dynamics of the fatty acid synthesis process and cellular heterogeneity.

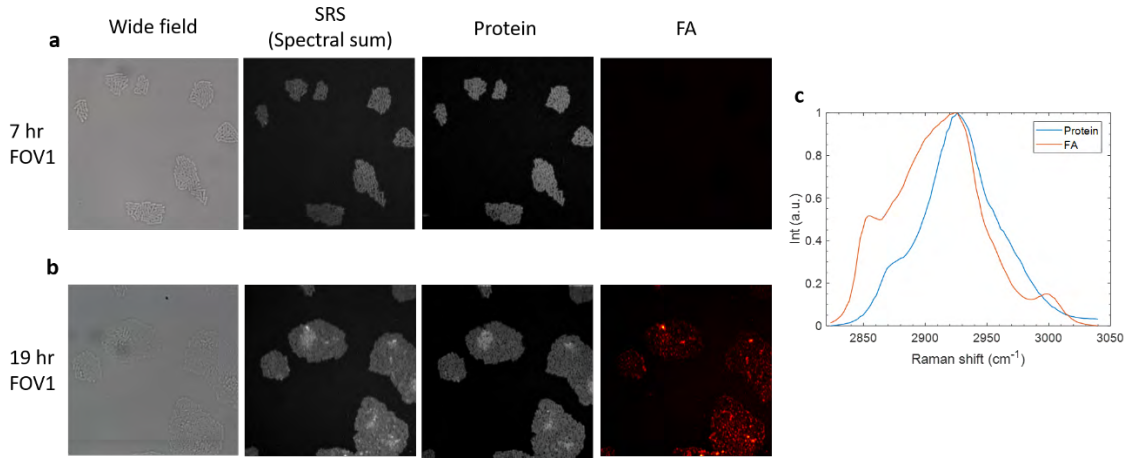


Figure 1. Live-cell imaging of *E. coli* with spectroscopic SRS imaging. (a-b) 7-hour and 19-hour imaging of the same field of view. From left to right: wide-field transmission image, SRS (spectral sum), protein concentration map, fatty acid, FA concentration map. (c) Spectral profiles of protein and fatty acid, FA.

This work is supported by a DOE grant BER DE-SC0019387 to MJD, WW and JXC.

Single Cell Chemical Imaging with Stimulated Raman Scattering for Biofuel Production Screening

Nathan Tague[1]* (ntague@bu.edu), Haonan Lin[1], Jonghyeon Shin[1], Jing Zhang[1], Jean-Baptiste Lugagne[1], Wilson Wong[1], Ji-Xin Cheng[1], **Mary J. Dunlop[1]**

[1]Boston University, Boston, MA

Project Goals:

The convergence of metabolic engineering and synthetic biology disciplines has led to increasingly efficient microbial production of valuable chemicals such as biofuels. Recent advances in genetic and protein engineering allow for more extensive strain optimization and diverse chemical production. However, state of the art screening techniques are destructive and lose single cell production data. Stimulated Raman scattering (SRS) imaging has the potential to be used as a single cell screening platform by directly detecting chemicals' Raman spectra. Here, we demonstrate the ability to differentiate production levels of several biofuels within E. coli based on unique Raman spectra footprints. Utilizing the spectral information, we showcase the ability to differentiate free fatty acids of different chain lengths, allowing for accurate enzyme specificity prediction. The results show the potential of this imaging platform to be used to guide strain optimization for metabolic engineering projects.

Identifying growth harness actuator genes for *Pseudomonas putida*

Shara Balakrishnan[1]* (sbalakrishnan@ucsb.edu), Rob Egbert[2] and Enoch Yeung[3]

[1]Dept. of Electrical and Computer Eng, University of California Santa Barbara, Santa Barbara, California;

[2]Earth & Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington;

[3]Dept. of Mechanical Eng, University of California Santa Barbara, Santa Barbara, California

Website: https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

Project Goals: *The Pacific Northwest National Laboratory Persistence Control Scientific Focus Area is focused on identifying controllable variables that govern the persistence of engineered microbial functions in the rhizosphere. Phenotype controllability is a new concept, integrating advances in machine learning and control theory with the latest measurement and biotechnological advances in synthetic and systems biology. With rigorous notions of phenotype control established, we elucidate novel design principles to control the environmental niche of candidate rhizosphere microbes. These principles will lead to secure plant–microbe biosystems that promote secure, stress-tolerant, and highly productive biomass crops.*

Advancing a new genetically modified microbial species from the laboratory to the field is a major obstacle in synthetic biology. When a synthetic microbe is introduced into a new environment, it interacts with native microbial fauna, subjecting itself to competitors and predators that constrain proliferation [1]. To control its persistence in the alien environment, we aim to design growth harness actuators that can potentially control the expression of genes to manipulate the microbial growth. A critical challenge in the design process is to identify the "fitness genes" which are the key genes that play a role in the growth phenotype.

Existing approaches primarily center on analysis of differential RNA-sequencing (dRNA-seq) data; comparative models contrast different phenotypes with different transcriptome profiles to create a genotype-to-phenotype map. To identify the fitness genes, dRNA-seq can be used by comparing the RNA expression levels of samples in the growth phase and the stationary phase. Random barcoded transposon sequencing (RB-TnSeq) is an approach that is tailored specifically to identify fitness genes. The fitness genes are identified by creating genetic variants of the base strain by deactivating a single gene in each variant and measuring the fitness values of the variant. The more negative the fitness value, the more important the gene. The fundamental premise of RBTnSeq methods is that the microbe of interest is genetically tractable and can be engineered with high enough transformation efficiency to make large, pooled libraries of knockdown variants. Thus, for emerging microbial species in the rhizosphere with undefined genetics, there is a need for data-driven approaches.

We propose a novel data-driven technique that integrates dynamical systems theory and machine learning to discover genotype-to-phenotype mathematical models that map transcriptomic signatures to fitness curves. These models quantify how the growth curve changes in response to change in expression of candidate genes, in a purely data-driven fashion, which enables analysis of de

novo genome sequence and transcriptomics data. These models are formulated precisely to solve a never before explored sensor fusion problem in Koopman operator theory, which through spectral techniques yields candidate genes that strongly influence growth phenotype.

We model the proliferation process of the microbe as a state space model from dynamical systems theory by assuming the state of the system to be the RNA expression as it characterizes the genetic activity of the microbe and the output to be the growth curve as that is the phenotype of interest. The state space model comprises the state equation which captures the evolution of the genetic activity as a function of time and the output equation which maps the genetic activity to the growth curve output. We developed an algorithm called the output constrained deep dynamic mode decomposition (OC-DeepDMD) algorithm 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. The high dimensional state is obtained by a state inclusive nonlinear transformation identified using multilayer feedforward neural networks as done in [2].

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 select 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 perform a time-series RNA sequencing experiment for the MAX and NC conditions while obtaining optical density (OD) measurements. We identified the Koopman operator representation of the state space model using OC-DeepDMD algorithm. Using modal decomposition, we club the genes into sparse modes and the dominant genes in the mode that contributes maximum to the output are the candidate genes of interest.

In summary, we proposed the OC-DeepDMD algorithm to identify state space models that serve as genotype-to-phenotype maps. Specifically, we used this algorithm to identify a set of genes for P.putida in R2A media that correspond to the growth phenotype. These genes form the recommendations for genetic targets for CRISPR-dCas9 knockdown to thereby control persistence of the microbe.

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*, *1*(1), pp.22-30

[2] Yeung, E., Kundu, S. and Hodas, N., 2019, July. Learning deep neural network representations for Koopman operators of nonlinear dynamical systems. In *2019 American Control Conference (ACC)* (pp. 4832-4839). IEEE.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012345 and by Pacific Northwest National Laboratory, grant no. 490521

Title: Serine integrase-assisted genome engineering (SAGE) enables efficient, iterative site-specific genome engineering in bacteria

Authors: Joshua Elmore[1]* (joshua.elmore@pnnl.gov), Gara Dexter[2], Ryan Francis[1], Lauren Riley[2], Jay Huenemann[2], Henri Baldino[1], Adam Guss[2], **Robert Egbert**[1]

Institutions: [1]Pacific Northwest National Laboratory, Richland, WA; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee

Website: https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

Project Goals: The Persistence Control Science Focus Area at PNNL 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 for the model bioenergy crop sorghum through data-driven genome reduction and engineered metabolic addiction to plant root exudates. These principles will lead to secure plant–microbe biosystems that promote secure, stress-tolerant, and highly productive biomass crops.

Abstract text: Sustainable enhancements to crop productivity and increased resilience to adverse conditions are critical for modern agriculture, and application of plant growth-promoting rhizobacteria (PGPR) is a promising method to achieve these goals. However, many desirable PGPR traits are highly regulated in their native microbe, limited to certain plant rhizospheres, or insufficiently active for agricultural purposes. Synthetic biology can address these limitations, but its application is hampered by the lack of appropriate tools for sophisticated, high-throughput genome engineering for operating outside of the laboratory. Here we present an orthogonal genome engineering system, Serine integrase-assisted Genome Engineering (SAGE), which enables iterative, site-specific integration of up to 10 distinct heterologous DNA constructs (Figure 1). Transformation efficiencies with SAGE are frequently on par with or higher than those observed with common

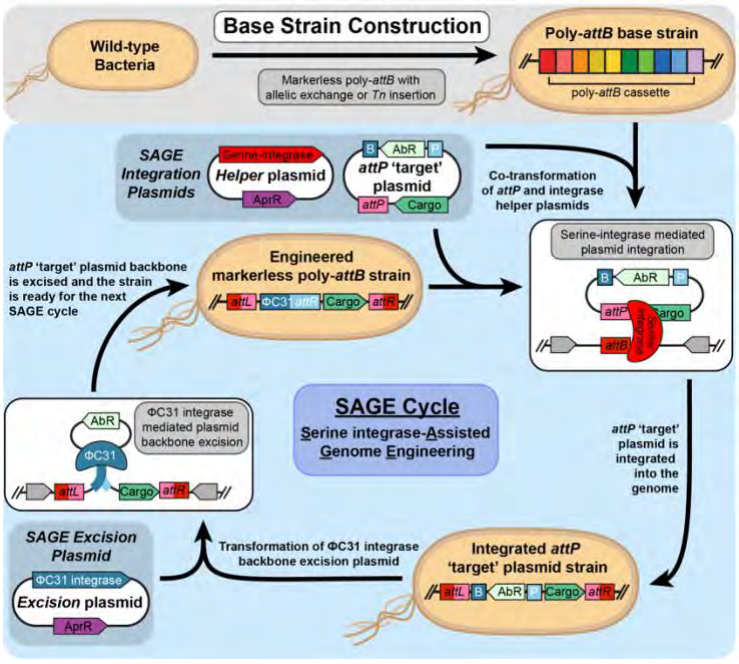


Figure 1. Overview of Serine integrase-assisted genome engineering

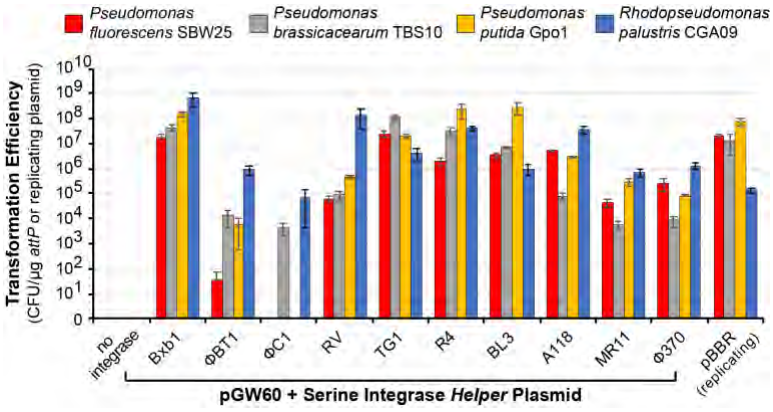


Figure 2. Serine recombinase performance in multiple bacteria.

replicating plasmids (Figure 2) – allowing rapid construction of large genome-integrated combinatorial strain libraries. Unlike other related genome engineering technologies, the machinery for SAGE only requires transient expression from a non-replicating *suicide helper* plasmid for maximal performance. Thus usable, plasmid-free strains are generated in a single step. A secondary serine integrase, ϕC31, can be utilized to excise undesired *E. coli* plasmid backbone DNA from the genome – allowing recycling of antibiotic markers for subsequent genome engineering steps. SAGE performs well in diverse bacteria. Currently, it has been established in a range of Gram-negative bacteria and is actively being developed for Gram-positive bacteria. Here we demonstrate the utility of SAGE by integrating a ~1400-member barcoded constitutive promoter library into 4 hosts. Promoter performance under diverse conditions for each bacteria was assessed using a custom variant of the barcode sequencing method developed by the Wang lab[2]. We identified a collection of *5' UTR insensitive* promoters with a ~40,000-fold dynamic range whose expression was consistent across all tested conditions. We intend to utilize SAGE to understand the genetic determinants of environmental persistence in phylogenetically diverse sorghum bacteria through the development of functional genomics tools and the rapid assessment of root exudate metabolism in heterologous hosts.

References/Publications

1. Elmore, J. R.; Dexter, G. N.; Francis, R.; Riley, L.; Huenemann, J.; Baldino, H.; Guss, A. M.; Egbert, R., The SAGE genetic toolkit enables highly efficient, iterative site-specific genome engineering in bacteria. *bioRxiv* **2020**.
2. Johns, N. I.; Gomes, A. L. C.; Yim, S. S.; Yang, A.; Blazejewski, T.; Smillie, C. S.; Smith, M. B.; Alm, E. J.; Kosuri, S.; Wang, H. H., Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nat Methods* **2018,** *15* (5), 323-329.

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: 3D Cartography of the Sorghum Rhizosphere

Authors: Pubudu Handakumbura*, Albert Rivas-Ubach, Tamas Varga, Anil Battu, Thomas Wietsma, Janet Jansson, Christer Jansson, **Robert Egbert**

Institution: Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory (PNNL), Richland, WA 99354 USA

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 for the model bioenergy crop sorghum through data-driven genome reduction and engineered metabolic addiction to plant root exudates. These principles will lead to secure plant–microbe biosystems that promote secure, stress-tolerant, and highly productive biomass crops.

Abstract Text: 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 recently demonstrated that the root surface is metabolically heterogeneous with hot spots for bacterial attachment[3]. To follow up on these findings and, more broadly, to elucidate details underlying metabolic hotspots for plant-microbe interactions in the rhizosphere we developed a 3D root cartography platform (Fig. 1). The platform maps root exudate metabolites and microbes onto a 3D image of the root, generated by X-ray computed tomography (XCT). Following 3D rendering of the root structure, the root and adhering soil is pulled up from the pot. Loosely adhered soil is removed from the root by gentle shaking while the rhizosphere remains attached to the root surface. The root with its rhizosphere is segmented into short fragments, assigned with a barcode and coordinate position, and separated into root and rhizosphere fractions for subsequent metabolite and microbial profiling using LC-MS and 16S/ITS amplicon sequencing, respectively. To preserve the integrity of the root system architecture once it is removed from the pot, we developed 3D-

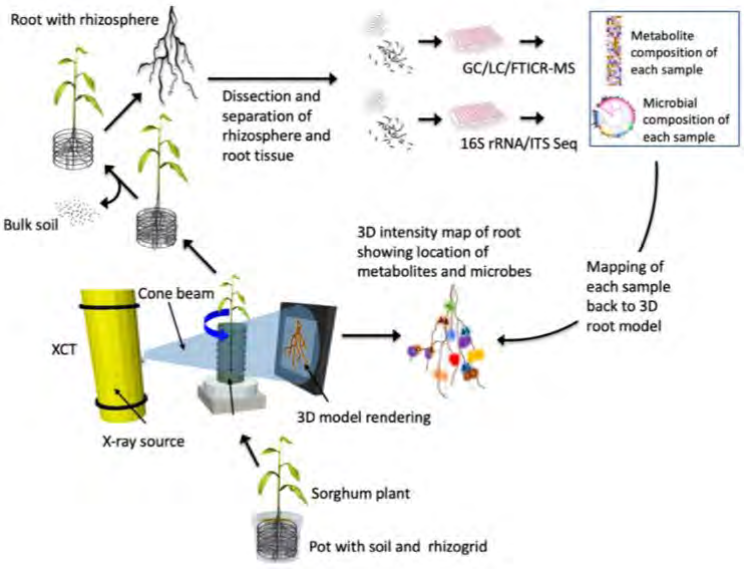


Figure 1. Schematics of the 3D root cartography platform.

printed polymer grids (rhizogrids) that are inserted into the pots prior to filling them with soil (Fig. 1). The rhizogrid also serves to provide coordinates for reconstructing the 3D root image from the excised root segments. Spatially resolved metabolite and microbial data on the root surface and in the rhizosphere are visualized using open-source software[4,5].

In the present study, XCT was used to obtain a 3D image of sorghum (*Sorghum bicolor* (L.) Moench). The imaged root was freed from bulk soil and dissected into 1-cm long segments. The rhizosphere was washed off from each root segment and the rhizosphere solution and corresponding root tissue were collected in micro-vials and kept at -80 °C for future analyses. 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.

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 Protsyuk, I. *et al.* 3D molecular cartography using LC-MS facilitated by Optimus and 'ili software. *Nature Protocols* **13**, 134-154 (2018).

5 Floros, D. J. *et al.* Mass Spectrometry Based Molecular 3D-Cartography of Plant Metabolites. *Front Plant Sci* **8** (2017).

Funding statement: This research was supported in part by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. **??** to the Pacific Northwest National Laboratory (PNNL), and part by Environmental Molecular Sciences Laboratory (EMSL), a DOE user facility at PNNL, as a contribution of the EMSL Strategic Science Area under project 22142. The Persistence Control SFA at PNNL 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 ??**. PNNL is managed by Battelle, LLC for the U.S. Department of Energy under contract no under contract DE-AC05-76RL01830.

Rapid Functional Annotation Using Degenerate Kmers

Jason E. McDermott (Jason.McDermott@pnnl.gov)[1,2], William Nelson[1], Christine Chang[1], Bryan Killinger[1], Joon Lee[1], Arif Khan[3], Sayan Ghosh[3], Mahantesh Halappanavar[3], **Robert Egbert[1]**
[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA;
[2]Department of Molecular Microbiology and Immunology, Oregon Health & Science University, Portland, OR
[3]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland WA

https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

Project Goals:
The Persistence Control Science Focus Area at the Pacific Northwest National Laboratory 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 for the model bioenergy crop sorghum through data-driven genome reduction and engineered metabolic addiction to plant root exudates. These principles will lead to secure plant–microbe biosystems that promote secure, stress-tolerant, and highly productive biomass crops.

Abstract:
Proteins enact the functionality encoded by genomes and so understanding protein function is critical to understanding and predicting how natural and synthetic bacteria would interact with communities like native rhizosphere. Prediction of protein function from sequence is possible because of evolutionary relationships between proteins with similar functions, and existing algorithms can identify the corresponding sequence similarity. However, many proteins have similar functions but diverse sequences, which thwart existing methods, and driven by advances in sequencing technology the number of protein sequences with no known function or similarity to proteins of known function is large and growing rapidly. Additionally, non-genomic data from high-throughput functional screens and multi-omics approaches can be invaluable to providing information about protein function, but currently no methods exist that integrate such information with sequence-based approaches to provide functional annotations for proteins.

Previously we have developed machine learning methods to predict functions for problematic protein families including Type III secreted effectors (*1, 2*), multidrug resistant efflux pumps (*3*), and ubiquitin ligase mimics from bacteria and viruses (*4*). In the current work we describe development of a general, modular pipeline to represent protein sequences as vectors of short peptide sequences (called kmers), using both native and degenerate amino acid encoding to provide flexibility. The pipeline we describe creates these vectors, clusters them based on similarity, analyzes the network structure of the resulting graph, and creates signatures of kmers which best characterize protein families. The modular nature of the pipeline will allow incorporation of information and relationships derived from non-genomic sources of data such as fitness data for multiple environmental conditions being produced by our project using RB-TnSeq mutagenesis, metatranscriptomics and metaproteomics under different environmental

conditions, and functional relationships derived from other sources such as sequence co-evolution.

We describe two applications of our novel pipeline. In the first application we have used high-performance computing to assess the similarities between over 20 million bacterial protein sequences. We encoded the sequences using our pipeline, then calculated pairwise similarity using a GPU-based algorithm (*5*), and used exascale graph analytics to identify clusters of closely related sequences (*6*). The resulting analysis provides a landscape analysis of the bacterial protein universe. We show that this method can recapitulate known relationships between proteins based on traditional means (such as BLAST and hidden Markov models), highlight inconsistencies in the underlying protein database, and provide hypotheses for functions of novel proteins thus providing a large-scale sequence landscape. In the second, we have developed flexible kmer signatures for rapid and accurate classification of nitrogen cycling gene families from metagenomes. We evaluated the performance of our method relative to standard annotation methods in terms of sensitivity and specificity and speed of application.

Though still under development, our open-source, modular pipeline represents an advancement for analysis of large sets of protein sequences, and determination of complementation landscapes for complex microbiomes. The results we present suggest that our approach can provide expanded functional information from metagenomes, and will support integration of multiple other sources of information such as functional screens and omics data.

References

1. R. Samudrala, F. Heffron, J. E. McDermott, Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type III secretion systems. *PLoS Pathog* **5**, e1000375 (2009).
2. J. E. McDermott *et al.*, Computational prediction of type III and IV secreted effectors in gram-negative bacteria. *Infect Immun* **79**, 23-32 (2011).
3. J. E. McDermott, P. Bruillard, C. C. Overall, L. Gosink, S. R. Lindemann, Prediction of multi-drug resistance transporters using a novel sequence analysis method. *F1000Res* **4**, 60 (2015).
4. J. McDermott *et al.*, Prediction of bacterial E3 ubiquitin ligase effectors using reduced amino acid peptide fingerprinting. *PeerJ* **7**, (2019).
5. J. Y. Lee, G. M. Fujimoto, R. Wilson, H. S. Wiley, S. H. Payne, Blazing Signature Filter: a library for fast pairwise similarity comparisons. *BMC Bioinformatics* **19**, 221 (2018).
6. S. Ghosh, M. Halappanavar, A. Tumeo, A. Kalyanaraman, A. Gebremedhin, in *2018 IEEE High Performance Extreme Computing Conference (HPEC)*. (Waltham, MA, 2018).

Funding statement: *This research was supported by the U.S. Department oEnergy (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: Identifying persistence control strategies in plant-growth promoting soil bacteria with RB-TnSeq, CRISPRi, and proteomics

Authors: Aaron J. Ogden[1*] (Aaron.Ogden@pnnl.gov), Ritu Shrestha[1*] (ritu.shrestha@pnnl.gov) Joshua R. Elmore[1], Valentine V. Trotter[2], Adam M. Deutschbauer[2], **Robert G. Egbert**[1]

Institutions: [1]Pacific Northwest National Laboratory, Richland, WA. [2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Website: https://genomicscience.energy.gov/research/sfas/pnnlbiosystemsdesign.shtml

Project Goals: The Persistence Control Science Focus Area at PNNL 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 for the model bioenergy crop sorghum through data-driven genome reduction and engineered metabolic addiction to plant root exudates. These principles will lead to secure plant–microbe biosystems that promote secure, stress-tolerant, and highly productive biomass crops.

Abstract text: Microbial amendments to agricultural soil is becoming an increasingly common strategy to improve crop yield, stress resilience, and carbon sequestration. To prevent escape of these microbes or their genetic contents into unintended agricultural, ecological, or industrial systems, the scientific community requires a better understanding of which genes are dispensable via genomic reduction to limit their ability to grow outside of their intended niche. We hypothesize that identification and genomic reduction of genes essential for growth on common rhizosphere carbon- and nitrogen-sources will result in bacterial strains with novel increased niche-specificity, and are therefore more likely to grow only in their intended environment (e.g. in association with a desired plant). To identify these niche-specific genes and validate their necessity in multiple environments we have implemented LC-MS/MS protein profiling, Random Barcoded Transposon-Sequencing (RB-TnSeq)[1] and CRISPR interference (CRISPRi)[2] using two model plant-growth-promoting soil bacteria, *Pseudomonas fluorescens* SBW25 and *Pseudomonas putida* KT2440. We first identified proteins that are differentially regulated in both strains using protein profiling upon growth in a defined lab media (M9), soil extract (SE), and soil. Our results indicate that each media type results in a unique protein-level metabolic response. However, we also demonstrated a strong correlation between proteins that change significantly during growth in SE and soil, suggesting SE is an appropriate surrogate for soil phenotypes. Further, we showed that KT2440 and SBW25 protein-level responses to both SE- and soil-growth are similar, suggesting both organisms adapt similarly to each environment. We also identified multiple proteins associated with nitrogen and central carbon metabolism were significantly altered in both SBW25 and KT2440 in SE and soil.

We employed RB-TnSeq, a technique that allows rapid and high-throughput screening of mutant populations, to identify genes required for utilization of rhizosphere-relevant carbon and

nitrogen sources, as well as for adaptation to common abiotic soil conditions (e.g. temperature stress). Media types chosen for RB-TnSeq include a defined lab media (RCH2), SE, and soil supplemented with different carbon, and nitrogen sources found in the rhizosphere. Of particular importance is N-acetylglucosamine (GlcNAc), an abundant breakdown product of chitin found ubiquitously in soil. Identification and elimination of GlcNAc catabolic genes is therefore likely to reduce microbial fitness in soil, limiting their niche to other carbon and nitrogen sources (e.g. root exudates of a specific plant). RB-TnSeq successfully identified two GlcNAc importers (Pflu5028 and 2096), as well as catabolic genes (Pflu5025 and 5026) as essential for GlcNAc utilization. We also identified genes necessary for utilization of carbon sources (e.g. mannitol, acetate, lactate, xylose) and nitrogen sources (e.g. urea, nitrate, nitrite), and demonstrate that orthologous genes in SBW25 and KT2440 exhibit similar mutant phenotypes.

To identify gene suppression combinations that control environmental persistence, we developed a multi-gene CRISPRi system in SBW25. Employing single-transcript, multi-guide RNAs for nuclease-dead ddCpf1, we optimized multiplex CRISPRi in SBW25, , including the use of truncated guides [3], by targeting fluorescent protein expression. In addition to green, red, and orange fluorescent proteins, we targeted proteins involved in nitrogen reduction (NarB), urea metabolism (UreC), nitrogen utilization (NtrC) and mucoid production (MucA). Our results indicate that two guides per gene provides higher gene repression compared to a single guide per gene. Also, truncated guides exhibited either improved or equivalent knockdown efficiency for multiple genes when compared to full length guides. We plan to use multiplex CRISPRi in SBW25 to repress up to six different soil-relevant phenotypes simultaneously and we expect these results will guide the creation of genome-reduced strains with engineered environmental niches. Looking forward, we expect our multi-omics and genetics platform to enable persistence control engineering in phylogenetically diverse bacterial isolates from the sorghum rhizosphere.

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).
2. Wang, Mugui., et al. “Multiplex Gene Editing in Rice Using the CRISPR-Cpf1 System.” Mol Plant, Vol 10, Issue 7, (2017): 1011-1013.
3. Fu, Yanfang, et al. "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs." Nature biotechnology 32.3 (2014): 279-284.

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.*

Transgenic Poplar Lines Reveal Host Genes Involved in Defense Against Rust

Gen Li[1], Liang Hu[1], James F. Parsons[1], Gary Coleman[1,2], Shunyuan Xiao[1,2] and **Edward Eisenstein**[1,3,*] (eisenste@umd.edu)

[1]Institute for Bioscience and Biotechnology Research (IBBR), University of Maryland, Rockville, MD; [2]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD; [3]Fischell Department of Bioengineering, University of Maryland, College Park, MD

Project Goals: The objective of this research is to investigate the molecular basis for the virulence of *Melampsora larici-populina* towards *Populus* spp. in order to address the formidable challenge of engineering durable resistance in poplar against leaf rust. The research plan builds on dual host systems and known host defense proteins and pathogen avirulent effectors along with genome-wide approaches to identify key pathogenesis effectors of *Melampsora larici-populina* that target poplar defense and nutrient acquisition. The overarching hypothesis of the project is that rust effectors with homology to other fungal effectors will either bypass the poplar immune system and/or suppress effector-triggered immunity through conserved mechanisms, and that an analysis of these interaction networks will provide for new approaches to develop rust-resistant poplar. The project goals are to mine key rust effectors that suppress host immunity, elucidate interaction networks in poplar targeted by key rust effectors, and to generate transgenic poplar resources for identification of genes involved in pathogen infection or host defense triggered by rust. Our ongoing work on the molecular mechanisms involved in pathogen recognition by the plant innate immune system, the physiological processes that haustorium-dependent pathogens use to commandeer nutrient acquisition systems from the host, are enabling us to construct novel poplar biotypes for use as bioenergy crops. This work is providing new resources to the scientific community to expand the potential for disease-resistant poplar as a bioenergy feedstock.

Abstract: *Melampsora larici-populina* contains a large number of genes that encode candidate secreted effector proteins (CSEPs), which are thought to play significant roles in promoting rust infection in *Populus* spp. CSEPs genes encode small, cysteine-rich secreted proteins that are specifically up-regulated for expression during infection, signifying a key role in host colonization by the pathogen. Two groups of effectors have been the target of this project: (1) 160 *Melampsora larici-populina* CSEPs belonging to 67 structural families, all of which share significant homology among pathogenic rust, *Septoria* or powdery mildew fungi; and (2) a unique 117-member family, whose members show virtually no similarity to any protein in sequence databases. A rapid, quantitative screen was developed to assess *Melampsora* CSEPs that affect poplar immunity. Attenuation (or enhancement) of salicylate levels stemming from changes in HR was measured quantitatively using LC-MS to assess the impact of *Melampsora larici-populina* CSEPs on mounting an immune response. Transient expression of CSEPs along with HR-promoting effectors in tobacco leaves, in poplar leaves, and in poplar protoplasts identified two effects on host immunity: many CSEPs suppress the host immune response, unexpectedly, several promote an immune reaction, triggering a hypersensitive response (HR). In an effort to characterize HR in poplar, pathogenesis related (PR) gene expression was

analyzed in poplar treated with salicylic acid, in transgenic poplar lines expressing HR-promoting *R* genes, and in transgenic poplar lines expressing *Melampsora* CSEPs. Biochemical analysis reveals that pathogenic effectors promote a burst of reactive oxygen species (ROS) that correlates with PR1 gene expression. These lines are being used to identify additional host genes involved in defense against rust, and to evaluate susceptibility mechanisms in compromised hosts.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0017886

Enhanced metadata standards supported by the National Microbiome Data Collaborative

Chris Mungall[1]* (cjmungall@lbl.gov), Faiza Ahmed[2], Anubhav[3], Jeffrey Baumes[2], Jonathan Beezley[2], Mark Borkum[3], Lisa Bramer[3], Shane Canon[1], Patrick Chain[4], Danielle Christianson[1], Yuri Corilo[3], Karen Davenport[4], Brandon Davis[2], Meghan Drake[5], William Duncan[1], Kjiersten Fagnan[1], Mark Flynn[4], David Hays[1], Bin Hu[4], Marcel Huntemann[1], Julia Kelliher[4], Sonya Lebedeva[1], Po-E Li[4], Mary Lipton[3], Chien-Chi Lo[4], Douglas Mans[3], Stanton Martin[5], Lee Ann McCue[3], David Millard[3], Kayd Miller[1], Nigel Mouncey[1], Paul Piehowski[3], Elais Player Jackson[4], Anastasiya Prymolenna[3], Samuel Purvine[3], TBK Reddy[1], Rachel Richardson[3], Migun Shakya[4], Montana Smith[3], Jagadish Chandrabose Sundaramurthi[1], Deepak Unni[1], Pajau Vangay[1], Bruce Wilson[5], Donny Winston[6], Elisha Wood-Charlson[1], Yan Xu[4], **Emiley Eloe-Fadrosh**[1]

[1] Lawrence Berkeley National Laboratory, Berkeley, CA; [2] Kitware, Clifton Park, NY; [3] Pacific Northwest National Laboratory, Richland, WA; [4] Los Alamos National Laboratory, Los Alamos, NM; [5] Oak Ridge National Laboratory, Oak Ridge, TN; [6] Polyneme LLC, New York, NY

https://microbiomedata.github.io/nmdc-metadata/

Project Goals: Short statement of goals. (Limit to 1000 characters)
The National Microbiome Data Collaborative (NMDC) is a pilot initiative launched to support microbiome data exploration and discovery through a collaborative, integrative science gateway. With a community-centered design approach, the NMDC team is building an open-source, integrated data science ecosystem that leverages existing data standards, data resources, and infrastructure within the DOE complex.

Abstract

To understand microbiomes we need to integrate, analyze, and query large amounts of data, including multi-omics data (e.g., metagenome, metatranscriptome, metaproteome, and metabolome) and environmental data. This is challenging because these data are heterogeneous and complex, and existing standards and ontologies are lacking or incomplete.

For the NMDC project, we created a FAIR (findable, accessible, interoperable, and reusable) schema for handling data and metadata of multiple aspects of microbiome data, including environmental metadata about a sample and a study, metadata and provenance for all processing and workflows, and searchable information arising from annotation workflows (for example, functional annotations and results of binning).

The schema leverages and maps to existing standards where appropriate. For describing sample metadata and environmental characteristics, we leveraged the Genomics Standards Consortium (GSC) MIxS (Minimal Information about any Sequence) and use a combination of ENVO

(Environment Ontology) and GOLD used for classifying environments. This includes a mechanism for uniquely identifying source samples using identifier systems such as IGSN (International Geo Sample Number), allowing us to link together data from different omics processing pipeline connected to the same source sample. We extended the W3C PROV standard (https://www.w3.org/TR/prov-overview/) for metadata about computational workflows. For outputs of genomics/transcriptomics workflows, we built on standards such as GFF3, using standardized systems such as KEGG for functional annotation; and for metabolomics/metaproteomics we map to existing ontologies such as PSI-MS where possible.

Our schema weaves together these different standards into a coherent whole. It is rendered as JSON-Schema which allows for precise validation of data input streams using standard validators, as well . We also aim for FAIR compliance by also providing an RDF (Resource Description Framework) version of the schema, including mappings to existing standards.

We used this schema to integrate multiple diverse types of data into JSON-LD files, and to drive search in a web portal.

Funding statement.

This work is supported by the Genomic Science Program in the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) under contract numbers DE-AC02-05CH11231 (LBNL), 89233218CNA000001 (LANL), DE-AC05-00OR22725 (ORNL), and DE-AC05-76RL01830 (PNNL).

Multi-omics workflows to support data integration for the National Microbiome Data Collaborative

Bin Hu[4]* (bhu@lanl.gov), Faiza Ahmed[2], Anubhav[3], Jeffrey Baumes[2], Jonathan Beezley[2], Mark Borkum[3], Lisa Bramer[3], Shane Canon[1], Patrick Chain[4], Danielle Christianson[1], Yuri Corilo[3], Karen Davenport[4], Brandon Davis[2], Meghan Drake[5], William Duncan[1], Kjiersten Fagnan[1], Mark Flynn[4], Marcel Huntemann[1], Julia Kelliher[4], Sonya Lebedeva[1], Po-E Li[4], Mary Lipton[3], Chien-Chi Lo[4], Douglas Mans[3], Stanton Martin[5], Lee Ann McCue[3], David Millard[3], Kayd Miller[1], Nigel Mouncey[1], Chris Mungall[1], Paul Piehowski[3], Elais Player Jackson[4], Anastasiya Prymolenna[3], Samuel Purvine[3], TBK Reddy[1], Rachel Richardson[3], Migun Shakya[4], Montana Smith[3], Jagadish Chandrabose Sundaramurthi[1], Deepak Unni[1], Pajau Vangay[1], Bruce Wilson[5], Donny Winston[6], Elisha Wood-Charlson[1], Yan Xu[4], **Emiley Eloe-Fadrosh**[1]

[1] Lawrence Berkeley National Laboratory, Berkeley, CA; [2] Kitware, Clifton Park, NY; [3] Pacific Northwest National Laboratory, Richland, WA; [4] Los Alamos National Laboratory, Los Alamos, NM; [5] Oak Ridge National Laboratory, Oak Ridge, TN; [6] Polyneme LLC, New York, NY

Project Goals: The National Microbiome Data Collaborative (NMDC) is a pilot initiative launched to support microbiome data exploration and discovery through a collaborative, integrative science gateway. With a community-centered design approach, the NMDC team is building an open-source, integrated data science ecosystem that leverages existing data standards, and data resources and infrastructure within the DOE complex.

Abstract

Standardized omics workflows drive the analysis of raw omics data and ensures the data stored in the National Microbiome Data Collaborative data portal[1] are processed in a uniform fashion and comparable across studies. The NMDC source code repository[2] offers workflows to perform Illumina paired-end reads quality control, metagenomic and metatranscriptomic, metabolomic and metaproteomic analysis. These best practice workflows are developed on top of decades of omics analysis experience gathered from participating institutions, with all computing environment dependencies removed, and coded in the workflow description language (WDL[2]). They are packaged as software containers[3] and documented[4] to enable microbiome researchers to install and run workflows locally, to understand the tools and uses for each workflow, and to further allow local workflow improvements or customisations to meet their specific requirements. By leveraging these workflows, researchers can analyze their data by themselves and expect the same results as if their data were processed by the NMDC portal. A web platform (NMDC EDGE) running these workflows interactively will be provided through the next version of the EDGE bioinformatics suite and similar integration is planned for the DOE KnoweldgeBase (KBase) in the future.

References

[1] https://data.microbiomedata.org/

[2] https://github.com/microbiomedata/

[3] https://www.commonwl.org

[4] https://hub.docker.com/u/microbiomedata

[5] https://nmdc-workflow-documentation.readthedocs.io/en/latest/

Funding statement

This work is supported by the Genomic Science Program in the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) under contract numbers DE-AC02-05CH11231 (LBNL), 89233218CNA000001 (LANL), DE-AC05-00OR22725 (ORNL), and DE-AC05-76RL01830 (PNNL).

The National Microbiome Data Collaborative community survey on microbiome research data

Pajau Vangay[1]* (pvangay@lbl.gov), Faiza Ahmed[2], Anubhav[3], Jeffrey Baumes[2], Jonathan Beezley[2], Mark Borkum[3], Lisa Bramer[3], Shane Canon[1], Patrick Chain[4], Danielle Christianson[1], Yuri Corilo[3], Karen Davenport[4], Brandon Davis[2], Meghan Drake[5], William Duncan[1], Kjiersten Fagnan[1], Mark Flynn[4], David Hays[1], Bin Hu[4], Marcel Huntemann[1], Julia Kelliher[4], Sonya Lebedeva[1], Po-E Li[4], Mary Lipton[3], Chien-Chi Lo[4], Douglas Mans[3], Stanton Martin[5], Lee Ann McCue[3], David Millard[3], Kayd Miller[1], Nigel Mouncey[1], Chris Mungall[1], Paul Piehowski[3], Elais Player Jackson[4], Anastasiya Prymolenna[3], Samuel Purvine[3], TBK Reddy[1], Rachel Richardson[3], Migun Shakya[4], Montana Smith[3], Jagadish Chandrabose Sundaramurthi[1], Deepak Unni[1], Bruce Wilson[5], Donny Winston[6], Elisha Wood-Charlson[1], Yan Xu[4], **Emiley Eloe-Fadrosh**[1]

[1] Lawrence Berkeley National Laboratory, Berkeley, CA; [2] Kitware, Clifton Park, NY; [3] Pacific Northwest National Laboratory, Richland, WA; [4] Los Alamos National Laboratory, Los Alamos, NM; [5] Oak Ridge National Laboratory, Oak Ridge, TN; [6] Polyneme LLC, New York, NY

https://microbiomedata.org/

Project Goals: The National Microbiome Data Collaborative (NMDC) is a pilot initiative launched to support microbiome data exploration and discovery through a collaborative, integrative science gateway. With a community-centered design approach, the NMDC team is building an open-source, integrated data science ecosystem that leverages existing data standards, data resources, and infrastructure within the DOE complex.

Abstract

The NMDC is building a science gateway to enable access to multidisciplinary microbiome data and standardized, reproducible data products by leveraging unique capabilities, expertise, and resources available across Lawrence Berkeley National Laboratory (LBNL), Los Alamos National Laboratory (LANL), Pacific Northwest National Laboratory (PNNL) and Oak Ridge National Laboratory (ORNL). To address the data challenges of the microbiome research community, the NMDC team is using a community-centered design approach, which involves seeking feedback from the community throughout its phases of iterative development. The NMDC team is also applying best practices in user experience design through consultation with the Science Gateways Community Institute (SGCI), an NSF-funded initiative to provide services, resources, community support, and education for creating and sustaining science gateways.

To understand the needs of the microbiome research community, the NMDC team conducted a community survey during December 2020 to January 2021 on how researchers work with data,

access and share data, and use existing resources. The anonymous online survey was distributed broadly with NMDC partner groups, and redistributed within the community.

We present preliminary results from n=768 survey respondents and share a preliminary qualitative analysis by SGCI. We also present several core user groups that emerged from the survey data and briefly discuss how these results will inform the development of the NMDC science gateway.

This survey is an important first step toward understanding how the microbiome research community works with data, and toward establishing an inclusive, bidirectional communication channel between the NMDC and the broader community.

Funding statement

This work is supported by the Genomic Science Program in the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) under contract numbers DE-AC02-05CH11231 (LBNL), 89233218CNA000001 (LANL), DE-AC05-00OR22725 (ORNL), and DE-AC05-76RL01830 (PNNL).

Infective Viruses and Inert Virions: Illuminating Abundant Unknowns in Terrestrial Biogeochemical Cycles

Anneliek M. ter Horst[1+], Christian Santos-Medellín[1+], Sara E. Geonczy[1], and **Joanne B. Emerson[1]*** (jbemerson@ucdavis.edu)

[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 degraded, 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 forests, chaparrals, grasslands, and wetlands. Leveraging a prescribed forest fire and a temperature and atmospheric CO_2 manipulation experiment, we will also explore feedbacks between soil viruses and carbon dynamics in response to environmental change. Finally, through laboratory experiments, we will investigate the composition, fate, and transport of viral particles in soil. By integrating field and laboratory experiments across a variety of soil chemical compositions and spatiotemporal scales, this project will expand our understanding of the soil virosphere and its influence on carbon and nutrient cycling.

Abstract: 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 phosphorous, 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 degraded virions and to assess their respective contributions to soil biogeochemical cycling.

With a focus on bacterial and archaeal viruses with dsDNA genomes, we use viral size-fraction metagenomics (viromics) as our primary approach, separating smaller virions from larger microbes via 0.2 μm filtration prior to DNA extraction and sequencing. By depleting sources of non-viral DNA, we have shown that a far greater diversity of soil viral populations can be recovered from viromes, compared to total metagenomes. However, we do not have a firm understanding of what it means, biologically and ecologically, to recover a viral genome in a virome. Was the virion containing that genome produced yesterday or last year, and is it still capable of infecting a new host, or has it decayed past the point of infectivity? Our comparisons of viromes from the same soils prepared with and without a DNase treatment (to remove free DNA prior to virion lysis) are yielding preliminary insights into the conditions and temporal scales over which virions are produced, remain infective, and decay in soil. Interestingly, when a DNase treatment is not applied, this post-0.2 μm metagenomics approach also seems to enrich for free DNA, allowing for the potential characterization of "relic" DNA in the environment.

Coincident with known decreases in microbial activity during the dry season, we find low virion abundance and diversity in dry compared to wet Mediterranean soils, and free DNA seems to be more abundant in dry soils. Though preliminary, these results are reproducible under field and laboratory conditions in a variety of soils, suggesting new virion production during wet-up and

virion decay during the dry season. Ongoing multi-omics investigations are testing the hypothesis that Mediterranean dry seasons are marked by low viral activity and the accumulation of free DNA, while viral blooms occur throughout the rainy season, with the onset of microbial and viral activity in wet soils fueled in part by free DNA and other necromass as substrates.

We hypothesize that high-temperature fires inactivate virions in near-surface soils. Preliminary results from chaparral and woodland habitats at Quail Ridge Natural Reserve, which burned in the California LNU Complex Fires in August 2020, include DNase-treated viromic DNA yields below detection limits throughout the dry season post-fire. Substantial increases in viromic DNA yields after rain suggest that a bloom of new viral particles appeared in these burned soils at the onset of the rainy season, similar to our results from unburned soils. We infer that, at least for viruses, the timing of fire (during the wet or dry season) and the associated soil moisture content are likely to be important for distinguishing the effects of fire from those of desiccation alone. We are continuing to follow soil viral community recovery in these burned chaparrals and woodlands, which experienced wildfire during the dry season. To assess the impacts of fire during the rainy season, common for management, we will analyze soils before and after a prescribed burn in Blodgett Forest (a mixed conifer forest), scheduled to occur in Spring 2021. Laboratory experiments are also being performed to tease apart the relative effects of fire, temperature, and soil moisture content 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.

To compare the chemical composition of soil virions, bacteria, and lysed host necromass, we are separating and/or enriching each of these fractions from four different soils (forest, chaparral, grassland, and wetland habitats) for a variety of laboratory analyses. Metabolomics and incubations with ^{32}P radiolabeled nucleotides and orthophosphate will be used to compare the chemical and phosphorous contents of these soil constituents. Virion isoelectric points will be measured with a zetasizer, revealing whether viral particles tend to have isoelectric points above, below, or near the pH of their native soils, indicating the relative degrees of potential sorption to minerals and/or transport within soil hydrological conduits. Soil viral community composition will also be compared across a range of buffer pHs, aggregate size fractions, and soil moisture contents to assign traits to viral populations, which will be tracked across our field experiments.

Results from this project will facilitate a better understanding of viral contributions to terrestrial biogeochemical cycling, both through their infections of hosts responsible for carbon and nutrient cycling and as components of soil organic matter.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021198.

Opportunities Linking Omics and Structural Biology at PNNL: Excelling at Cryo-EM

Trevor Moser,[1] Irina Novikova,[1] Amar Parvate,[1] Samantha Powell,[1] and **James E. Evans**[1*]

[1]Pacific Northwest National Laboratory, Richland, WA

Project Goals: This project is focused on the purchase, installation and operation of a new state-of-the-art cryogenic Transmission Electron Microscope (Krios G3i) at EMSL to advance DOE-BER user research in protein/small molecule structural biology and whole cell ultrastructure. The microscope is now fully operational and has started imaging samples from EMSL and DOE-BER users.

Abstract: The acquisition, commissioning and operation of the new Krios G3i instrument at EMSL was a 3-year joint funding venture between EMSL and DOE-BER. This project was designed to rejuvenate cryo-EM research at EMSL by replacing an outdated 25-year old instrument that was limited by analyzing single samples in a manual format to now allowing 12 samples to be loaded and imaged in a semi-automated fashion. Due to the joint funding approach, the new microscope is now available to the general EMSL user community and DOE/BER researchers in a 50/50 split allocation. EMSL users continue to access this new instrument via the normal EMSL user proposal calls which permit combining cryo-EM with other capabilities at EMSL - such as mass spectrometry or super-resolution fluorescence microscopy. Access for DOE-BER users is free of charge to the users as it is funded by this current project which also allows for an expedited submission and review process for cryo-EM only projects.

The KriosG3i microscope has complete screening, data collection and image processing workflows for: 1) micro-electron diffraction of small molecule or protein crystals, 2) single particle analysis of soluble and membrane protein complexes and 3) electron tomography of whole cells or isolated organelles. It is equipped with a K3 direct electron detector, Ceta-D camera, phase plate and Bioquantum energy filter. In addition to semi-automated data collection, we have installed automated image processing workflows for real-time monitoring feedback of session quality and full 3D reconstruction of all workflows. To date we have demonstrated sub-2 angstrom 3D reconstructions for single particle and micro-electron diffraction workflows and sub-nanometer resolution for whole cell tomography. While we can provide very rapid access for samples that arrive pre-frozen on clipped and pre-screened grids, we can also begin with samples that arrive in buffer and require all steps of the cryo-EM workflow. In a subset of cases, we can also start from a provided gene of interest and employ or cell-free expression system to produce enough protein for structural characterization. If users are interested in access to the Krios cryo-EM capability at EMSL, please contact the team listed above or join the poster session to discuss the various mechanisms of access.

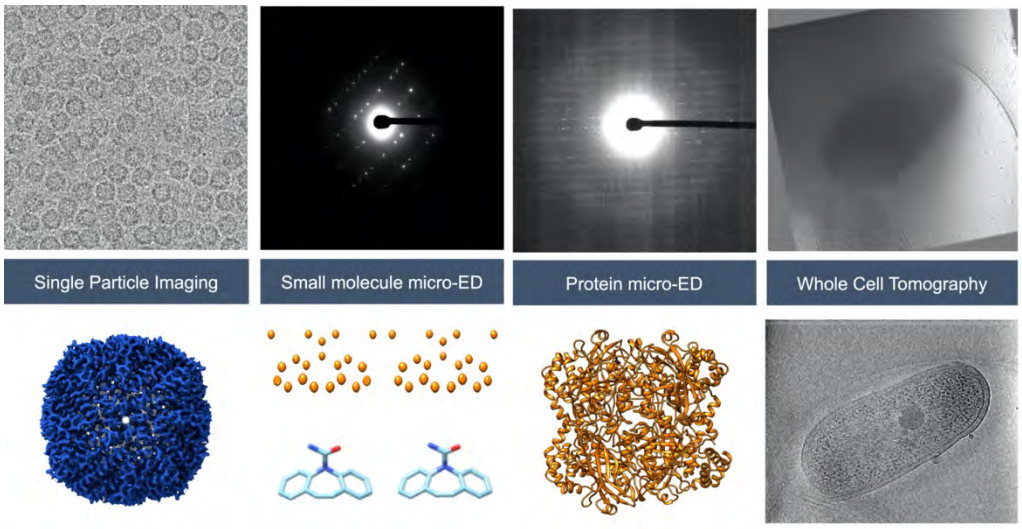


Figure Caption: Benchmarking examples of each of the cryo-EM sample workflows available to users on the new Krios G3i microscope at EMSL. Top row shows raw data while bottom row shows reconstructed volumes.

Funding statement: Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract 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 FWPs 74194 and 74195. The work was performed at EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program.

Nitrogen Addition Impacts on Structure and Function of the Switchgrass Root-associated Diazotrophic Community

Darian Smercina [1,2*] (darian.smercina@pnnl.gov), Lisa Tiemann [1], Alan Bowsher [4], Maren Friesen [5,6], Elizabeth Eder [7], David Hoyt [7], **Sarah Evans [3]**

[1] Michigan State University, Dept. of Plant, Soil and Microbial Sciences
[2] Earth and Biological Sciences Directorate, PNNL
[3] Michigan State University, Dept. of Integrative Biology
[4] Michigan State University, Dept. of Microbiology and Molecular Genetics
[5] Washington State University, Dept. of Plant Pathology
[6] Washington State University, Dept. of Crop and Soil Sciences
[7] Environmental Molecular Sciences Lab, PNNL

http://rhizosphere.msu.edu/

Project Goals: This work aims to improve our understanding of plant-microbe interactions in switchgrass (*Panicum virgatum*) bioenergy cropping systems and their impact on soil biogeochemistry. In particular, we aim to address the role of associative nitrogen fixing organisms (diazotrophs) in meeting switchgrass nitrogen demands.

Associative nitrogen fixation (ANF), the biological conversion of atmospheric dinitrogen gas to bioavailable forms by heterotrophic bacteria, is an important terrestrial N source that occurs under diverse environmental conditions.[1] ANF likely occurs predominately in the rhizosphere, where labile carbon (C) is readily accessible and competition for N between plants and soil microorganisms reduces N availability.[1] Switchgrass, an important bioenergy crop, harbors a diverse community of diazotrophic bacteria in association with its roots and may rely on these organisms as a significant N source when grown on marginal lands.[2,3] It is increasingly clear that diazotrophs are present and fixing N in association with switchgrass, however the impact of soil N availability on these potential N contributions, particularly in fertilized cropping systems, is not known.

In order to understand how soil N availability may impact switchgrass-diazotroph associations and potential N contributions from ANF, we evaluated the switchgrass root-associated diazotroph community and potential ANF rates under long-term and short-term fertilizer N additions.[4] We grew switchgrass in three Michigan marginal land soils in the greenhouse under these N addition treatments for four months before harvest. At harvest, belowground material was subsampled for *nifH* functional gene sequencing and potential ANF rates. In a separate experiment, we also examined the impact of N availability and diazotroph presence (inoculation with *Azotobacter vinelandii*) on switchgrass rhizosphere metabolite chemistry, using data from hydroponically grown switchgrass.[5] Switchgrass seedlings were grown for two-weeks in ¼ strength Hoagland's nutrient solution under high or low N availability. Growth media was then collected and rhizosphere metabolite chemistry was measured via NMR.

We found the switchgrass rhizosphere to exert strong selective pressure on the root-associated diazotroph community. Beta diversity of diazotroph communities in the three Michigan field soils suggest these communities were initially distinct ($R^2 = 0.543$, $p = 0.001$), but the root-associated communities showed little evidence of these site histories ($R^2 = 0.073$, $p = 0.015$). Long-term N addition was not a strong driver of diazotroph community structure ($R^2 = 0.037$, $p = 0.043$), but communities tended to separate by short-term N treatment ($R^2 = 0.0799$, $p = 0.001$). This response to short-term N may be driven by changes in root exudate chemistry as we observed significant differences in switchgrass rhizosphere metabolites under high vs. low N availability. In particular, we noted that high N rhizospheres tended to be dominated by carbohydrates (55.1 %), while organic acids (28.7 %) were the most abundant compounds in low N rhizospheres. Our findings indicate that N availability is likely a driver of diazotroph community structure as well as the forms of C available in the rhizosphere. However, N availability was not a driver of ANF rates in our study. ANF rates were overall highly variable across field soils and N additions, ranging from below detection to over 10 μg N fixed g^{-1} rhizosphere day^{-1}, suggestive of ANF as a hot spot/hot moment process. And while we find evidence for potential association between specific diazotroph community members and ANF rates, community composition was not generally a driver of ANF.

References

1. Smercina, D. N., Evans, S. E., Friesen, M. L., & Tiemann, L. K. (2019). To fix or not to fix: controls on free-living nitrogen fixation in the rhizosphere. Appl. Environ. Microbiol., 85(6), e02546-18.
2. Ruan, L., Bhardwaj, A. K., Hamilton, S. K., & Robertson, G. P. (2016). Nitrogen fertilization challenges the climate benefit of cellulosic biofuels. Environmental Research Letters, 11(6), 064007.
3. Roley, S. S., Duncan, D. S., Liang, D., Garoutte, A., Jackson, R. D., Tiedje, J. M., & Robertson, G. P. (2018). Associative nitrogen fixation (ANF) in switchgrass (Panicum virgatum) across a nitrogen input gradient. PloS one, 13(6), e0197320.
4. Smercina, D. N., Evans, S. E., Friesen, M. L., & Tiemann, L. K. (2020). Impacts of nitrogen addition on switchgrass root-associated diazotrophic community structure and function. *FEMS Microbiology Ecology*, *96*(12), fiaa208.
5. Smercina, D. N., Bowsher, A. W., Evans, S. E., Friesen, M. L., Eder, E. K., Hoyt, D. W., & Tiemann, L. K. (2020). Switchgrass rhizosphere metabolite chemistry driven by nitrogen availability. *Phytobiomes Journal*, PBIOMES-09.

This work was conducted under the MMPRNT project, funded by the Department of Energy (DOE) Biological and Environmental Research (BER) Office of Science award DE-SC0014108. This work was also funded in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409 and DE-FC02-07ER64494), by the National Science Foundation Long-Term Ecological Research Program (DEB 1832042) at the Kellogg Biological Station, and by Michigan State University AgBioResearch. A portion of the research was performed under the Facilities Integrating Collaborations for User Science (FICUS) program and used resources at the DOE Joint Genome Institute (JGI) and the Environmental Molecular Sciences Laboratory (EMSL; grid.436923.9), which are DOE Office of Science user facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under contract numbers DE-

AC02-05CH11231 (JGI) and DEAC05-76RL01830 (EMSL). In particular, the metabolomics profiling measurements were made using EMSL instrumentation (Proposal ID 49977).

PhytoOracle: Leveraging Open-Source Tools for Phenomic Data Processing at Scale

Michele Cosi[1*] (cosi@email.arizona.edu), Emmanuel Gonzalez,[1] Ariyan Zarei,[2] Travis Simmons,[1,3] Duke Pauli,[1] Eric Lyons,[1] and **Andrea Eveland**[4]

[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]Donald Danforth Plant Science Center, St. Louis, MO

https://github.com/LyonsLab/PhytoOracle
https://github.com/phytooracle
https://phytooracle.readthedocs.io/

Project Goals: Within the last decade, phenomics research has seen larger and higher-dimensional data sets. These data sets exceed the capacity for standard computational and analytical frameworks requiring the development of workflows capable of handling these vast amounts of data. To address these challenges, we developed PhytoOracle, aimed at improving processing and analysis of phenomics data. PhytoOracle speeds up analyses by leveraging distributed computing resources, and improves data reproducibility through containerization of computational code. PhytoOracle's distributive abilities and containerized code allow for handling of larger data volumes and modalities making it customizable and scalable. As a result, PhytoOracle can halve the time required to process 1TB of data, accelerating the overall extraction of morphological and physiological parameters. Ultimately, derived phenotypes are quantified and associated to causal genetic components in order to study abiotic stress tolerance.

Abstract: Plant phenomics is the scientific field that aims to quantify and study phenotypic traits through the application of sensor technology and machine learning algorithms. As sensor technology advances, data volume and processing times increase. Due to the lack of open source phenomic pipelines, our team developed PhytoOracle, a modular, scalable pipeline that aims to improve analysis of phenomics data. Our pipeline addresses the issues through (1) accelerating analysis tasks by integrating distributed computing resources and managing high throughput data; (2) containerization of computational code for improved ease-of-use and reproducibility. PhytoOracle expedites data processing by distributing tasks to either local, cloud, or high-performance computing (HPC) systems using CCTools (Albrecht, Donnelly, Bui, & Thain, 2012). Pipeline components are available as Docker containers, providing portability and modularity. Containerized code allows users to execute code without the need to install additional software dependencies, proving to be an efficient solution for deployment. PhytoOracle was developed for, but is not limited to, processing data primarily originating from the world's biggest agricultural robot, the Scanalyzer, located at the University of Arizona's Maricopa Agricultural Center. The Scanalyzer is equipped with sensors that are able to capture a variety of sensor data at sub-millimeter resolution, outputting up to 10 terabytes (TB) of data per

day. These sensors include a laser line scanner, a hyperspectral imager, and three cameras: thermal infrared, RGB, and chlorophyll fluorescence. The capacity of data generated by the Scanalyzer easily surpasses the processing capacity of standard laboratory computers, therefore requiring a quicker data processing solution. PhytoOracle is able to process 1TB of data in half the time a 64-core laboratory computer requires, by distributing jobs across 2,700 HPC cores. The advancement of phenomics demands algorithms capable of processing increasingly large data volumes within a reasonable timeframe. As a result of these key capabilities, PhytoOracle can efficiently process data in a timely manner to extract phenotypic information, which in turn enables faster elucidation of the genetic components of complex traits.

This work is supported by grants DOE-BER #DE-SC00248484, DOE ARPA-E #DE-AR0001101.

References:

Albrecht, M., Donnelly, P., Bui, P., & Thain, D. (2012). Makeflow: A portable abstraction for data intensive computing on clusters, clouds, and grids. SWEET ’12, 1–13. doi:10. 1145/2443416.2443417

Elucidating the Molecular Mechanisms Underlying Drought Resilience in Sorghum

Maxwell Braud,[1] Yuguo Xiao,[1] Indrajit Kumar[1], Philip Ozersky,[1] Rajdeep S. Khangura[2], Vanessa Johnson[1], Abby Stylianou[3], Eric Lyons[4], Brian Dilkes,[2] Duke Pauli,[4] Todd C. Mockler,[1] and **Andrea L. Eveland**[1*] (aeveland@danforthcenter.org)

[1]Donald Danforth Plant Science Center, St. Louis, MO; [2]Purdue University, West Lafayette, IN; [3]Saint Louis University, St. Louis, MO; [4]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.**
- **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 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. Last summer, a sorghum mutant population was 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. State-of-the-art phenotyping data analytics pipelines have been developed as part of this project and DOE-funded initiatives (*see poster by Cosi et al.*) and will be extended to define stress-related phenotypes at multiple scales. Advanced genomics methods are being used to construct network maps that will provide a framework for predicting and investigating gene functions and interrogating differences in the gene regulatory architectures of diverse sorghum genotypes. For example, natural variation in a prioritized candidate gene from a drought-responsive gene regulatory network showed associations with whole-plant water use efficiency in controlled environment (Parvathaneni et al., 2021) and field conditions.

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.

Publications

Parvathaneni RK, Kumar I, Braud M, Ozersky P, Mockler TC, Eveland AL (2021) Regulatory signatures of drought response in stress resilient *Sorghum bicolor*. *bioRxiv*. **doi:** https://doi.org/10.1101/2020.08.07.240580

This work is funded by DOE BER award #DE-SC00248484 and by DOE ARPA-E #DE-AR0001101

Genesearch: A Sequence Similarity Search Service for Genomics Workflows

Jeffrey Johnson[2*], Richard Shane Canon[1], I-Min A. Chen[1], Ken Chu[1], Paramvir Dehal[1], David Lyon[1], Torben Nielsen,[1] Hugh Salamon[1], Elisha Wood-Charlson[1]
Kjiersten Fagnan[1*]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Cohere Consulting, LLC, Seattle, WA
https://code.jgi.doe.gov/jgi-kbase/genesearch

Project Goals: Deploy a suite of composable services to support dynamic genomics data analysis for JGI, NMDC, and KBase Users (Limit to 1000 characters)

Abstract text. Please limit to 2 pages.
As genomics datasets continue to grow precipitously, it has become more difficult for researchers to incorporate new sequences into their workflows in a consistent and reproducible way. One approach to addressing this difficulty is to create a set of "primitive" software services from which one can construct reliable research workflows. Sequence similarity search is a ubiquitous element in these workflows and an obvious candidate for a reusable and composable service. In fact, many institutions already offer such services through web interfaces ([1], [2]).

JGI and KBase both aim to help DOE scientists by allowing them to measure and analyze their data in new ways. JGI and KBase share the need to express increasingly sophisticated relationships in biological data, such as phylogenetic, chemical and environmental similarities, expressed in terms of taxonomy, gene homology, chemical similarity, etc. This common goal requires significant computing resources that increase as we add more organisms, interactions, and environmental parameters to our databases. Focusing first on our users' zero order data analysis use cases, we have identified core pieces of the JGI and KBase infrastructures that can be unified and shared.

JGI and KBase have created Genesearch, a service that provides a similarity search capability with a selection of alignment search tools and databases. Genesearch is structured as a microservice[3] in the sense that it provides exactly one function and is easy to deploy and maintain. Genesearch is currently deployed and available to DOE researchers within KBase and the Joint Genome Institute, and is also available as open source software for research groups that use their own sequence databases. The software can run normally or in a Docker container, and can be accessed by web clients and through a Python interface.

Genesearch is only the first in a suite of composable services for analyzing and manipulating large datasets on DOE and non-DOE resources. We are also working on a service that maps sequence identifiers between databases used by KBase, JGI, UniProt, and NCBI. Here we hope

to solicit input from the research community and to illustrate how these reusable elements can allow scientists to overcome the data deluge problem, build confidence in their work, and focus on answering previously intractable questions.

References

1. NCBI Web BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi (retrieved 1/25/2021)
2. Mirdita M, Steinegger M and Soeding J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics, doi: 10.1093/bioinformatics/bty1057 (2019)
3. Newman S, Building Microservices. O'Reilly Media, Inc, Sebastopol, CA (2015)

Funding statement.

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.

• **References** can be **Publications** instead, if needed. Use any common style for these citations.

Genesearch: A sequence similarity search service for genomics workflows

As genomics datasets continue to grow precipitously, it has become more difficult for researchers to incorporate new sequences into their workflows in a consistent and reproducible way. One approach to addressing this difficulty is to create a set of “primitive” software services from which one can construct reliable research workflows. Sequence similarity search is a ubiquitous element in these workflows and an obvious candidate for a reusable and composable service. In fact, many institutions already offer such services through web interfaces.

We have created Genesearch, a service that provides a similarity search capability with a selection of alignment search tools and databases. Genesearch is structured as a microservice

in the sense that it provides exactly one function and is easy to deploy and maintain. Genesearch is currently deployed and available to DOE researchers within KBase and the Joint Genome Institute, and is also available as open source software for research groups that use their own sequence databases. The software can run normally or in a Docker container, and can be accessed by web clients and through a Python interface.

Genesearch is only the first in a suite of composable services for analyzing and manipulating large datasets on DOE and non-DOE resources. We are also working on a service that maps sequence identifiers between databases used by KBase, JGI, UniProt, and NCBI. Here we hope to solicit input from the research community and to illustrate how these reusable elements can allow scientists to overcome the data deluge problem, build confidence in their work, and focus on answering previously intractable questions.

Cross-Kingdom Characterization of Community Dynamics and C flow in Grassland Soils

Javier A. Ceja-Navarro[1]* (jcnavarro@lbl.gov), Katerina Estera-Molina[2], Mengting Yuan[2], Alyssa Byer[1,2], Aaron Chew[2], Daliang Ning[3], Kateryna Zhalnina[1], Ricky Lewis[4], **Nhu Nguyen**[4], Trent Northen[1], **Jizhong Zhou**[3], **Jennifer Pett-Ridge**,[5] and **Mary K. Firestone**[1,2]

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]University of California, Berkeley, California; Berkeley, California; [3]University of Oklahoma, Norman, Oklahoma; [4]University of Hawaii, Manoa, Hawaii; [5]Lawrence Livermore National Laboratory, Livermore, California.

Website: https://nature.berkeley.edu/crosskingdominteractions/

Project Goals: Our project asks how cross-kingdom and within-kingdom interactions (involving viruses, bacteria, archaea, fungi, protists, microfauna, and plant roots) provide a functional framework for nitrogen (N) cycling in grassland soils. We are using stable isotope probing, NanoSIMS, metagenomic and metatranscriptomic sequencing, exometabolomics, network analysis, and ecosystem modeling to unravel how biotic interactions shape N availability and loss pathways and how these interactions and pathways differ among soil compartments (rhizosphere, detritusphere, hyphosphere, and bulk soil). Our primary goals are to: 1) determine how biotic interactions control key N- cycle transformations, such as depolymerization of macromolecular organic N compounds, N mineralization and immobilization, nitrification, and denitrification, and 2) assess how spatial compartmentalization and transfer between soil compartments determine the occurrences and rates of N-cycling processes.

Decades of research have identified key microbial mediators of terrestrial nutrient cycling, their edaphic sensitivities, and the functional genes and enzymes involved. While many individual aspects of bacterial, fungal, and microfaunal mediation of nutrient cycling are reasonably well understood, these organisms mediate biogeochemical processes in a complex biotic milieu. While soil biotic interactions are recognized, their impacts on nutrient cycling are poorly understood. We posit that cross-kingdom interactions shape carbon (C) and nitrogen (N) cycling and resulting plant/microbe nutrient availability and loss pathways.

In a Californian Mediterranean annual grassland undergoing simulated drought (-50% average precipitation), we collected soil at multiple timepoints corresponding to different plant phenological stages. Total DNA was extracted, and we analyzed communities of bacteria, fungi, and microfauna using amplicon sequencing. Our results show that the diversity of bacterial and fungal communities declined during the growing season and reached its lowest levels during the plants' exponential growth, only to recover to levels similar to the early vegetative stage during the plants' later phenological stages. Community composition analyses showed that sampling time had a significant effect on community structure, while the simulated drought influenced only the composition of bacterial groups. To infer community assembly mechanisms, we used a phylogenetic bin-based null model (iCAMP)[1], which indicated that homogenous selection, dispersal limitation, and drift were the key processes controlling bacterial community assembly, while dispersal limitation and drift were more influential in shaping the assembly of fungal communities. The analysis of microfauna dynamics is ongoing.

Microbial communities can be influenced by many factors, but rhizosphere metabolites are thought to play a particularly significant role in the assembly of root-associated microbiomes. We used liquid chromatography mass spectrometry-based exometabolomics (LC-MS) to analyze rhizosphere metabolite profiles of our mixed annual grassland communities collected in the field. We identified specific metabolites that were more abundant at the early stages of plant development, including organic acids and nucleotides/nucleosides (e.g., salicylate, lactate, 5-methylcytosine). Fifty-six metabolites, including sugars and amino acids, were more prevalent during later plant development (e.g., sucrose, maleic, L-proline). We found that amino acids, organic acids, and nucleotides/nucleosides (e.g., p-coumaric acid, L-threonine, guanine) decreased in abundance in water-limited treatments.

In the same drought simulated plots, we pulse labeled the annual grassland plants with $^{13}CO_2$ and then collected rhizosphere soil for DNA extraction and density-gradient stable isotope probing (SIP). Labeled DNA fractions are being used to identify bacterial, fungal, protozoan and metazoan groups that had access to labeled C through either direct uptake of plant-derived C or predation of microbial cells. Our results from the first five-day labeling event suggest that different groups of organisms consumed ^{13}C plant-derived substrates in the simulated drought versus normal moisture treatment soils. Highly enriched groups in the rhizosphere of soils under drought treatment included several members of the family Burkholderiaceae and the genus *Skermanella*, *Pseudonorcardia*, *Modestobacter*, and *Ramlibacter*. A fungal ASV (Chytridiomycetes) and the bacterivore protist *Cryptodyfflugia* were significantly ^{13}C-enriched in the drought treatment. In the rhizosphere of the normal moisture plants, we observed a larger number of significantly ^{13}C-enriched taxa, from multiple trophic levels. These bacterial and fungal groups included *Parviterribacter*, *Conexibacter*, *Geodermatophilus*, *Microvirga*, *Methylobacterium*, *Agaromycetes*, *Morteriella*, and *Olpidium*. ^{13}C-enriched microfauna in the normal moisture treatment included both protists and nematodes. The protists included the bacterivores *Pseudocyrtolophosis*, *Paracercomonas*, the omnivore *Cercomonas*, the eukaryvore *Bressalua*, green algae *Stichococcus*, and a plant pathogen from the Peronosporales. ^{13}C-enriched nematodes included bacterivores from the class Chromadorea and the plant parasite *Filenchus.*

Our results suggest an important connection between plant phenology and cross-kingdom soil microbial community interactions and demonstrate that multiple trophic groups participate in the movement of C from the rhizosphere into the soil ecosystem. We are measuring gross soil N fluxes and multiple C pools that will be used to further define the contributions of cross-kingdom interactions to biogeochemical cycles through ecological modeling.

1. Ning, D. *et al.* A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* **11**, 4717 (2020).

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 under Award Number DE-SC0016247 and DE-SC0020163 to UC Berkeley. Work at Lawrence Berkeley National Laboratory was performed under the auspices of the U.S. Department of Energy Contract No. DE-AC02-05CH11231, and work at Lawrence Livermore National Laboratory under award SCW 1678 and U.S. Department of Energy Contract DE-AC52-07NA27344.

Title: Belowground Allocation and Dynamics of Recently-Fixed Plant Carbon in a California Annual Grassland Soil

Authors: Christina Fossum[1*] (chfossum@berkeley.edu), Katerina Estera-Molina[1], Maggie Yuan[1], Donald Herman[1], Ilexis Chu-Jacoby[1], Peter Nico[2], Jennifer Pett-Ridge[3], **Mary Firestone[1,2]**

Institutions: [1]University of California, Berkeley; [2]Lawrence Berkeley National Laboratory, [3]Lawrence Livermore National Laboratory

Website: 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. We are using stable isotopes, solid state ^{13}C-NMR spectroscopy, density gradient-based separation, and ecosystem modeling to unravel how biotic interactions shape the flow and fate of C and N in soil. The primary goals of the work discussed here are to: 1) follow the temporal and spatial dynamics of recently fixed plant C and 2) explore the incorporation of plant-derived carbon into the soil organic carbon stock in soil.

The flow and fate of plant fixed carbon (C) was followed for two years after a five-day $^{13}CO_2$ field labeling of a Northern California annual grassland. Soil and plant samples were collected immediately after the labeling pulse, and again at three days, three weeks, six months, one year, and two years. Soil organic matter was fractionated using a sodium polytungstate density gradient to separate the free-light fraction (FLF), occluded-light fraction (OLF), and heavy fraction (HF). Using isotope ratio mass spectrometry, ^{13}C enrichment and total C content was determined for plant shoots, roots, soil, soil dissolved organic carbon (DOC), and the FLF, OLF, and HF fractions. The HF fraction was analyzed by solid state ^{13}C NMR spectroscopy.

At the end of the labeling period, the largest amount of ^{13}C was recovered in plant shoots, but a substantial amount was also found belowground in roots, soil, and soil DOC. Density fractionation of 3-week soil samples (from which living roots were removed) indicated that the highest isotope enrichment occurred in the mineral-rich heavy fraction. After 6 months, when plant shoots and roots had died during the dry summer period, the amount of label in the FLF was equal to that in the HF. By the 1-year sampling, ^{13}C in the FLF had declined substantially and by the end of the 2-year period, 69% of label was in HF, 18% in FLF and 13% in OLF.

While the total ^{13}C content of the HF did not change measurably from the 3-week sample to the 2-year sample, ^{13}C NMR spectroscopic analysis of spring HF samples from 2018, 2019, and 2020 suggests that the relative proportion of aliphatic/alkyl functional groups declined over the 2-year period. Simultaneously, aromatic and carbonyl functional groups increased and the proportion of carbohydrate groups remained relatively constant.

In summary, plant photosynthate C appeared in soil rapidly after being fixed, and by 3 weeks, a substantial amount of the total plant-derived ^{13}C had become associated with the heavy

Fraction (HF) of soil. While the amount of ^{13}C associated with this fraction did not then change measurably over the following 2 years, functional group characterization by ^{13}C NMR spectroscopy showed changing chemical characteristics of the mineral associated organic matter, including declines in aliphatic/alkyls and increases in aromatic and carbonyl functional groups.

References/Publications:

Fossum, Christina, et al. (2020). *Allocation and dynamics of recently fixed plant carbon in a California rangeland soil* (Master's thesis, University of California, Berkeley).

This research was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Awards DE-SC0016247 and DE-SC0020163 to UC Berkeley with subawards to Lawrence Berkeley National Lab and Lawrence Livermore National Lab (award SCW1678).

Carbon Transported by Arbuscular Mycorrhizal Fungi to Soil Alters the Characteristics of Soil Organic Carbon as well as the Soil Microbial Community

Anne Kakouridis[1*] (annekakouridis@berkeley.edu), Maggie Yuan[1], John A. Hagen[1], Christina A. Fossum[1], Rachel A. Neurath[3], Madeline L. Moore[3], Donald J. Herman[1], Peter Nico[3], Peter Weber[2], Jennifer Pett-Ridge[2], and **Mary K. Firestone**[1,3]

[1]University of California Berkeley, Berkeley, CA.[2]Lawrence Livermore National Lab, Livermore, CA. [3]Lawrence Berkeley National Laboratory, Berkeley, CA.

Website: https://nature.berkeley.edu/crosskingdominteractions/

Project Goals: **Our project asks how cross-kingdom and within-kingdom interactions (involving viruses, bacteria, archaea, fungi, protists, microfauna, and plant roots) provide a functional framework for nutrient cycling in grassland soils. We are using stable isotope probing, NanoSIMS, metagenomic and metatranscriptomic sequencing, exometabolomics, network analysis, and ecosystem modeling to unravel how biotic interactions shape nutrient availability and loss pathways and how these interactions and pathways differ among soil compartments (rhizosphere, detritusphere, hyphosphere, and bulk soil). In the work presented here, our primary goals are to: (1) illuminate how arbuscular mycorrhizal fungi (AMF) move carbon (C) beyond the plant rhizosphere into bulk soil and affect the early stages of carbon stabilization, and (2) investigate the influence of AMF on soil bacterial communities and bacterial facilitation of nutrient availability to plants.**

Arbuscular mycorrhizal fungi (AMF) consume up to 20% of plant photosynthetic carbon (C) and grow extensive hyphal networks into the soil; AMF thus have the potential to influence soil C dynamics. Plant root litter, exudates, and microbial residues are thought to be the primary sources of organic C in soil, and they are subsequently transformed into soil organic matter (SOM) by a series of chemical and microbial processes. However, the mechanisms responsible for the persistence of SOM are complex and the roles of specific microbial groups are not well characterized. AMF represent an important pathway for the flow of C from plants into the soil and may thereby alter surrounding soil bacterial communities. These bacteria can colonize hyphae, consume hyphal exudates, and help AMF mobilize nutrients in soil.

We used ^{13}C stable isotope tracing to measure the transfer of C from the host plant *Avena barbata*, a widespread annual grass, into the soil via the AMF, *Rhizophagus intraradices*. In a greenhouse experiment, we used a two-chamber microcosm design to distinguish the fluxes of C from AMF from those of roots. To illuminate how AMF affect the early stages of C accumulation, we tracked $^{13}CO_2$ as it was fixed by host plants and transferred to AMF over the course of six weeks during the exponential phase of plant growth. We assessed changes in C chemistry with solid state ^{13}C nuclear magnetic resonance (NMR) spectroscopy in soil accessible to AMF only. We characterized the form of AMF-contributed C by density gradient fractionation into three pools: a heavy fraction (likely mineral associated), a free light fraction (likely still as free hyphae), and an occluded light fraction (likely hyphae protected in aggregate structures).

Organic C in these soil fractions has distinct rates of biochemical and microbial degradation and these fractions are widely thought to represent ecologically relevant soil subunits.

Using nanoscale secondary ion mass spectrometry (NanoSIMS), we determined that hyphae and roots had a similar level and distribution of ^{13}C enrichment. By isotope ratio mass spectrometry (IRMS), we found that after six weeks of labeling, 26.7 mg of AMF-transported ^{13}C remained in the soil, which accounted for 1.1% of the total soil C pool. Of the ^{13}C that remained in the soil, 17.8 mg or 67% of the in the free light fraction, 2.2 mg or 8% of the total soil C in the occluded fraction, and 6.7 mg or 25% in the heavy fraction. Thus, after six weeks, 33% of the C transported by AMF was in a potentially protected form (mineral associated or aggregate-occluded). ^{13}C-NMR spectra showed a larger carbohydrate peak in the spectra for soil with AMF compared to soil without AMF. This suggests that AMF produced organic compounds in the form of hyphae and/or released metabolites that contained a large proportion of carbohydrates.

We also investigated the influence of AMF hyphae on soil bacterial communities beyond the direct influence of roots, finding that AMF significantly modified the soil bacterial community composition but not diversity. Out of a total of 3019 amplicon sequence variants (ASVs), nineteen ASVs significantly increased and seventeen ASVs significantly decreased in relative abundance in the presence of AMF by DESeq analysis. Over half of the ASVs (and sequences) that responded to the presence of AMF, either positively or negatively, were *Proteobacteria*. A number of the ASVs that increased in relative abundance in the presence of AMF in our study match bacterial taxa that are often found in the rhizosphere, including *Arthrobacter crystallopoietes*, *Caulobacter sp*, *Rhizobium sp*, *Dongia sp*., and two *Verrucomicrobia* taxa.

In summary, AMF moved a substantial amount of C from plants beyond the rhizosphere compartment into the bulk soil. About a third of that C occurred as mineral-associated and aggregate-occluded forms, which may comprise C forms with some initial stability. NMR spectroscopy indicated an increased presence of organic C compounds (such as chitin) that can be associated with AMF hyphae. AMF C inputs modified the bacterial community, potentially enhancing AMF access to N and P.

This research was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Awards and DE-SC0016247 and DE-SC0020163 to UC Berkeley with subcontracts to Lawrence Berkeley National Laboratory and Lawrence Livermore National Laboratory. Work conducted at Lawrence Berkeley National Laboratory was performed under the auspices of the U.S. Department of Energy Contract No. DE-AC02-05CH11231 and work at LLNL, under the auspices of the U.S. Department of Energy under Contracts No. DE-AC52-07NA27344 and DE-AC02-05CH11231.

Soil Viral Community Composition Differs Spatially and in Response to Wet-up in Mediterranean Grasslands

Christian Santos-Medellín[1]* (cmsantosm@ucdavis.edu), Katerina Estera-Molina[2], Mengting Yuan[2], Jennifer Pett-Ridge[3], Mary Firestone[2,4], and **Joanne B. Emerson[1]**

[1]University of California, Davis, California; [2]University of California, Berkeley, California; [3]Lawrence Livermore National Laboratory, Livermore, California; [4]Lawrence Berkeley National Laboratory, Berkeley, California

Website: 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 impacts of drought in a Mediterranean grassland, we harvested rhizosphere-influenced soil samples from 15 experimental plots encompassing two multi-year watering treatments (100% and 50% precipitation since 2017). Collections were performed twice during the 2020 growing season of *Avena barbata*, the annual grass that dominates the ecosystem. To profile the dsDNA viral diversity associated with our 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. By depleting sources of non-viral DNA, this viromic approach facilitates the recovery of a greater richness of viral populations (vOTUs) compared to the recoverable viral diversity from total metagenomes [1, 2].

While precipitation regime and collection time point had significant, albeit minor, effects on overall soil viral community composition, beta diversity trends were largely driven by the location of sampled plots in the field. The observed spatial structuring was defined by a steady turnover of viral populations (vOTUs) along a 16 m transect, with 65% of vOTUs displaying differential abundance patterns impacted by plot position. This distance-decay relationship

highlights potential constraints on the distribution of soil viruses and, possibly, on virus-host interactions at a local scale in these soils. Ongoing characterizations of bacterial and archaeal diversity in these samples should reveal whether the presumed hosts for these viruses exhibit similar spatial patterns, but preliminary data suggest that the observed spatial structuring may be restricted to viral communities.

In a second study, we have focused on rewetting of dry soils, which in Mediterranean ecosystems drives a pulse of CO_2 emissions and a release of inorganic N at magnitudes with global climate implications [3]. Viral activity has been proposed as a contributor to the microbial processes behind this biogeochemical burst, but the specific virus and infected host populations involved have not been identified. To investigate soil viral activity and virus-host dynamics before and after wet-up, while controlling for some of the variation inherent in our field experiments, we have performed laboratory simulations of wet-up. Our preliminary data suggest that viral particle (virion) abundance and diversity are low in dry soils compared to wet soils and, interestingly, that degraded ("relic") DNA may be more abundant in dry soils. For example, 10 days after rewetting dry grassland soils, we observed a >10-fold increase in the number of viral populations detected in viromes and a >8-fold increase in the number of reads recruited to viral contigs in total metagenomes. This viral bloom seems to be a conserved feature of grassland soils, as evidenced by substantial spikes in viromic DNA yields and decreases in free DNA yields as early as 24 hours after rewetting four compositionally distinct grassland soils. Ongoing temporal surveys will identify the microbial populations and metabolic processes impacted by viral predation during wet-up, along with the relative contributions of dormant viruses and desiccation-resistant virions to these dynamics. Future integration of these results with our ongoing studies of bacterial, fungal, and microfaunal mediation of nutrient cycling will bring us closer to an understanding of viral contributions to terrestrial microbial ecology and biogeochemical cycling.

References

1. Santos-Medellin C, Zinke LA, ter Horst AM, Gelardi DL, Parikh SJ, Emerson JB. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *bioRxiv*. 2020. , 2020.08.06.237214 *The ISME Journal (in press)*
2. ter Horst AM, Santos-Medellín C, Sorensen JW, Zinke LA, Wilson RM, Johnston ER, Trubl GG, Pett-Ridge J, Blazewicz SJ, Hanson PJ, Chanton JP, Schadt CW, Kostka JE, and Emerson JB. Minnesota peat viromes reveal terrestrial and aquatic niche partitioning for local and global viral populations. *bioRxiv*. 2020. , 2020.12.15.422944 *Microbiome (in review)*
3. Barnard RL, Blazewicz SJ, Firestone MK. Rewetting of soil: Revisiting the origin of soil CO_2 emissions. *Soil Biol Biochem* 2020; **147**: 107819.

Funding statement: 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 under Award Numbers DE-SC0016247 and DE-SC0020163 to UC Berkeley and UC Davis and award SCW1678 at LLNL. Work at Lawrence Livermore National Laboratory was performed under the auspices of the U.S. Department of Energy Contract DE-AC52-07NA27344.

Identification and Characterization of a New Set of Monocot BAHD Monolignol Transferases

Rebecca A. Smith,[1*](rasmith29@wisc.edu), Emily Beebe,[1] Craig Bingman,[1] Kirk Vander Meulen,[1] Steven D. Karlen,[1] John Ralph,[1] and **Brian G. Fox**.[1]

[1]Department of Biochemistry and Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI

Project Goal: Identify new FMT and PMT BAHD acyltransferases from bioenergy monocot species.

Plants have large BAHD acyltransferase families that perform a wide range of enzymatic tasks in primary and secondary metabolism. Acyl-CoA monolignol transferases, which couple a CoA substrate to a monolignol through an ester linkage, represent a newer class of such acyltransferases. The resulting conjugates may be used for plant defense, but are, importantly, also used as 'monomers' for lignification, in which they are incorporated into the growing lignin polymer chain. These conjugates can add value to the lignin as the phenolate esters can be easily clipped off to provide a valuable aromatic commodity chemical stream. *p*-Coumaroyl-CoA monolignol transferases (PMT) increase the production of monolignol *p*-coumarates, thereby increasing the value of lignin with *p*-coumarate and its byproducts. Other conjugates can improve cell wall digestibility by incorporating mild-alkali-cleavable ester bonds into the lignin polymer backbone. Feruloyl-CoA monolignol transferases (FMT) improve cell wall saccharification, after mild pretreatments, by catalyzing the production of monolignol ferulate conjugates; their incorporation into the lignin generates so-called "zip-lignins". The prevalence of monolignol ferulate conjugates in cell walls throughout the angiosperm clade, particularly throughout the commelinids, suggests that there are more FMT enzymes yet to be uncovered. We used phylogenetics to find potential FMT and PMT enzymes from *Sorghum bicolor* and *Panicum virgatum* based on their similarity to previously identified rice FMT and PMT enzymes. The enzymes were synthesized using the wheat germ cell-free translation system and tested for monolignol transferase activity. Based on these results, we have identified putative FMT and PMT genes for sorghum and switchgrass and have compared their activities to those of known monolignol transferases. Enzyme kinetics of the new and previously identified FMT and PMT enzymes was performed to ascertain which enzyme(s) had the highest catalytic efficiency. These putative FMTs and PMTs were transformed into *Arabidopsis thaliana* to test their activities and to discern the plants' ability to biosynthesize monolignol conjugates *in planta*. Arabidopsis does not naturally produce monolignol conjugates, which simplifies the detection of the novel compounds. The presence of ferulates and *p*-coumarates on the lignin of these transformants indicated that the putative FMTs and PMTs are acting as functional, and efficient, feruloyl-CoA and *p*-coumaroyl-CoA monolignol transferases within plants.

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-SC0020349 to B.G.F.

Extension of Genetic Circuit Design Optimization to Industrially-relevant Organisms

Hamid Doosthosseini[1]* (hdoosth@mit.edu), Jai Padmakumar[1], Ye Chen[1], Bin Shao[1], Yongjin Park[1], Amin Espah Borujeni[1], Adam Arkin[2], Christopher Voigt[1], Carrie Eckert[3,4], and **Ryan Gill[3,5]**

[1]Massachusetts Institute of Technology, Cambridge, MA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Renewable and Sustainable Energy Institute, University of Colorado, Boulder CO; [4]National Renewable Energy Laboratory, Golden, CO; [5]Danish Technical Institute, Copenhagen, Denmark.

web.mit.edu/voigtlab, www.cellocad.org

Project Goals:

- Extend genetic circuit design software (Cello) to species used in bio-production, including yeast (*S. cerevisiae*), *E. coli*, and fast-growing bacteria (*V. natriegens*)
- Develop gates that can be carried stably on the genome without antibiotic selection, as opposed to plasmids
- Design new gates that are more modular and easier to connect by mining small orthogonal repressors from bacteriophage
- Develop methods to integrate –omics data with simulations to identify and correct the causes of failures, predict the impact of carrying the circuit, and determine the impact on metabolic precursors

Abstract text:

Strains used in bio-production, such as for fuels and chemicals, typically are engineered to continuously express the introduced pathways. Genetic circuits have the potential to optimize production by timing when genes are turned on, differentiating cells to perform different tasks, implement feedback control, respond to stress and re-direct metabolic flux based on need. However, there are several problems with genetic circuits that limit their use: 1. they are usually built in laboratory strains, 2. they are carried on plasmids, 3. they lead to a growth defect, and 4. they are difficult to build and often break. Over the last year, we have developed new gates that can be carried in the genomes of industrially-relevant organisms, including *S. cerevisiae*, E. coli and *V. natriegens*. They are carried in "landing pads" that are designed to carry the circuits without interfering with host processes. These data were put into the design automation software package Cello 2.0[1], which includes updated capabilities to be able to work with these organisms. A user can now specify a new circuit design using Verilog and then automatically "re-compile" it for these new hosts simply by selecting them from a drop-down menu. We demonstrate that the genome-encoded circuits can be carried for many generations – over weeks – without breaking or imposing a growth defect.

We are still limited in the scale of the circuits that can be constructed (about 6 repressors) because larger circuits faced unpredictable breakage due to their burden on cellular resources. Several approaches are being taken to overcome this problem. First, new gates are being designed using repressor proteins identified from phage. This has led to the identification of 18 new orthogonal gates that produce near-identical and large responses when carried in the genome. We have also formulated computational models, based on transcriptomics data, to quantify cellular resource usage and predict breakage[2,3]. These tools can also simulate the dynamics of circuit transitions and elucidate transient malfunctions[4].

Collectively, these advances make it easier for a strain engineer to routinely incorporate genetic circuits into their designs. Genetic circuits with significantly more control units can now be automatically designed and assembled, with predictable dynamics and resource usage. More stable circuits can now be easily built with high fidelity on the genome and in new strains supporting a wider range of biosynthesis processes.

References/Publications

1 Nielsen, A. A. et al. Genetic circuit design automation. Science 352, aac7341, doi:10.1126/science.aac7341 (2016).

2 Gorochowski, T. E. et al. Genetic circuit characterization and debugging using RNA-seq. Mol. Syst. Biol. 13, 952, doi:10.15252/msb.20167461 (2017).

3 Borujeni, A. E., Zhang, J., Doosthosseini, H., Nielsen, A. A. & Voigt, C. A. Genetic circuit characterization by inferring RNA polymerase movement and ribosome usage. Nat. Commun. 11, 1-18 (2020).

4 Fontanarrosa, P. et al. Genetic circuit dynamics: Hazard and Glitch analysis. ACS Synth. Biol. 9, 2324-2338 (2020).

5 Park, Y., Espah Borujeni, A., Gorochowski, T. E., Shin, J. & Voigt, C. A. P recision design of stable genetic circuits carried in highly-insulated E. coli genomic landing pads. Mol. Syst. Biol. 16, e9584 (2020).

6 Chen, Y. et al. Genetic circuit design automation for yeast. Nat. Microbiol. 5, 1349-1360 (2020).

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), under contract no. DE-SC0018368

Engineering of Regulatory Networks for Complex Phenotypes in *E. coli* and *S. cerevisiae*

Rongming Liu[1], Liya Liang[1], Emily Freed[1*] (emily.freed@colorado.edu), Alaksh Choudhoury[1], David Romero-Suarez[2], Linas Tamošaitis[2], Brian Weinert[2], Adam Arkin[3], Christopher Voigt[4], Carrie Eckert[1,5], and **Ryan Gill**[1,2]

[1]Renewable and Sustainable Energy Institute, University of Colorado, Boulder CO; [2]Danish Technical Institute, Copenhagen, Denmark; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Massachusetts Institute of Technology, Cambridge, MA; [5]National Renewable Energy Laboratory, Golden, CO.

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.

The sustainable production of biofuels and bioproducts is of continued importance in light of increasing concerns about climate change and energy security. Advances in metabolic engineering, synthetic biology, and systems biology have provided a number of strategies for the more rapid design, construction, and testing of model systems for the production of next generation industrial compounds. However, the titer and productivity of engineered strains are still below those required for economic production. The rate limiting step is no longer our ability to construct designer strains, but rather how to design and engineer increasingly complex networks of combinatorial phenotypes required for the economic and sustainable production of these biofuels and other bioproducts. The core challenges are: i) the general lack of mechanistic understanding required to predictably rewire targeted phenotypes; and ii) the size of the combinatorial mutational space spanning complex phenotypes is much larger than the size that can be searched on laboratory timescales.

To address these issues, we developed a foundation for forward engineering of regulatory control architectures, which combines CRISPR Enabled Trackable Genome Engineering (CREATE; developed in the Gill lab) and forward engineering of *E. coli* and *S. cerevisiae* regulatory networks to access complex targeted phenotypes. We designed, constructed, and mapped libraries of more than 100 regulatory genes containing more than 100,000 specific mutations to perturb the *E. coli* regulatory network. We performed growth competition experiments for library mutants conferring increased tolerance to a variety of industrially-relevant compounds including

furfural, styrene, acetate, isopropanol, and isobutanol. We additionally identified *E. coli* regulatory gene mutants that had increased styrene or isobutanol production. We performed in-depth analyses of the mutations conferring increased tolerance and/or production phenotypes to gain a better understanding of global regulation in *E. coli*.

We additionally mapped over 83,000 mutations in 47 regulatory proteins in *S. cerevisiae* and used these data to identify mechanisms conferring tolerance to isopropanol and isobutanol in yeast. Future work, done in collaboration with the Cell Architecture Lab at the Novo Nordisk Foundation Center for Biosustainability, will focus on designing novel strategies to engineer regulatory networks in *S. cerevisiae* by investigating post-translational modification signaling, such as phosphorylation, ubiquitylation, and acetylation. Post-translational modifications provide a rapid and reversible method to regulate protein functions; therefore, they are an attractive target for modifying protein activities. We additionally aim to build designer libraries to replace endogenous promoters with synthetic promoters to investigate the functional impact of gene dose on organismal traits. Yeast CREATE libraries will be evaluated for industrially relevant stress resistance under model bioreactor conditions.

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science under contract number DE-SC0018368.

Integration of Random Barcode Transposon Sequencing Applications into 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:

In order to accelerate the pace of gene annotation for the numerous novel genes found in recently sequenced organisms, there need to exist widely-available methods that allow us to probe multiple gene functions at once. RB-TnSeq (Random Barcode Transposon Sequencing), a protocol created by the Arkin Lab, allows us to do this by knocking out thousands of genes at random and testing mutant growth under dozens of conditions simultaneously[1]. Our objective is to allow others to easily use and understand RB-TnSeq by making its software available on KBase[2].

Next-generation sequencing has provided biologists with millions of gene sequences with no known function. An ongoing challenge is to find ways to test the functions of multiple genes at once. A proven method to approach this challenge is RB-TnSeq, which uses barcoded transposon insertions to get a library of organisms with different genes knocked out and barcodes to represent the abundance of those organisms. With the recent addition of the RB-TnSeq data-analysis applications to KBase, scientists have a large part of the process simplified (with included data visualizations). Additionally, having the application included in KBase means that the results can be used downstream in applications such as Cello (MIT), which together have the potential to streamline production of high-performing inducible biofuel strains.

The RB-TnSeq applications are divided into three components, each of which returns a statistical summary of the results, a visualization to interpret the results, and all the newly generated files. The statistical summary and visualization make quality assessment simple, and the other files returned to the user allow for custom analysis.

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). https://doi.org/10.1038/nbt.4163

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

Functional genetic screening in the thermotolerant yeast *Kluyveromyces marxianus*

Maria Priscila Lacerda[1]* (mariapriscila.lacerda@colorado.edu), Emily Freed[1], Ian Wheeldon[2], Carrie Eckert[1,3], and **Ryan Gill**[1,4].

[1]Renewable and Sustainable Energy Institute, University of Colorado, Boulder, CO; [2]Chemical and Environmental Engineering, University of California Riverside, Riverside, CA; [3]National Renewable Energy Laboratory, Golden, CO; [4]Danish Technical Institute, Copenhagen, Denmark.

http://gillgroup.org/research/

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. While substantial research has been put toward advancing model hosts such as *Saccharomyces cerevisiae* for chemical biosynthesis, less effort has been made with non-conventional hosts. In this context, we seek to create a functional CRISPR-based genomic screening tool to explore the genetics and metabolism of the yeast *Kluyveromyces marxianus*. This yeast is targeted for this project because it is thermotolerant and natively possesses the ability to uptake and metabolize a range of C5, C6, and C12 sugars. The outcomes of this work will not only help advance our understanding of *K. marxianus* and enable the engineering of new strains but may also lead to new genome editing technologies and improved our fundamental understanding of other non-model organisms.

The key element to enable the utilization of the low pH yeast *K. marxianus* in industrial applications is the development of improved genetic tools to more quickly and effectively manipulate the genome for genotype-phenotype discovery and metabolic engineering. There are efforts to develop these tools in *K. marxianus*, mainly aiming at the modification of metabolic pathways. *K. marxianus* has advantages over other yeast species, mainly because it has greater thermal stability (>40°C), low pH tolerance, a relatively fast growth rate, and a highly resistant cell membrane. Fermentative processes, when carried out at higher temperatures, reduce cooling costs in addition to reducing problems caused by contamination. *K. marxianus* also has a high secretory capacity in relation to *S. cerevisiae*, due to properties such as appropriate glycosylation and strong signal peptides. This strain has been viewed as an alternative to *S. cerevisiae* in 2nd generation ethanol processes as it is able to naturally assimilate a variety of sugars in addition to glucose, such as pentose, hexose, arabinose, cellobiose, lactose and xylose, as well as some toxic compounds present in some sources of lignocellulosic biomass.

In collaboration with JGI's DNA Synthesis team and Ian Wheeldon at UC-Riverside, a genome-wide CRISPR-Cas9 library has been created. This new tool will then be used to define essential genes that support growth on a range of substrates as well as to identify genes essential to high temperature and low pH growth. Library validation and growth screen experiments will require next-generation sequencing amounting to 10 sets. Essential genes necessary for growth on glucose, xylose, glycerol, and lactose at high temperatures (37^{o}- 50^{o}C) will be identified. In addition, other conditions as tests under low pH, such as 2.5, 3 and 3.5, and tolerance to isobutanol will be performed. These screening experiments will require up to 8 different experiments each requiring up to 400 million next-generation sequencing reads. The outcomes of this work will not only help advance our understanding of *K. marxianus* and enable the engineering of new strains, but may also lead to new tools, technologies, and fundamental understanding of other non-model organisms.

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

IMAGINE BioSecurity: Integrative Modeling and Genome-Scale Engineering for Biosystems Security

Katie L. Arnolds[1], Lukas R. Dahlin[1], Jeffrey G. Linger[1], Chao Wu[1], Jianping Yu[1], Wei Xiong[1], Lin Ding[2], Yo Suzuki[2], Cristal Zuniga[3], Karsten Zengler[3], and **Michael T. Guarnieri**[1*] (**Michael.Guarnieri@nrel.gov**)

[1]National Renewable Energy Laboratory, Golden, CO; [2]J. Craig Venter Institute, La Jolla, CA, [3]University of California, San Diego, La Jolla, CA

http://www.nrel.gov/bioenergy/imagine-biosecurity.html

Project Goal: The IMAGINE SFA team proposes to establish a design knowledgebase and achieve predictive control in phylogenetically-diverse, DOE-relevant microbes to enable secure biosystems designs while maintaining optimal performance in controlled environments.

Abstract: Genetically modified organisms (GMOs) have emerged as an integral component of a sustainable bioeconomy, with an array of applications in agriculture and bioenergy. However, the rapid development of GMOs and associated synthetic biology approaches raises a number of biosecurity concerns related to environmental escape of GMOs, detection thereof, and impact upon native ecosystems. In order to establish a secure bioeconomy, novel biocontainment strategies — informed by a fundamental understanding of systems level governing mechanisms — are needed. We have established the Integrative Modeling and Genome-scale Engineering for Biosystems Security SFA Team (**IMAGINE BioSecurity**) to achieve predictive control of engineered systems to enable secure biosystems design. Our team integrates core capabilities in synthetic and applied systems biology to develop a high-throughput platform for the design, generation, and analysis of biocontainment strategies in industrially-relevant and next-generation biocatalysts. IMAGINE will leverage NREL's metabolic engineering and multi-scale omics capabilities in industrial microbial hosts and unique pilot-scale deployment capacity to expand DOE's knowledgebase into deployment-relevant systems. These capabilities are complemented by expertise in synthetic genomics and genome-scale and community metabolic modeling to enable predictive design strategies for next-generation microbial production platforms. The IMAGINE SFA is establishing an extensive library of biocontainment modules and strains, testing platform, and systems knowledgebase. These outputs will lay the foundation for predictive computational design of biocontainment strategies with enhanced stability and resilience in diverse bacterial and yeast hosts, while maintaining maximal fitness and bioproductivity. Combined, these efforts will reduce the risk associated with deployment of engineered biosystems, ultimately enabling a secure bioeconomy.

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.

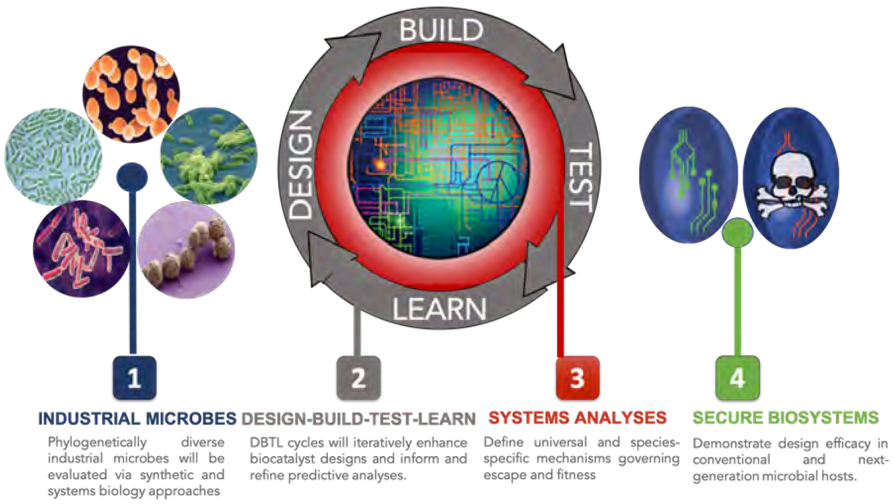


Figure 1. SFA Overview. We will evaluate combinatorial biocontainment strategies in diverse microbial hosts (1). Design-Build-Test-Learn cycles (2) and Systems Level Analyses (3) will be iteratively integrated to enable predictive design and generation of biocontainment modules that reduce microbial escape frequency while minimizing metabolic burden. Successful design will be demonstrated in a series of industrial microbes and next-gen biocatalysts (4).

Nanoscale Dynamics of Cellulase *Tr*Cel7A Digesting Cellulose

Authors: Zachary K. Haviland,[1] Daguan Nong,[1] Zachary K. Haviland[1*] (zbh5056@psu.edu), Kate L. Vazquez Kuntz,[2] Thomas J. Starr,[2] Dengbo Ma,[3] Ming Tien,[3] Charles Anderson,[2] and **William O. Hancock,[1]**

Institutions: [1]Pennsylvania State University, Department of Biomedical Engineering, University Park, PA; [2]Pennsylvania State University, Department of Biology, University Park, PA; and [3]Pennsylvania State University, Department of Biochemistry and Molecular Biology, University Park, PA

https://sites.psu.edu/hancocklab/research/cellulase-motility/

Project Goals: To study the dynamics of TrCel7A as it actively hydrolyzes cellulose using a single-molecule approach. A custom built SCATTIRSTORM microscope was designed to allow sufficient temporal and spatial resolution to accurately study cellulases with nanometer resolution. The overall goal is to create a better understanding of the mechanisms involved when cellulases interact with cellulose by creating a model to characterize the various states the enzymes can enter when binding and degrading the substrate.

Abstract text:

Cellulose is an abundant polysaccharide found in plant cell walls that is digested to cellobiose subunits by cellulase enzymes. Optimizing the digestion of lignocellulose for biofuels applications is of great interest, motivating a need to better understand the molecular mechanism of cellulases, and is under constant investigation due to its ability to be hydrolyzed into glucose subunits. The addition of yeast allows these glucose molecules to be turned into useable biofuels, which can assist in the current energy crisis when performed at an industrial scale. Currently, this process is extremely costly and inefficient, since the energy used to purify the enzymes is only slightly less than the energy these enzymes can produce. Several glucoside hydrolases (GH) are used in this reaction to produce glucose, including the cellobiohydrolase Cel7aCel7A from Trichoderma reesei is a model which is an exoglucanase that degrades cellulose from the reducing-end of strands by cleaving individual cellobiose units as it processes along the cellulose surface. Despite TrCel7aCel7A being one of the most studied glucose cellobiohydrolases, the binding and hydrolysis mechanisms are still not fully understood. Here we took a single-molecule tracking approach to study the kinetics of quantum dot-labeled TrCel7aCel7A on immobilized acetobacter cellulose. 11,116 enzyme molecules were tracked with spatial precision of a few nanometers for hundreds of seconds as they bound to and moved processively along their cellulose substrate. During processive segments, enzymes moved at 3.2 nm/s on average for a distance of 39 nm. The static episodes preceding and following processive runs were of similar duration to one another and were similar to the duration of the static binding events that lacked any processive movement. Although transient jumps of >20 nm were observed, no diffusive behavior indicative of a diffusive search of the enzyme for a free end of a cellulose strand were observed. These data were integrated into a three state model, in which TrCel7A can bind from solution into either a static or a processive state, and the molecules switch between processive and static states

before dissociating. From these data, we conclude that the rate limiting step for this cellulase is transition out of the static state either by dissociating from the cellulose surface or by initiating a processive run.

Funding statement: This work was supported by the Department of Energy Office of Science grant number DE-SC0019065.

Characterizing the degradation of cellulose by combinations of cellulolytic enzymes

Authors: Thomas J. Starr, Kate L. Vazquez Kuntz, and Charles T. Anderson.

Institution: Pennsylvania State University, Department of Biology, University Park, PA.

Project Goals: To study the dynamics and interactions of cellulases during the degradation of cellulose. The overarching goal is to create biochemical models that describe the synergistic interactions between cellobiohydrolases and accessory enzymes as they create cellobiose from cellulose.

http://www.personal.psu.edu/cta3/

Abstract Text:

To better understand factors that hinder cellulose degradation by cellobiohydrolases, we characterized accessory enzymes that act on cellulose in combination with *Trichoderma reesei* Cel7A, a cellulase that targets the reducing ends of glucan chains in cellulose to produce soluble cellobiose. These accessory enzymes include endo-1,4-beta-D-glucanase (Cel7B) from *Trichoderma longibrachiatum* and Cel6A, a cellobiohydrolase II. Cel7B is known for its ability to randomly cleave cellulose to produce additional binding sites for other cellulases, whereas Cel6A cleaves cellobiose units from the nonreducing ends of glucan chains in cellulose. By measuring reducing end production in reactions between cellulose from *Acetobacter* and these enzymes, we found that although Cel7B significantly increases cellulose hydrolysis by Cel7A, the enzyme appears to act in an additive manner rather than a synergistic one. Conversely, Cel6A appears to act synergistically with Cel7A at varying concentrations, contributing to more cellulose hydrolysis than either of the two enzymes alone. Future experiments will involve testing three-enzyme cocktails containing each of the previous enzymes to test whether these combinations allow for greater synergy in degrading cellulose. These data will be used to develop biochemical models to help explain single-molecule observations made for each enzyme acting on cellulose made using a newly developed SCATTIRSTORM microscope.

Funding Statement: This work was supported by the Department of Energy Office of Science grant number DE-SC0019065.

Integrating BER Facilities and Resources for Multi-Omics to Reactive Transport Modeling Workflows

Timothy D. Scheibe[1*] (tim.scheibe@pnnl.gov), Nancy Hess[1], EMSL Summer School Consortium[2], and **Christopher S. Henry**[3]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA; [2]Instructors of the 2020 EMSL Summer School (multiple institutions); [3]Argonne National Laboratory, Lemont IL

https://www.kbase.us/multiscale-microbial-dynamics-modeling/

Microbial community dynamics strongly affect the function of natural ecosystems, but their impacts are difficult to quantify and predict because of challenges related to the wide range of spatial and temporal scales and multidisciplinary integration. Multiple BER user facilities and resources, spanning research programs across both divisions of BER, can be brought to bear on these challenges. This potential was demonstrated in a week-long summer school organized in July 2020 by the Environmental Molecular Sciences Laboratory (EMSL) in close collaboration with KBase, JGI, ESS-Dive, WHONDRS, and individual researchers funded by BER programs. The BER community came together to construct a multi-omics to reactive transport modeling workflow, much of it embodied within the KBase platform, and to demonstrate its use through application to community environmental datasets developed through the WHONDRS consortium employing JGI, EMSL, ESS-DIVE and other BER resources.

This multiscale approach to microbial dynamics modeling combines environmental metabolomics and metagenomics data to formulate complex microbial community reaction networks. These can be solved using metabolic modeling methods such as flux balance analysis, resulting in reaction models and rate formulations that can be incorporated into ecosystem-scale reactive transport codes such as PFLOTRAN. Alternative pathways through the workflow were also demonstrated in the summer school, including the use of thermodynamics-based theories to define reaction rates from elemental composition of organic matter in natural samples provided from high-resolution metabolomics datasets. Public presentations and demonstrations were provided to over 500 globally distributed participants, and 28 selected students participated in intensive hands-on sessions using WHONDRS datasets. Course materials are openly available on KBase at the URL above.

Funding statement: This work is supported as part of the Genomic Sciences Program DOE Systems Biology Knowledgebase (KBase) funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) under Award Numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886. The 2020 EMSL Summer School was sponsored by EMSL, a BER scientific user facility, and by the River Corridor SFA funded by the Environmental System Science Program of BER.

Quantitative Analysis of the Fate of Microbial Residues in Biofuel Soils

Kirsten S. Hofmockel,[1*] Sheryl L. Bell[1], and Christopher P. Kasanke[1]
[*] (kirsten.hofmockel@pnnl.gov)
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 (1) identifying the microbial functions and biopolymers of microbial necromass that contribute to soil C accumulation under controlled conditions, (2) characterizing microbial necromass accumulation in response to crop selection and edaphic factors in situ and (3) generating long-term, cross-site data that can be used to model C cycling in bioenergy cropping systems under different soil conditions.

Molecular-scale biology underpins organismal, community and ecosystem dynamics; yet linking molecular-scale biology to ecosystem outcomes remains a major frontier in deploying sustainable bioenergy feedstock production systems. The overarching objective of this research is to identify how crop selection and soil properties influence the accumulation and stabilization of microbial residues in soil, contributing to soil fertility, reduced CO_2 emissions, and enhanced long-term soil carbon (C) storage. Evidence increasingly supports the concept that most soil organic matter (SOM) is of microbial origin. However, identifying which microbial residues contribute to SOM and what controls their accumulation remains an active area of investigation. Crop selection and soil texture influence the physicochemical attributes of the soil, which structure microbial communities and can influence SOM formation, cycling and long-term storage. To examine the relative importance of biological and geochemical processes, we used a long-term soil incubation and ^{13}C-labeling to test the hypothesis that *microbial community composition and biomass influences microbial derived SOM pool sizes, chemical composition, and ultimately microbial derived SOM accumulation.* By integrating lab to field studies, we aim to identify the molecules and biochemical pathways that contribute to long-term organic matter stabilization in bioenergy soils.

To test this hypothesis, we conducted a long-term lab incubation of soils derived from switchgrass (*Panicum virgatum*) and corn (*Zea mays* L.) plots from the DOE the Great Lakes Bioenergy Research Center (GLBRC) Intensive Biofuel Cropping System Experiments in MI and WI, USA. Soils were collected from silty loams from the Arlington Agricultural Research Station (AARS) in WI and sandy loams from the Kellogg Biological Station (KBS) in MI. Plant-free soil incubations were amended with ^{13}C-labeled glucose, which was rapidly incorporated into microbial biomass; microbial membrane lipids had measurable ^{13}C enrichment after just 24 hours of incubation. The turnover time of microbial cells in soil is estimated to range from 30-50 days; therefore, we harvested the incubation after 2 months to quantify the production and

accumulation of microbial residues. By quantifying ^{13}C accumulation in microbial biomarkers, we can test crop and soil minerology/texture effects on pools of microbial residues.

After 2 months, the incubations were harvested and extracted to measure C concentration and ^{13}C incorporation into microbially derived metabolite, protein and lipid pools as well as unextractable residues associated with solids. We leveraged the sensitivity of PLFA-SIP to quantify the mass and composition of the soil community; microbial DNA was not enriched enough for community profiling DNA-SIP (~1.6 atom % ^{13}C). Microbial biomass was greater in switchgrass compared to corn soils and in the silty compared to sandy soils ($p< 0.001$ for both). Cropping system strongly influenced the profile of the enriched microbial community ($R^2 = 0.33$, $p = 0.002$) and ^{13}C accumulation in microbial biomass ($p < 0.01$). Soils from sandy corn plots had the lowest ^{13}C fungal, G+, and G- bacterial biomass ($p < 0.05$ all). Saprotrophic fungal lipids were the most enriched in ^{13}C, but due to their greater overall biomass, G- bacteria accumulated the most total ^{13}C. Cropping system had the opposite effect on microbial protein, with greater ^{13}C protein in corn compared to switchgrass soils at both sites ($R^2 = 0.82$, $p < 0.001$). The composition of the lipidomes (intact lipids profiles) were markedly different between treatments, with cropping system ($R^2 = 0.28$, $p < 0.001$) having a stronger influence than soil type ($R^2 = 0.11$, $p < 0.001$). Cropping system strongly influenced microbial community and lipidome composition, community biomass and microbial protein concentrations, yet this did not scale to influence overall ^{13}C retention measured in whole soil. Instead, site differences dominated ^{13}C retention, with greater microbial derived ^{13}C in silty (58.7 ± 1.2 μg ^{13}C g soil^{-1}) than sandy soils (49.1 ± 1.8 μg ^{13}C g soil^{-1}). These site effects were largely driven by the small but highly enriched metabolite pool. Although the metabolite pool was the lowest mass pool (< 0.04%) and is generally considered ephemeral, ~13% of the residual ^{13}C accumulated in the metabolite pool after 2 months. Preliminary results suggest the labeled metabolites include a persistent pool of highly ^{13}C enriched trehalose. Our findings reveal that soil texture strongly influences retention of microbial C and cropping systems may enhance organic carbon accumulating pathways indirectly through microbial community selection. Differences in community structure appear to influence the overall metabolic signature, resulting in highly concentrated pools of metabolites and proteins that differ between crop-associated communities. Ongoing efforts include confirming community composition differences using amplicon sequencing, assessing metabolite chemistry differences, and further quantify highly labeled compounds within each pool, in order to understand the mechanisms driving differences in microbial residue retention.

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.

The temperature sensitivity of soil: microbial biodiversity, growth, and carbon mineralization.

Chao Wang [1,2], Ember M. Morrissey [1,*], Rebecca L. Mau [3,4], Michaela Hayer [c], Juan Piñeiro [1], Michelle C. Mack [3], Jane C. Marks [3,5], Sheryl L. Bell [f], Samantha N. Miller [3,5], Egbert Schwartz [3,5], Paul Dijkstra [3,5], Benjamin J. Koch [3,5], Bram W. Stone [3], Alicia M. Purcell[3], Steven J. Blazewicz [6], Kirsten S. Hofmockel [7,8], Jennifer Pett-Ridge [6], **Bruce A. Hungate**[3,5,]

[1] Division of Plant and Soil Sciences, West Virginia University, Morgantown; [2] CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, LN, China;[3] Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff;[4] Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff; [5] Department of Biological Sciences, Northern Arizona University, Flagstaff; [6] Biological Sciences Division, Pacific Northwest National Laboratory, Richland; [7] Physical and Life Sciences Directorate, Lawrence Livermore National Lab, Livermore; [8] Ecology, Evolution and Organismal Biology Department, Iowa State University, Ames

Project Goals: Our overarching objective is to develop and apply 'omics approaches to investigate microbial community processes involved in nutrient cycling; interrogating community and taxon-specific microbial controls over key biogeochemical processes in terrestrial environments, testing quantitative ecological and biogeochemical principles using 'omics data, and including theories of element limitation of microbial growth, growth efficiency, and nutrient use efficiency. Our work will therefore "form and test hypotheses on underlying ecological principles", and "facilitate scaling of concepts and data across multiple levels of biological organization".

Abstract
Microorganisms drive soil carbon mineralization and changes in their activity with increased temperature could feed back to climate change. Variation in microbial biodiversity and the temperature sensitivities (Q_{10}) of individual taxa may explain differences in the Q_{10} of soil respiration, a possibility not previously examined due to methodological limitations. Here, we show phylogenetic and taxonomic variation in the Q_{10} of growth (5-35 °C) among soil bacteria from four sites, one from each of Arctic, boreal, temperate and tropical biomes. Differences in the temperature sensitivities of taxa and the taxonomic composition of communities determined community-assembled bacterial growth Q_{10} which was strongly predictive of soil respiration Q_{10} within and across biomes. Our results suggest community assembled traits of microbial taxa may enable enhanced prediction of carbon cycling feedbacks to climate change in ecosystems across the globe.

This work was supported by the U.S. Department of Energy, Program in Genomic Sciences (DE-SC0020172). Work at LLNL was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344 and Awards SCW1590 and SCW1679

The Phylogenetic Organization of Tundra Bacterial Growth in Response to Short-Term and Long-Term Warming

Jeffrey Propster[1,2]* (jeffrey.propster@nau.edu), **Egbert Schwartz[1,2]**, Michaela Hayer[1], **Ember Morrisey[3], Michelle Mack[1,2]**, **Bruce Hungate[1,2]**

[1]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff; [2]Department of Biological Sciences, Northern Arizona University, Flagstaff; [3]Division of Plant and Soil Sciences, West Virginia University, Morgantown

Project Goals: This work focuses on quantifying the response of soil bacteria to the effects of climate change in the Arctic tundra. Growth rates of individual bacterial taxa are measured using a stable isotope label, and ecological and phylogenetic organization of the warming response is assessed. The experiment is part of a larger project aiming to integrate stable isotope techniques in the field with –omics approaches in order to measure microbial growth and metabolic rates *in situ*, thereby providing an ecological perspective of soil nutrient cycling at the taxon level. We are implementing these new methods to investigate the effects of elevated temperature on soil microbial nutrient transformations across four ecosystems from the Arctic to the tropics. At the Arctic site at Toolik Lake Research Station, specifically, future work will also investigate the effects of climate change on ecological interactions of bacteria and fungi and use stable-isotope-informed metagenomics to identify which soil taxa and metabolic processes are most temperature sensitive.

Increases in Arctic temperatures have thawed permafrost and accelerated soil microbial activity, releasing greenhouse gases that amplify climate warming. Additionally, warming has accelerated shrub encroachment into tundra, altering litter quality and causing further changes in soil microbial processes. We quantified the growth response of individual bacterial taxa to the long-term effects of warming using a 30-year field experiment in tussock tundra. To uncouple the direct effects of warming on bacterial growth rates from effects mediated by other ecosystem responses to warming, we also conducted a parallel short-term experiment (3 months). Intact, active layer soil was assayed in the field for 30 days with ^{18}O-labeled water. Rates of ^{18}O incorporation into DNA were estimated for each taxon as a proxy for growth. Warming doubled bacterial growth rates summed across entire assemblages, with long-term and short-term experiments exhibiting similar aggregate responses, but driven by very different taxa in the short- and long-term warming experiments. The increase in short-term warming was most attributable to emergent growing taxa not detected in other treatments, but the increase for long-term warming was attributable to taxa that co-occurred in the control. Phylogenetic organization of growth was detected to the order level in all treatments. Growth was phylogenetically conserved in a deeper level in control plots (variation explained by class: 38%, order: 14%) than the warming treatments (class: < 1%, order: 54% and 51%), but both the short-term and long-term warming responses (i.e. change in warmed growth rate compared to control) were not

explained by phylogeny and occurred among unrelated taxa that varied widely in their metabolic and physiological traits. Most significant responses were positive and occurred in the long-term warming treatment. In conclusion, tundra bacterial activity greatly increased in response to short-term and long-term warming, and taxa growing in each treatment exhibited deep phylogenetic organization with bacterial orders tending to have similar growth rates. The responders to climate change appear unconstrained by taxonomy nor functional role. Identifying how individual taxa and cohesive phylogenetic groups are affected by climate change may help move toward a mechanistic understanding of ecosystem processes that will influence how much Arctic soils are likely to feed back to the changing climate.

This work was supported by the US Department of Energy, Program in Genomic Sciences, Award Number: DE-SC0020172.

Title: Scaling of ^{18}O-informed Microbial Growth Rates Links Microbial Activity to Soil Carbon Flux

Authors: Bram W Stone,[1,2]* (bram.stone@pnnl.gov), Junhui Li[1], Benjamin J Koch[1,3], Steven J Blazewicz[4], Michaela Hayer[2], Kirsten S Hofmockel[1], Xiao-Jun Allen Liu[5], Rebecca L Mau[6], **Ember Morrissey[7], Jennifer Pett-Ridge[4], Egbert Schwartz[2,3]**, and **Bruce A Hungate[2,3]**

Institutions:

[1]Earth and Environmental Sciences Directorate, Pacific Northwest National Lab, Richland, WA

[2]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ

[3]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ

[4]Physical and Life Sciences Directorate, Lawrence Livermore National Lab, Livermore, CA

[5]Department of Microbiology, University of Massachusetts, Amherst, MA

[6]Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ

[7]Division of Plant and Soil Sciences, West Virginia University, Morgantown, WV

Website: www.nau.edu/ecoss

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. Green'omics sentence here. 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: Global climate projections depend on estimates of soil carbon accumulation and decomposition[1-3], processes driven by microorganisms[3-6]. Given the vast diversity of soil microorganisms, different microbial taxa may have individualistic effects on carbon (C) fluxes in soil, yet testing this idea has been challenging. We identify the growth of individual soil microorganisms using ^{18}O-H_2O quantitative stable isotope probing (qSIP)[7,8] from soils collected along an elevation gradient in Northern Arizona. To understand the microbial ecology of nutrient uptake and nutrient use, we supplied soils with ^{18}O-H_2O either alone (control), or in combination with glucose (C amended) or glucose with $[NH_4]_2SO_4$ (C + N amended). We then converted relative measures of growth to quantitative rates of carbon flux for individual bacteria. Taxon-specific productivity (μg C g soil^{-1} week^{-1}) was then modeled as a function of per-capita growth

rate, taking into account relative abundance, 16S content per unit soil, as well as 16S copy number and genome size (calculated using a bioinformatic approach[9]) to estimate taxon-specific cell size and carbon content. We then modeled bacterial respiration rate (μg C g soil^{-1} week^{-1}) as a function of taxon-specific growth rate and taxon-specific carbon use efficiency (CUE).

Here, we show strong differences in the bacterial taxa responsible for respiration from four ecosystems, indicating the potential for taxon-specific control over soil carbon cycling. Trends in functional diversity, defined as the richness of bacteria contributing to carbon flux and their equitability of carbon use, paralleled trends in taxonomic diversity although functional diversity was lower overall. Nutrient amendment diminished functional diversity, consolidating carbon flow through fewer bacterial taxa. Among genera common to all ecosystems, *Bradyrhizobium*, the Acidobacteria genus *RB41*, and *Streptomyces* together composed 45-57% of carbon flow through bacterial productivity and respiration. We conclude that the bacterial taxa that used the most carbon amendment (glucose) were also those that used the most native soil carbon, suggesting that the behavior of key soil taxa may influence carbon balance. Mapping carbon flow through different microbial taxa as demonstrated here is a crucial step in developing taxon-sensitive soil carbon models that may reduce the uncertainty in climate change projections.

References/Publications

1. Bodman, R. W., Rayner, P. J. & Karoly, D. J. Uncertainty in temperature projections reduced using carbon cycle and climate observations. *Nat. Clim. Chang.* 3, 725–729 (2013).
2. Hurrell, J. W. et al. The Community Earth System Model: A Framework for Collaborative Research. *Bull. Am. Meteorol. Soc*. **94**, 1339–1360 (2013).
3. Luo, Y. et al. Toward more realistic projections of soil carbon dynamics by earth system models. *Global Biogeochem. Cycles* 3**0**, 40–56 (2016).
4. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive earth's biogeochemical cycles. *Science* **80**, 320, 1034–1039 (2008).
5. Schlesinger, W. H. & Bernhardt, E. S. The biosphere - the carbon cycle of terrestrial ecosystems. in Biogeochemistry: an analysis of global change 135–172 (Elsevier, 2013). doi:10.9774/gleaf.9781315625478_4
6. Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Chang*. **3**, 909–912 (2013).
7. Hungate, B. A. et al. Quantitative microbial ecology through stable isotope probing. *Appl. Env. Microbiol.* **81**, 7570-7581 (2015).
8. Koch, B. J. et al. Estimating taxon-specific population dynamics in diverse microbial communities. *Ecosphere* **9**, doi: 10.1002/ecs2.2090.
9. Li, J. et al. Predictive genomic traits for bacterial growth in culture versus actual growth in soil. *ISME J.* **13**, 2162-2172 (2019).

Funding statement: This research was supported by grants from the Department of Energy's Biological Systems Science Division Program in Genomic Science (DE-SC0016207 and DE-SC0020172), and by the National Science Foundation (No. DEB-1645596). Research conducted at Lawrence Livermore National Laboratory was supported by the U.S. Department of Energy Office of Science, via awards SCW1679 and SCW1590, conducted under the auspices of DOE Contract DE-AC52- 07NA27344.

Differential Regulation of Maize and Sorghum Orthologs in Response to the Fungal Pathogen *Setosphaeria turcica*

Pragya Adhikari[1*] (adpragya@illinois.edu), Santiago X. Mideros[1], and
Tiffany M Jamann[1]

[1]University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, 61801

Project Goals: The goals of this project are to understand the relationship between the sorghum and maize defense responses to a common pathogen, *S. turcica,* to identify candidate genes involved in the reaction to *S. turcica* in both maize and sorghum, and to examine the extent of conservation in gene expression in compatible and incompatible interactions between maize and sorghum.

Abstract: Host jumps are a threat to food security, and host resistance is critical for ensuring food security. *Setosphaeria turcica* infects both maize and sorghum and the isolates are host-specific, offering a unique system to examine compatible and incompatible interactions. We hypothesized that resistance mechanisms are conserved between hosts. We conducted transcriptional analysis of maize and sorghum in response to maize-specific and sorghum-specific *S. turcica* isolates and identified functionally related co-expressed modules. Maize had a larger transcriptional response than sorghum. *S. turcica* responsive genes were enriched in core orthologs in both crops, but only up to 16% of core orthologs showed conserved expression patterns. Most changes in gene expression for the core orthologs, including hub genes, were lineage-specific, suggesting that resistance in maize and sorghum evolved largely independently. We identified several defense-related shared differentially expressed orthologs with conserved expression patterns between the two crops, suggesting a role for parallel evolution of those genes in both crops. Many of the differentially expressed genes during the incompatible interaction were related to quantitative disease resistance. This work can inform how to engineer an incompatible interaction and offer insights into how different hosts with relatively recent divergence interact with a common pathogen.

This research was funded by the Department of Energy (DOE) Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0019189 (Plant Feedstocks Genomics) to Tiffany M. Jamann.

Genomic regions associated with pathogenicity in *Exserohilum turcicum* identified by linkage mapping in a biparental population

Pummi Singh*(pummi@illinois.edu), Santiago X. Mideros, and **Tiffany M. Jamann**

University of Illinois at Urbana-Champaign, Urbana

Project Goals:

Our research seeks to identify host-specificity genes in *E. turcicum* with the goal to understand how pathogens evolve to become pathogens of bioenergy crops. Our approach includes linkage mapping of a biparental population of *E. turcicum* strains with specificity to sorghum and maize and analysis of *in planta* and *in vitro* transcriptomes of the host-specific strains.

Abstract text

The filamentous fungus *Exserohilum turcicum* is a significant pathogen of sorghum and maize and is the causal agent of sorghum leaf blight (SLB) and northern corn leaf blight (NCLB). Both diseases can cause significant yield losses, and this pathosystem serves as a model for understanding host jumps. Strains of *E. turcicum* are typically host specific and infect either sorghum or maize. Pathogen host specificity is attributed to a single locus for each host, which facilitates host jumps between maize and sorghum in regions where maize and sorghum are co-cultivated. To test the hypothesis that a single locus in *E. turcicum* underlies specificity to maize and a second locus underlies specificity to sorghum, we generated a biparental population of *E. turcicum.* We crossed strains specific to maize and sorghum, phenotyped the population for virulence on sorghum and maize, genotyped the population to create a linkage map of *E. turcicum,* and located candidate pathogenicity genes. A total of 190 ascospores from 35 pseudothecia were isolated from the biparental cross. Greenhouse phenotyping of the biparental population (n = 144) showed independent inheritance of virulence, as indicated by a 1:1:1:1 segregation for virulence to maize, sorghum, both maize and sorghum, and avirulence to both crops (p=0.09). The population and host specific parent strains were genotyped using genome skim sequencing on an Illumina NovaSeq platform, resulting in over 780 million reads. A total of 32,635 variants including single nucleotide polymorphisms and indels were scored. A genetic map consisting of 17 linkage groups spanning 3,069 centimorgans was constructed. The maize and sorghum host specificity genes mapped to distinct loci on different linkage groups, confirming that a single locus in *E. turcicum* strains confers pathogenicity to sorghum and maize. Fungal transcriptome analysis from *in planta* infections and axenic cultures is in progress. The results from current research are expected to have significant impact on targeted breeding efforts towards improved host resistance and reduction in yield losses of biofuel crops along with a better understanding of the *E. turcicum*- Andropogoneae pathosystem.

Funding statement- This work is funded by DOE award number DE-SC0019189 (Plant Feedstocks Genomics).

Fungal Nutrient Acquisition and Transport in Soil Micromodels is Regulated by Organic Acid Chelation and Specific Membrane Transporters

Arunima Bhattacharjee*[1](arunimab@pnnl.gov), Christopher R. Anderton[1], Dušan Veličković[1], Jocelyn Richardson[2], Odeta Qafoku[1], Lindsey Anderson[1], Sneha Couvillion[1], Zihua Zhu[1], **Kirsten S. Hofmockel[1], Janet K. Jansson[1]**

[1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA; [2]Stanford Synchrotron Radiation Lightsource, Menlo Park, CA

Website: https://www.pnnl.gov/projects/soil-microbiome-science-focus-area

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 are designed to confront spatial challenges and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions. These 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. Data is captured and shared through an optimized data management pipeline. Knowledge gained will provide 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 fungi are integral for soil nutrient cycling and transport due to their widespread interactions with plants and bacteria. Nonetheless, the precise mechanisms by which soil fungi uptake and transport specific mineral nutrients remains unknown. We previously illuminated the mechanisms underpinning directional growth into soil micropore like spaces by the soil fungus *Fusarium sp. DS 682*[1] using a custom designed mineral doped soil micromodel. Here, we expand on this work and reveal the mechanisms underlying fungal mineral derived nutrient transport, where mineral K translocation through fungal hyphae occurs via specific fungal proteins that facilitate passage of organic acid-chelated K.

To investigate fungal nutrient transport proteins, we analyzed *Fusarium sp. DS 682* grown on agar with and without contact with specific soil minerals, such as K - Feldspar and mica. From these samples, several families of cell membrane and mitochondrial transporters were detected in both the presence of minerals (+M) and the absence of minerals (-M). Transporters common under both + M and - M included the major facilitator superfamily transporters, solute carrier transporters, ATP-binding cassette type transporters, and aquaporins. In addition, nine specific transmembrane transporter proteins, such as a drug transporter (K03446), electrochemical potential driven transporters (K16261, K07300), a sugar transporter (K08145), and a xenobiotic transporter (K05658) were detected in our *Fusarium sp.* in

the presence of minerals. Moreover, several fungal proteins associated with organic acid transport (K23630, K03448) were enriched under +M conditions, perhaps for transport of mineral nutrients.

The underlying biology regulating micronutrient acquisition by soil fungi has not yet been identified. Our finding of enriched expression of organic acid transporters in the +M condition suggests that uptake of mineral nutrients was facilitated by fungal secreted organic acids, as has been previously suggested[2] through correlative evidence. To confirm our findings, we grew *Fusarium sp. DS 682* in our previously developed soil micromodels that are compatible with matrix-assisted laser/desorption ionization-mass spectrometry imaging (MALDI-MSI). These devices permitted us to map the location of small molecular weight molecules (e.g., organic acids) in an untargeted fashion using MALDI-MSI. Here, we detected several organic acids, such as citric acid, tartaric acid, malic acid and fumaric acid, on the micromodel surface after fungal growth, and found that the spatial distribution of citric acid and tartaric acid were notably different after fungal growth. Moreover, we detected K-citrate in fungal biomass grown in mineral doped micromodels using micro-X-ray fluorescence (μ-XRF) and X-ray absorption near edge structure (XANES), through our collaboration with the Stanford Synchrotron Radiation Lightsource (SSRL). K-citrate was not detected in fungal hyphae in micromodels without minerals. These results provide direct evidence of a previously uncharacterized mechanism of inorganic K uptake and transport, where fungal hyphae produce organic acids (i.e., citric acid) for uptake of mineral nutrients through organic acid transporter proteins expressed for translocation of mineral derived K+. Moreover, this study provides a pathway for discovery of specific fungal transport mechanisms that contribute toward nutrient translocation under different moisture conditions in soil.

References

1. Bhattacharjee, A. *et al.* Draft Genome Sequence of Fusarium sp. Strain DS 682, a Novel Fungal Isolate from the Grass Rhizosphere. *Microbiol Resour Announc* **10**, (2021).
2. Wallander, H. & Wickman, T. Biotite and microcline as potassium sources in ectomycorrhizal and non-mycorrhizal Pinus sylvestris seedlings. *Mycorrhiza* **9**, 25–32 (1999).

Funding statement: *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (OBER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Soil Microbiome Scientific Focus Area "Phenotypic Response of the Soil Microbiome to Environmental Perturbations." 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.*

Unraveling the Molecular Mechanisms Underlying the Microbiome Response to Soil Rewetting

Mary Lipton[1*] (mary.lipton@pnnl.gov), Karl Weitz[1], Montana Smith[1], Tom Metz[2], Sneha P. Couvillion[2], Kent J. Bloodsworth[1], Vanessa Paurus[1], **Janet Jansson[2], Kirsten Hofmockel[2]**

[1]Environmental Molecular Sciences Division; [2]Biological Sciences Division, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

URL: https://www.pnnl.gov/projects/soil-microbiome-science-focus-area

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 are designed to confront spatial challenges and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions. These 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. Data is captured and shared through an optimized data management pipeline. Knowledge gained will provide 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.

Environmental stress, such as drought, is increasing in frequency with unknown outcomes for soil microbiomes. Therefore, the mechanisms by which microbial communities respond to dehydration are very important to understand. This knowledge is particularly needed for arid, marginal lands, and for otherwise fertile lands that are experiencing reduced precipitation due to climate change. As soils dry we posit that microbes adapt by accumulation of compounds that offset changes in osmotic potential. Alternatively, they die and lyse releasing their intracellular compounds to the environment, thus creating different pools of organic carbon that are available to viable microbes upon rewetting. Of particular interest is the impact of shifts in soil moisture on soil lipids. Despite the central role that microbial lipids play in cellular homeostasis and phenotype, intact lipids have not been widely studied in the context of soil ecology.

Here, we aimed to understand how the physiology, metabolism, and interactions between soil microorganisms change in response to changes in soil moisture, and to use this understanding as a basis for predicting the soil metaphenome. Soil samples were collected from our irrigated field site that is managed by Washington State University, in Prosser, WA. The soil was used to establish soil microcosms that were desiccated to simulate drought. Using a series of incubation experiments, we evaluated the response of the soil microbiome to water addition, by measuring real-time respiration, microbial community shifts and metabolic and physiological adaptation through changes in microbial lipids.

Our results reveal that the soil lipidome is a robust indicator of the microbial community's functional response, even at short time-scales during which community composition may not undergo substantial change. A total of 838 unique lipids were identified from the soil samples across 6 timepoints, representing the largest number of lipids identified in soil to date[1]. Under dry conditions we observed an increase in lipids implicated in mediating

heat, osmotic and oxidative stress and nutrient deprivation. We also found an abundance of lipids containing fatty acid moieties that were characteristic of fungal metabolism. These included higher abundances of non-phosphorus membrane lipids (sulfoquinovosyl diacylglycerols, betaine lipids), ceramides and polyunsaturated fatty acids with longer chain lengths in glycerophospholipids and triacylglycerols. Shifts in the lipidome were not accompanied by significant changes in the fungal community composition. The bacterial community structure was, however, more sensitive to drought and rewetting, as indicated by significant increases in Firmicutes and decreases in Verromicrobia. This shift was accompanied by an increase in lipids with fatty acid characteristics typical of bacteria, such as unsaturated and monosaturated fatty acids with shorter chain lengths, suggesting rapid metabolic reactivation in the bacterial community.

Subsequent experiments targeted the impact of soluble vs solid substrates following rewetting and the relative importance of cellularly associated carbon compounds, or dehydrated DOC that becomes bioavailable after rewetting. By comparing the rate of CO_2 production from desiccated soils that were rewet with water (control) to those where ^{13}C labeled glucose was added in solution and solid forms, we observed the strong influence of bioavailability (soluble v solid) on catabolic metabolism using real-time mass spectrometry[1]. The rate of $^{12}CO_2$ production was 8x that of the $^{13}CO_2$ production in the soluble ^{13}C glucose amendment, indicating that the soil microorganisms were not immediately metabolizing extracellular compounds. The rate of $^{13}CO_2$ production was 3x faster under souble compared to soild ^{13}C glucose addition indicating that highly soluble carbon substrates were the first extracellular compounds metabolized followed by less soluble substrates. Our results suggest that soil microbes were initially accessing carbon native to the soils, followed by soluble, then solid ^{13}C glucose. These findings demonstrate that rewetting dry soil results in preference for easily accessible, perhaps intercellular C, followed by substrates that are dissolved or require enzymatic depolymerization for assimilation. In summary, this study illuminates physiological and metabolic responses to soil drying and-rewetting. In particular, the lipid data suggest that soil microbial communities can quickly react to changes in soil moisture by shifting their lipid compositions. This may have important implications for carbon allocation within and between organisms and the soil environment.

References:

1. Couvillion, S.Pl, Naylor, D., Smith, M.L., Paurus, P.L., Bloodsworth, K.J., Stratton, K.G., Metz,T.O., Lipton, M.S., Jansson, J.K., Hofmockel, K.S. Influence of soil moisture stress on the soil microbiome and lipidome (2021, in preparation).
2. Weitz, K.K., Smith, M.L., Hixson, K.K., Hill, E. A., Jansson, J.K., Hofmockel, K.S., Lipton, M.S. (2020) Real-Time Mass Spectrometry Measurements of Respiration Rates in Biological Systems. *Journal of the American Society for Mass Spectrometry*. Dec 1. https://doi.org/10.1021/jasms.0c00251

***Funding statement*:** This work was supported by the Department of Energy (DOE) Office of Biological and Environmental Research (BER) and is a contribution of the Scientific Focus Area "Phenotypic response of the soil microbiome to environmental perturbations". PNNL is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830. Portions of this work was performed in the W. R. Wiley Environmental Molecular Sciences Laboratory (EMSL) (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory (PNNL).

Title: Interspecies Interactions During Chitin Decomposition in a Naturally Enriched Model Soil Microbial Consortium.

Authors: Ryan McClure[1*] (ryan.mcclure@pnnl.gov), Yuliya Farris[1], Michelle Davison[1], Natalie Sadler[1], Hyun Seob Song[2,3], Robert Danczak[1], William Nelson[1], Ruonan Wu[1], Dan Naylor[1], **Kirsten S. Hofmockel[1,4], Janet K. Jansson[1]**.

Institutions: [1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA; [2]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA; [3]Nebraska Food for Health Center, Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, NE, USA; [4]Department of Agronomy, Iowa State University, Ames, IA, USA

Website: https://www.pnnl.gov/projects/soil-microbiome-science-focus-area

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 are designed to confront spatial challenges and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions. These 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. Data is captured and shared through an optimized data management pipeline. Knowledge gained will provide 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.

Chitin is one of the most abundant carbon and nitrogen sources available in nature. However, its complex chemistry means that complete breakdown of this polymer into constituent n-acetyl-glucosamine (NAG) monomers or other compounds is energetically expensive. As a result, interspecies interactions are central to both the complete breakdown of chitin and dispersion of the carbon and nitrogen within this molecule throughout the soil microbiome. While these interactions are likely critical to organic carbon cycling in soil the molecular mechanisms are not well understood. A better understanding of the role of soil microorganisms in nutrient breakdown is critical if we wish to harness the microbiome to promote plant and ecological health and to gain a better understand of large-scale carbon and nitrogen cycling. Here, we focused on specific species and metabolic interactions that lead to the breakdown of chitin. To study these interactions we developed a naturally evolved simplified model soil consortium (MSC-1) containing ~20 species that were enriched under high chitin conditions (1). Network analysis of this community indicated the potential for several interspecies interactions. Here, we explored specific chitin degrading phenotypes by isolating several constituent strains from MSC-1 and carrying out genomic and phenotypic analyses to decipher species interactions during growth on chitin.

We isolated several axenic strains from MSC-1 and sequenced their genomes to enable us to predict the metabolic capacity of each isolate member. Eight of the isolates were chosen to represent the diversity of the original MSC-1 community and designated as Model Soil Consortium 2 (**MSC-2**). MSC-2 contains *Streptomyces venezuelae*, *Neorhizobium galegaae*, *Dyadobacter fermentans*, *Sphingopyxis fribergensis*, *Ensifer adhaerens*, *Variovorax paradoxus*, a *Sinorhizobium* species. and a *Rhodococcus* species. This collection of species represents six different families, five different orders and three different phyla. Highly abundant and central members of our original MSC-1 consortium, including *Streptomyces venezuelae*, *Ensifer adhearens* and *Rhodococcus* sp., are included in MSC-2. Subsequent genomic analyses of members

of MSC-2 demonstrated that all MSC-2 strains contained a subset of chitinase genes. However, the genetic potential for complete chitin degradation varied between strains. In addition, a subset of strains contained metabolite transporters for chitin breakdown products, including NAG. These observations point to the potential for metabolic complementarity regarding chitin breakdown by MSC-2.

We tested the ability of each MSC-2 strain to grow using chitin as the sole source of carbon. Each strain was grown axenically in M9 media supplemented with ammonium and chitin. Under these conditions we found that only *S. venezuelae, N. galegaae* and *S. fribergensis* were unable to grow on chitin. While the remaining six strains did show an increase in biomass under chitin, the kinetics of each strain's growth were very different. We next examined co-culturing pairs of MSC-2 species to determine if there was greater growth (as measured by optical density) when species were cultivated together vs. axenically. We found that when *S. venezuelae* (a non-grower on chitin) was cultured with *V. paradoxus* (which can grow on chitin) the O.D. of the two-member community was higher than either species grown together. In addition, when two species, both found to be unable to grow on chitin (*S. venezuelae* and *N. galegaae*), were cultured together growth was observed in the co-culture. These observations speak to metabolic interdependencies between species of MSC-2 during chitin degradation. These interdependencies were further explored through fluorescently tagging two of the strains with GFP variants: the *Rhodococcus* and *Sinorhizobium* sps. Fluorescently tagged *Rhodococcus* sp. and *Sinorhizobium* sp. were co-cultured with chitin then treated with NAG-ELF, a chemical tool designed to label functionally active chitinase. Hydrolyzed NAG-ELF was found to be nearly exclusively co-localized with *Rhodococcus* sp. cells, indicating that the *Rhodococcus* sp. chitinases are cell associated and that the *Rhodococcus* sp. is supplying both members with chitin degradation products.

The knowledge gained about interspecies interactions during chitin decomposition that we are obtaining in our model consortia are relative to understanding details of SOM degradation in natural soil ecosystems because chitin is an abundant source of both carbon and nitrogen in soil. It is important to note that all strains of MSC-2 were derived from our naturally evolved MSC-1 so the interactions we observe here likely reflect a subset of those occurring in soil. Our results indicate that even in a community where chitin degradation potential is ubiquitous only a subset of species actually carry out chitin degradation. To provide more mechanistic understanding of these results, we initiated species-resolved community metabolic networks using the KBase pipeline that were based on individual genome-scale metabolic models for the nine MSC-2 isolates. These models revealed that the predicted growth patterns were largely consistent with experimental observations. Our future research is focused on refining details of metabolic and signaling interactions during chitin decomposition, and the role of changes in soil moisture on those interactions.

References:

1. **McClure R, Naylor D, Farris Y, Davison M, Fansler SJ, Hofmockel KS, Jansson JK.** 2020. Development and Analysis of a Stable, Reduced Complexity Model Soil Microbiome. Front Microbiol **11:**1987.

Funding statement**:** *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (OBER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Soil Microbiome Scientific Focus Area "Phenotypic Response of the Soil Microbiome to Environmental Perturbations". 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.*

Prediction of Spatial Assemblies in Soil Microbiomes Governed by Interspecies Interactions and Environmental Gradients

Hyun-Seob Song,[1]* (hsong5@unl.edu), Natalie C. Sadler,[2] Joon-Yong Lee,[2] Arunima Bhattacharjee,[2] Christopher R, Anderton,[2] Ryan S. McClure,[2] Elias Zegeye[2,3], **Kirsten S. Hofmockel**,[2] and **Janet K. Jansson**[2]

[1]University of Nebraska-Lincoln, Lincoln, NE; [2]Pacific Northwest National Laboratory, Richland, WA; [3]The Gene and Linda Voiland School of Chemical Engineering and Bioengineering, Pullman, WA

https://www.pnnl.gov/projects/soil-microbiome-science-focus-area

Project Goals: PNNL's Phenotypic Response of Soil Microorganisms 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 are designed to confront spatial challenges and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions. These 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. Data is captured and shared through an optimized data management pipeline. Knowledge gained will provide 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.

Heterogeneous micro-patches in soil lead to spatial distributions in microbial communities. The formation of spatial patterns is not random but governed by numerous biotic and abiotic factors, including metabolic and chemotactic traits of individual members, interspecies interactions, and environmental factors. It is critical to understand how environmental constraints control the formation of specific spatial patterns through the modification of interspecies spatial interactions. However, elucidating key biological and environmental parameters governing specific spatial patterns of interacting species remains challenging, partly due to the lack of reliable computational and instrumental tools that can be effectively used for this purpose. Toward filling this gap, here we present coordinated modeling and experimental approaches for predicting and visualizing spatial formation of microbial assemblies.

First, we developed an agent-based model to predict spatiotemporal evolution of interacting species at the microscale (μm to mm)[1]. The simulated spatial patterns obtained from the agent-based model were subsequently used to train deep learning models to infer microbial interactions. The resulting machine learning model accurately inferred the spatial variation in microbial interactions in heterogeneous environments as demonstrated through a case study of *Pseudomonas fluorescens* and *Escherichia coli*. This co-culture developed context-dependent interactions when treated with polymeric chitin or N-acetylglucosamine (NAG, the hydrolysis product of chitin) because *P. fluorescens* can use both substrates for growth, while *E. coli* lacks the ability to degrade chitin.

To extend the modeling scope of spatial interactions to larger scales (up to cm), we developed a multi-component, multi-species reaction-diffusion model. Unlike agent-based models where microbial cells are treated as discrete particles, reaction-diffusion models provide a continuous description of the distribution of microorganisms as well as chemical nutrients in a multi-dimensional space. We formulated this model to simulate spatial interactions and assemblies of multiple isolates derived from our Model Soil Consortia[2]. Key variables we accounted for this purpose included chemotactic motion of microorganisms, spatial gradients of nutrients (such as chitin or NAG), and moisture levels. We parameterized basic metabolic and chemotactic features of species based on (1) genome-scale metabolic network models of isolates built using the DOE's KBase pipeline[3], and (2) microbial growth data using microchip experiments.

Predicted spatial interactions and microbial assemblies were also experimentally validated using visualization techniques. As a proof-of-concept study, we developed a dynamic model of a *Cellvibrio japonicus* and *E. coli* consortium to simulate the interactions among degraders (chitin-degrading *C. japonicus*), lazy friends (non-chitin-degrading *C. japonicus*), and cheaters (*E. coli*) in the homogeneous environment. In addition we tested and optimized conditions for labeling chitinase enzymes *in vivo* with a chitinase activity-based probe (chitotriose-ABP). Beyond predicting the optimal proportion of degraders in the presence and absence of cheaters, this model enabled the evaluation of the ecological benefit of spatial assemblies of soil microorganisms in promoting community function.

New developments of modeling and visualization methods in this work provide an improved capability of (1) predicting the evolution of soil microorganisms in spatially constrained environments, and (2) inferring interspecies interactions from the resulting image data. Incorporation of genome-scale metabolic networks into agent-based and reaction-diffusion models for more mechanistic simulations are currently in progress.

References:

1 Lee, J.-Y. *et al.* Deep learning predicts microbial interactions from self-organized spatiotemporal patterns. *Computational and structural biotechnology journal* **18**, 1259-1269 (2020).

2 McClure, R. *et al.* Development and Analysis of a Stable, Reduced Complexity Model Soil Microbiome. *Front Microbiol* **11**, ARTN 1987, doi:10.3389/fmicb.2020.01987 (2020).

3 Henry, C. S. *et al.* Microbial Community Metabolic Modeling: A Community Data-Driven Network Reconstruction. *J Cell Physiol* **231**, 2339-2345, doi:10.1002/jcp.25428 (2016).

Funding statement: *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (OBER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Soil Microbiome Scientific Focus Area "Phenotypic Response of the Soil Microbiome to Environmental Perturbations". 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.*

Title: Soil Moisture Impacts Composition, Activity and Ecosystem Functions of Soil DNA and RNA Viruses

Authors: Ruonan Wu[1] (ruonan.wu@pnnl.gov), Michelle R. Davison[1], William C. Nelson[1], Yuqian Gao[1], Carrie D. Nicora[1], Jason E. Mcdermott[1], Kristin E. Burnum-Johnson[1], **Kirsten S. Hofmockel[1, 2], Janet K. Jansson**[1]

Institutions: [1]Earth and Biological Sciences Directorate, Pacific Northwest National Lab, Richland, WA, USA; [2]Department of Agronomy, Iowa State University, Ames, IA, USA

https://www.pnnl.gov/projects/soil-microbiome-science-focus-area

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 are designed to confront spatial challenges and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions. These 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. Data is captured and shared through an optimized data management pipeline. Knowledge gained will provide 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.

Soil is known to harbor a large diversity of viruses, but the majority are uncharacterized and how the soil environment impacts the composition, activity and ecosystem functions of soil viruses is unknown. Here, we screened for viral sequences in some of the largest soil metagenomes produced to date (>1 Tb each) from a range of grassland soils with different historical precipitation patterns - from arid soils in eastern Washington to wet soils in Iowa and intermediate soils in Kansas. Due to the unprecedented size of the metagenomes, we collaborated with NERSC to assemble them[1]. Screening the assemblies resulted in a total of 2,631 viral contigs including 14 complete circular viral genomes[2], which were grouped into 214 clusters based on the protein sharing matrix and tetranucleotide frequency. A comparison of the viral communities across each site revealed that soil with lower historical moisture harbored significantly higher viral diversity and abundance, while also displaying less evidence of virus-host interactions (through CRISPR-Cas spacers). These seemly contradictory findings may reflect a predominance of viruses with lysogenic strategies in drier soils. Across the diverse grassland DNA virosphere, 45 of the viral clusters were found in all three grasslands. These viral 'common clusters' also featured high abundances, targeted the dominant bacterial taxa, contained putative viral generalists and were more frequently targeted by host CRISPR-Cas spacers. We hypothesize that these 'common clusters' contain viruses that are well adapted to grassland soils and represent the potential viral keystone species in these soils. In addition, auxiliary metabolic genes (AMGs) involved in 18 pathways were identified, with many detected on the complete viral genomes we assembled from the grassland metagenomes, suggesting potential novel viral contributions to carbon metabolism and energy acquisition in soil.

To further investigate the response of DNA and RNA viruses to changes in soil moisture, we collected samples from the Kansas grassland with a historical intermediate soil moisture profile. Three replicate samples from two field locations were either wet to saturation or air-dried for 15 days at 21°C to represent experimental wet and dry soil treatments, respectively. Paired expression data

(metatranscriptomes and metaproteomes) were mapped back to the assembled viral contigs to identify active DNA viruses and to determine their phenotypic responses to soil moisture extremes. The vast majority of actively transcribed DNA viruses were bacteriophage, but some were assigned to eukaryotic hosts, mainly insects[3]. We observed that higher soil moisture increased the transcription of a subset of DNA viruses. In contrast, there was a lower overall level of transcription in drier soil, but across a larger range of DNA viruses. A higher percentage of non-coding RNAs and more transcripts of lysogenic markers (i.e. integrases and excisionases) were also detected in dry soil, further supporting a higher prevalence of lysogenic viruses in more arid soils. We also detected peptides encoded by viral early and late genes that are known to be upregulated during an active viral infection. To our knowledge, this is the first use of bulk soil metaproteome data to detect viral peptides for identification of potentially active viruses. The metatranscriptome data was also used to assemble RNA viruses and illuminated a high sequence diversity in the soil RNA virosphere, revealing a high abundance of *Reoviridae* sequences and a highly diverse population of *Leviviridae*. A hallmark of *Leviviridae* is their lytic lifestyle and thus their higher abundance in wet soil suggests that there was greater potential for host lysis under high moisture conditions.

In summary, our combination of deep metagenomic sequencing with a multi-omics evaluation of viral gene expression provided new details about the influence of changing soil moisture, historical or experimental, on the composition, activity and potential ecosystem functions of soil viruses.

References:

1 Nelson, W.C., Anderson, L.N., Wu, R., McDermott, J.E., Bell, S.L., Jumpponen, A., Fansler, S.J., Tyrrell, K.J., Farris, Y., Hofmockel, K.S. and Jansson, J.K.. Terabase metagenome sequencing of grassland soil microbiomes. *Microbiology Resource Announcements* **9** (2020).

2 Wu, R., Davison, M.R., Nelson, W.C., Graham, E.B., Fansler, S.J., Farris, Y., Bell, S.L., Godinez, I., Mcdermott, J.E., Hofmockel, K.S. and Jansson, J.K. DNA viral diversity, abundance and functional potential vary across grassland soils with a range of historical moisture regimes. (2021, in preparation).

3 Wu, R., Davison, M.R., Gao, Y., Nicora, C.D., Mcdermott, J.E., Burnum-Johnson, K.E., Hofmockel, K.S. and Jansson, J.K. Moisture modulates soil reservoirs of active DNA and RNA viruses. (2021, in review)

***Funding statement*:** This work was supported by the Department of Energy (DOE) Office of Biological and Environmental Research (BER) and is a contribution of the Scientific Focus Area "Phenotypic response of the soil microbiome to environmental perturbations". PNNL is operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830. Portions of this work was performed in the W. R. Wiley Environmental Molecular Sciences Laboratory (EMSL) (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory (PNNL). The metaproteomic work was funded by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research (BER), Early Career Research Program to KEBJ.

Title: Volatile Emissions and Cell Wall Ester Signatures of Abiotic Stress in Poplar

Authors: Rebecca A Dewhirst[1]* (radewhirst@lbl.gov), Joseph Lei[1], Cristina Castanha[1], Robert P Young[2], Pubudu Handakumbura[2], Hardeep Mehta[2], Chaevien S Clendinen[2], Yu Gao[3], Miguel Portillo-Estrada[4], John E Mak[5], Luping Su[6], Allen H Goldstein[7], Silvano Fares[8], Jenny C Mortimer[9] and **Kolby J Jardine**[1]

Institutions: [1]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [2] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA; [3]Joint BioEnergy Institute, Emeryville, CA, USA; [4]Department of Biology, University of Antwerp, Wilrijk, Belgium; [5]Stony Brook University, Long Island, New York, USA; [6]Tofwerk USA, Boulder, CO, USA, [7]Department of Environmental Science, Policy and Management, UC Berkeley, Berkeley, CA, USA; [8]National Research Council of Italy, Rome, Italy; and [9]School of Agriculture, Food and Wine, University of Adelaide, Glen Osmond, SA, Australia

Website: http://cellwallesters.pbworks.com/w/page/127623629/FrontPage

Project Goals:

Cell wall polysaccharides can be heavily decorated with methyl and acetyl esters, which can impact fermentation yields of poplar biomass. The hydrolysis of these ester groups results in the production of volatile methanol and acetic acid, which were generally considered waste products of cell wall metabolism but have recently been shown to be tightly coupled to plant growth, stress and senescence processes. However, methanol and acetic acid are not captured by traditional metabolomics analysis and thus represents an important knowledge gap within cell wall metabolism and its interaction with the environment. The 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 goal of this research is to modify the expression of key genes involved in cell wall metabolism in order to modify the amount of methyl and acetyl groups present on cell walls. These genetic modifications will be evaluated for potential impacts on plant 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:
A common thread among many of the biochemical and physiological processes that determine plant responses to climate change variables are alterations in plant cell wall chemical composition, structure, and function. A large proportion of the plant cell wall can be modified with methyl and *O*-acetyl ester groups which may play important roles in cell growth, tissue development, proper xylem and stomatal functioning, central carbon and energy metabolism, and stress communication and signaling. While the hydrolysis of these esters leads to rapid physiological changes in the cell wall and the emission of methanol (meOH) and acetic acid (AA) to the atmosphere, little is known about their changes in response to abiotic stress. Here we provide evidence that drought and high

temperature stress induce a coordinated change in plant cell wall esters composition, central energy metabolism, and leaf–atmosphere fluxes of methanol, acetic acid, carbon dioxide (CO_2), and water (H_2O). $^{13}C_2$-acetate feeding of poplar branches (*Populus trichocarpa*) resulted in emissions of $^{13}CO_2$ from illuminated leaves, suggesting the utilization of free acetate as a respiratory substrate in the light via activation to acetyl-CoA. Moreover, ^{1}H-NMR analysis of leaf cell walls from branches exposed to $^{13}C_2$-acetate suggests that unlike previously assumed, *O*-acetylation of cell wall polysaccharides is reversible, potentially allowing plants to rapidly modify cell wall acetylation patterns. Continuous branch gas exchange observations demonstrated diurnal patterns of methanol and acetic acid emissions that were dominated by methanol in physiologically active control plants (AA/meOH < 10 %). Branch feeding with a 50:50 solution of $^{13}C_2$-acetate:^{13}C-methanol also revealed an AA/meOH ratio dominated by methanol (AA/meOH < 1 %). In contrast, experimental drought treatments resulted in a suppression of methanol emissions and a strong enhancement in acetic acid emissions together with metabolites of the acetate fermentation pathway, acetaldehyde and ethanol. These drought-induced changes in emission patterns lasted > 6 days with their initiation coinciding with a reduction in leaf water potential, stomatal conductance, transpiration, and photosynthesis. The strong enhancement in AA/meOH emission ratios during drought (up to 500 %) was associated with an increase in leaf cell wall *O*-acetylation. Moreover, AA/meOH emission ratios were found to increase with temperature in physiologically active poplar branches and detached leaves and stems. The temperature dependence of AA/meOH emissions ratio was also observed at the ecosystem scale using eddy covariance above a managed poplar plantation in Belgium, a citrus orchard in California, and a diverse forested ecosystem in Alabama. The results are consistent with the activation of the acetate fermentation pathway and acetate photoassimilation as an evolutionarily conserved drought and high temperature survival strategy with important implications for understanding acetate-mediated drought responses to cell wall *O*-acetylation patterns and plant hydraulics, transcription, cellular metabolism, and hormone signaling and its associated changes in carbon cycling and water use from individual plants to whole ecosystems. We suggested that AA/meOH emission ratios could be used as a new quantitative ecosystem sensor to discriminate between plant growth (enriched in meOH) and stress responses including reductions in conductance, transpiration and gross primary productivity (enriched in acetic acid).

References/Publications

1. Dewhirst RA, Afseth CA, Castanha C, Mortimer JC, Jardine KJ (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
2. Dewhirst RA, Mortimer JC, Jardine KJ (2020) Do cell wall esters facilitate forest response to climate? Trends in Plant Science, 25(8), 729-732

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), 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.

Closing the Carbon Cycle: Design, Optimization and Scaling-Up Production of Carbon-Negative Platform Chemicals

Steven D. Brown,[1*] (steve.brown@lanzatech.com), Fungmin (Eric) Liew,[1] Robert Nogle,[1] Tanus Abdalla,[1] Blake J. Rasor,[2] Christina Canter,[1] Rasmus O. Jensen,[1] Lan Wang,[1] Jonathan Strutz,[1] Payal Chirania,[3] Sashini De Tissera[1], Alexander P. Mueller[1], Zhenhua Ruan,[1] Allan Gao,[1] James Daniell,[1] Robert Conrado,[1] Timothy J. Tschaplinski,[3] Richard J. Giannone,[3] Robert L. Hettich,[3] Ashty S. Karim,[3] Séan D. Simpson,[1] Ching Leang,[1] and Michael Köpke[1], **Michael C. Jewett[2]**

[1]LanzaTech, Skokie, IL; [2]Northwestern University, Evanston, IL; [3]Oak Ridge National Laboratory, Oak Ridge, TN.

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 such as acetone, butanol, 3-hydroxybutyrate (3-HB), 1,3-butanediol (1,3-BDO) or monoethylene glycol (MEG).

Acetone and isopropanol are important industrial bulk and platform chemicals, exclusively produced from fossil resources today. We have developed a sustainable and commercially relevant route from abundant, low-cost waste feedstocks—such as industrial waste gases or biomass syngas—by engineering autotrophic acetogen, *Clostridium autoethanogenum*. To achieve this, we constructed and screened a combinatorial biosynthetic pathway library using genes derived from a historical industrial strain collection and enzyme engineering. To optimize flux, we performed strain engineering using omics analysis, kinetic modelling, and cell-free prototyping to identify competing interactions between heterologous enzymes and native metabolism. We developed and scaled up a continuous fermentation process in an industrial pilot plant, consistently demonstrating high selectivities (~90%) and productivities (~3 g/L/h) for extended periods (>3 weeks). Life cycle analysis confirmed significant (>165%) greenhouse gas

savings. We show that acetogens, despite living on the edge of life, can be efficient cell factories for chemicals production.

References

Simpson, S. D. & Köpke, M. Pollution to products: Recycling of 'above ground' carbon by gas fermentation. Curr. Opin. Biotechnol. (2020).

Karim, A. S. et al. In vitro prototyping and rapid optimization of biosynthetic enzymes for cellular design. Nat. Chem. Biol. 16, (2020).

Krüger, A. et al. Development of a clostridia-based cell-free system for prototyping genetic parts and metabolic pathways. Metab. Eng. 2020.03.11.987826 (2020) doi:10.1016/j.ymben.2020.06.004.

We acknowledge the U.S. Department of Energy Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program for funding of this project under Contract No. DE-SC0018249.

Multi-omics analyses reveal temperature-induced metabolic changes in the industrially relevant microbe *Clostridium autoethanogenum* affecting its product profile

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.

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.

Even with the advent of next-generation sequencing, there are limited options for many key metabolic genes considered for next-generation biofuel and bioproduct synthesis, and most genes found in public repositories are derived from type strain or environmental sequences with unproven performance. To expand the pool of available sequences that are likely to result in high performance, we have sequenced and mined the largest collection of industrially deployed Clostridium strains, evolved over several decades of intense development. To rapidly prototype the performance of identified genes and to develop improved, industrial-robust production strains for conversion of lignocellulosic biomass to next generation biofuels and bioproducts, we are establishing a new interdisciplinary venture, the cBioFAB that combines advancements in cell-free and Clostridium engineering metabolic engineering.

The utilization of industrial waste gases or syngas (a mixture of $H_2/CO/CO_2$) is a sustainable alternative for the production of commodity chemicals and biofuels[1]. Syngas is readily fermented to valuable solvents via the Wood-Ljungdahl pathway inherent in the microbial catalyst, *Clostridium autoethanogenum*; an organism that has been deployed commercially for ethanol production. However, for achieving high ethanol productivity and process stability, factors affecting fermentation, such as pH, temperature, media composition, etc., must be understood and controlled. Studies on other *Clostridium* species have reported that both substrate consumption and solvent profiles are significantly impacted by changes in temperature. For example, increases in ethanol production are observed at temperatures lower than an organism's optimum[2]. While studies have determined the optimal growth temperature for *C. autoethanogenum*, few have explored its response outside its optimal range. Hence, studies investigating the effect of temperature on the growth, metabolism, and membrane fluidity of this industrially-relevant microbe are needed to inform future optimization efforts. To that end, *C. autoethanogenum* cultures grown at varying temperatures (30°C and 40°C) were characterized at a molecular level

using a multi-omics approach (metabolomics, proteomics, and lipidomics). Initial results show that while CO uptake was similar at the two temperatures, significant differences in the product profiles were evident. At 40°C, a smaller proportion of the product was ethanol (56% vs. ~73% at 30°C), as more carbon (C) flux was diverted towards acetate (27%). Proteomics analyses revealed that the increased ethanol production could be driven by enzymes considered critical for diverting C flux from acetate to ethanol by reduction of acetate to acetaldehyde (Aldehyde ferredoxin oxidoreductases)[3]. These enzymes were ~4X more abundant in the 30°C cultures. Metabolomics analyses revealed the accumulation of multiple glycerol-conjugated fatty acids at 40°C, suggesting a shift in C flux to fatty acid biosynthesis and glycerolipid metabolism pathways via acetate at higher temperatures. Additionally, differences in the constituent membrane phospholipid species were also observed at the two temperatures, which along with metabolomics results, suggest that adaptive alterations to the membrane fluidity could be occurring. Interestingly, glycerolipids accumulating at 40°C such as 1-myristoyl-glycerol, can be used as food emulsifiers and are commercially relevant. Beyond these, several other proteins (621 of the 1,831 proteins quantified) and metabolites (including some newly identified in this study) were differentially abundant between the two conditions, indicating that temperature substantially impacts the metabolism of *C. autoethanogenum*. 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, enable the generation of a variety of valuable products.

References

1. Fackler, N., Heijstra, B. D., Rasor, B. J., et al. (2021). Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. Annual review of chemical and biomolecular engineering, 12, Accepted.
2. Kundiyana, D. K., Wilkins, M. R., Maddipati, P., & Huhnke, R. L. (2011). Effect of temperature, pH and buffer presence on ethanol production from synthesis gas by "Clostridium ragsdalei". Bioresource technology, 102(10), 5794–5799. https://doi.org/10.1016/j.biortech.2011.02.032
3. Liew, F., Henstra, A. M., Köpke, M., Winzer, K., Simpson, S. D., & Minton, N. P. (2017). Metabolic engineering of Clostridium autoethanogenum for selective alcohol production. Metabolic engineering, 40, 104–114. https://doi.org/10.1016/j.ymben.2017.01.007

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 of Novel Bioproduct Pathways in an Acetogen through Integrated Computational Modeling and High-throughput Candidate Screening

Zach Cowden,[1*] (zach.cowden@lanzatech.com), Rasmus O. Jensen,[1] Joseph Ni,[2] Samuel Brown,[2] Blake J. Rasor,[2] Ching Leang,[1] Sean D. Simpson,[1] and Michael Köpke,[1] **Michael C. Jewett[2]**

[1]LanzaTech, Skokie, IL; [2]Northwestern University

Project Goals: The Clostridia Foundry for Biosystems Design (cBioFAB) pulls together new capabilities in computational modeling, cell-free metabolic engineering (CFME), genetic engineering of *Clostridium*, and laboratory automation to streamline the development of industrial bioprocesses for next-generation bioproducts from gasified carbon substrates. In one aim of this project, we set out to develop a comprehensive set of enzymatic parts using methods for automated identification pulling from a wide net of databases and genome-scale models. A tailored set of enzyme parts is then applied towards the generation of novel metabolic pathways to targeted bioproducts. High-throughput assembly and characterization of predicted pathways allows for screening data to be fed back into the computational framework to optimize and predict improved novel pathways.

Mono-ethylene glycol (MEG) is a chemical used predominantly as a building block in the synthesis of polyethylene terephthalate (PET), a polymer with major applications in the plastics and textiles industries. MEG derived from gas fermentation would have several advantages over traditional chemical synthesisincluding reduced carbon emissions and the ability to utilize low-cost waste carbon feedstocks. Currently the only known biological route to MEG occurs through the metabolism of C5 sugars, which is less attractive from the perspective of carbon recycling and feedstock cost, and has not scaled commercially. *Clostridium autoethanogenum*, an acetogen used in the industrial production of chemicals, is being investigated as an engineered host strain for MEG production from synthesis gas (1). To accomplish this, we are employing CFME alongside high-throughput *in-vivo* screening to rapidly prototype novel pathways to MEG that were computationally predicted by the BNICE framework (2). Enzymes are produced via cell-free protein synthesis (CFPS) in *E. coli* cell extracts and then combined *in vitro* to recapitulate predicted routes to MEG. High-performing candidates are assembled in large combinatorial plasmid libraries with varying promoter strengths and transformed into *C. autoethanogenum* to be screened on a high-throughput automated biofoundry platform (3). This approach allows us to rapidly test, optimize, and rank novel predicted pathways. Using these methods, we were able to generate strains demonstrating production of MEG in small-scale gas fermentation.

References/Publications

1. Köpke, M., Simpson, S.D. Pollution to products: recycling of 'above ground' carbon by gas fermentation. Curr Opin Biotechnol 65,180-189. https://doi.org/10.1016/j.copbio.2020.02.017

2. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ. Exploring the diversity of complex metabolic networks. Bioinformatics. 2005 Apr 15;21(8):1603-9. doi: 10.1093/bioinformatics/bti213. Epub 2004 Dec 21. PMID: 15613400.
3. Karim, A.S., Liew, F., Garg, S. et al. (2020). Modular cell-free expression plasmids to accelerate biological design in cells. Synthetic Biology 5(1): ysaa019.

We acknowledge the U.S. Department of Energy Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program for funding of this project under Contract No. DE-SC0018249.

Modeling Tools to Predict Metabolic Network Behavior in Non-Model Systems

Jacob Martin* (jacobmartin2022@u.northwestern.edu), Kevin Shebek* (kevinshebek2016@u.northwestern.edu), Joseph Ni, Blake Rasor, Ashty Karim, Linda Broadbelt, **Keith Tyo**

Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL

tyolab.northwestern.edu; broadbelt.northwestern.edu

Project Goals: We aim to develop computational tools to aid in the design and understanding of metabolic pathways for the production of novel compounds. Our work utilizes promiscuous reaction rules to enumerate candidate pathways and filters these pathways based on network topology, thermodynamics, and theoretical yield. Once candidate pathways are prototyped in cell-free systems, we utilize kinetic modeling to determine rate-limiting steps in the pathway along with system-level interactions which would not be detected through steady-state models alone.

Non-model organisms present many opportunities for the production of small molecules due to a wide variety of metabolic capabilities, including the anaerobic fermentation of waste gas streams or plant-based byproducts to valuable chemicals. However, these non-model microbes are slow growing, less well-characterized, and more difficult to genetically modify than model organisms, making it difficult to quickly test potential pathways. Cell-free systems provide a platform to overcome these challenges by allowing rapid prototyping of novel pathways in the context of a cell's native metabolism. Despite this ability to rapidly test pathways, there is still a need to efficiently design, filter, and interpret the immense number of potential pathways. To accomplish this task, we are focused on two areas: novel pathway design, and kinetic modeling of cell-free metabolism. The successful development of these tools will aid in the utilization of cell-free prototyping to efficiently design and implement pathways into non-model organisms.

To identify promising novel pathways, we are developing a computational workflow that generates, analyzes, and filters pathways from specified starting compounds to a set of desired target compounds. Metabolic networks are generated by using Pickaxe, a reaction network generation software that utilizes promiscuous reaction rules to transform compounds, allowing for the generation of pathways consisting of both known and novel reactions. These pathways are then analyzed to score important features such as their yield determined by flux balance analysis, number of known enzymes, pathway length, and thermodynamics. Thermodynamic feasibility is calculated by determining the max-min driving force using eQuilibrator, which has been modified to accept arbitrary compounds for inputs. These pathway features are then used to determine overall pathway feasibility and potential performance to rank the pathways into a set of the most promising candidates for experimental consideration.

Once a pathway has been designed and is prototyped in cell-free systems, there remains a need for methods to uncover the kinetics of this system. To this end, we are developing a framework to construct dynamic kinetic models to understand how the pathway of interest interacts with native metabolism. While past kinetic models have been parameterized from literature kinetic

values or extensive fluxomics datasets, the modeling of these cell-free systems do not have these data due to limited characterization of non-model organisms and the lack of detailed multi-omics from rapid prototyping. Therefore, this method first uses a limited set of metabolomics, proteomics, and thermodynamics to constrain a system of linear equations, which guide and constrain the sampling of underlying fluxes and metabolic states. We subsequently apply the metabolic ensemble modeling (MEM) framework to sample and prune best-fit parameters. By modeling this system with a dynamic metabolic model, we capture complex time-dependent interactions between the novel pathway and native metabolism, providing detailed understanding of these connections and allowing for better control and optimization of the pathway of interest. This work will ultimately allow for both deeper understanding of cell-free metabolism as well as improved engineered pathway titers in both *in vitro* systems and translated to *in vivo* production strains.

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.

Spacer2PAM: an R Package for CRISPR Protospacer Adjacent Motif Prediction from Spacer Sequences

Grant A. Rybnicky[1]* (grantrybnicky2023@u.northwestern.edu), Michael Köpke[2], and **Michael C. Jewett**[3]

[1]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL; [2]LanzaTech, Skokie, IL; [3]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL

http://jewettlab.northwestern.edu/; https://www.lanzatech.com/

Project Goals: We are addressing the challenge of designing, building, and optimizing biosynthetic pathways in cells in an interdisciplinary venture that establishes the clostridia Foundry for Biosystems Design (cBioFAB). Working both *in vitro* and *in vivo*, the goal is to interweave and advance state-of-the-art computational modeling, genome editing, omics measurements, systems-biology analyses, and cell-free technologies to expand the set of platform organisms that meet DOE bioenergy goals. cBioFAB will (i) reconceive how we engineer complex biological systems by linking pathway design, prospecting, validation, and production in an integrated framework, (ii) enable systems-level analysis of the David T. Jones collection, one of the largest collections of clostridia strains in the world, to uncover novel metabolic pathways, regulatory networks, and genome editing machinery, and (iii) open new paths for synthesis of next-generation biofuels and bioproducts from lignocellulosic biomass.

Advancement in CRISPR-based genome editing tools have greatly accelerated the ability to modify and domesticate a variety of microorganisms, but some bacteria remain recalcitrant to these technologies. Heterologous expression of CRISPR nucleases often cause cellular toxicity by unclear mechanisms in many bacteria, limiting the ability to edit and modify these organisms. However, endogenously encoded CRISPR-Cas systems are broadly present among bacteria and archaea. Harnessing these endogenous systems presents away to circumvent the drawbacks associated with heterologous nuclease expression and enables access organisms that were previously inaccessible. One hurdle to using endogenous systems is identification of functional protospacer adjacent motif (PAM) sequences, a requirement for nuclease targeting. Most current experimental methods determine PAM sequences via selection of a pooled randomized PAM library in the presence of a CRISPR system and use of next generation sequencing data as a readout of PAM frequencies. However, this technique is not feasible in organisms with very low transformation efficiencies and Cas proteins that are difficult to recombinantly express. Computational approaches attempt to address this same problem by reversing the process of spacer acquisition *in silico*. Since spacers in CRISPR arrays originated in organisms encoding a PAM next to the spacer sequence, sequence alignment can be used to link spacers to potential PAMs. A few tools have used this approach, but efforts thus far have not been continuous from spacer to PAM prediction or do not allow user input of curated data. Here we present Spacer2PAM, an R package that allows users to predict PAM sequences and design targeted PAM libraries from user-provided CRISPR array spacer sequences. The

package includes functions to standardize and manipulate sequence and alignment data as well as predict and visualize PAM sequences from those data. We identify two regimes in which the tool is effective; a quick method to predict a PAM likely to be functional and a comprehensive method to design targeted PAM libraries. The quick method was able to identify functional PAM sequences for 8 out of 10 model CRISPR systems tested, with the other 2 systems yielding partial predictions. The comprehensive method was able to inform targeted library design that would achieve a functional PAM in 16 transformations or fewer for 9 out of the 10 model CRISPR systems. Using this tool, we apply the quick and comprehensive methods to organisms with unique carbon metabolism and suggest functional PAMs as well as targeted PAM libraries. We anticipate this consolidated and improved computational pipeline will enable faster domestication of endogenous and novel CRISPR systems, especially in organisms that have poor transformation efficiencies.

We acknowledge the Department of Energy grant DE-SC0018249 for funding of this project.

From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms

Dan Park*[1](park36@llnl.gov), Mimi Yung,[1] Jonathan Allen,[1] Jose Marti Martinez,[1] Jeffrey Kimbrel,[1] Ali Navid,[1] Adam Zemla,[1] Matthew Coleman,[1] Jennifer Pett-Ridge,[1] Harris Wang,[2] Tomasz Blazejewski,[2] Guillaume Urtecho,[2] Carlotta Ronda,[2] William Bentley,[3] Gregory Payne,[3] Kayla Chun,[3] Zhiling Zhao,[3] Jeffrey Gralnick,[4] Jennifer Listgarten,[5] **Yongqin Jiao**[1]

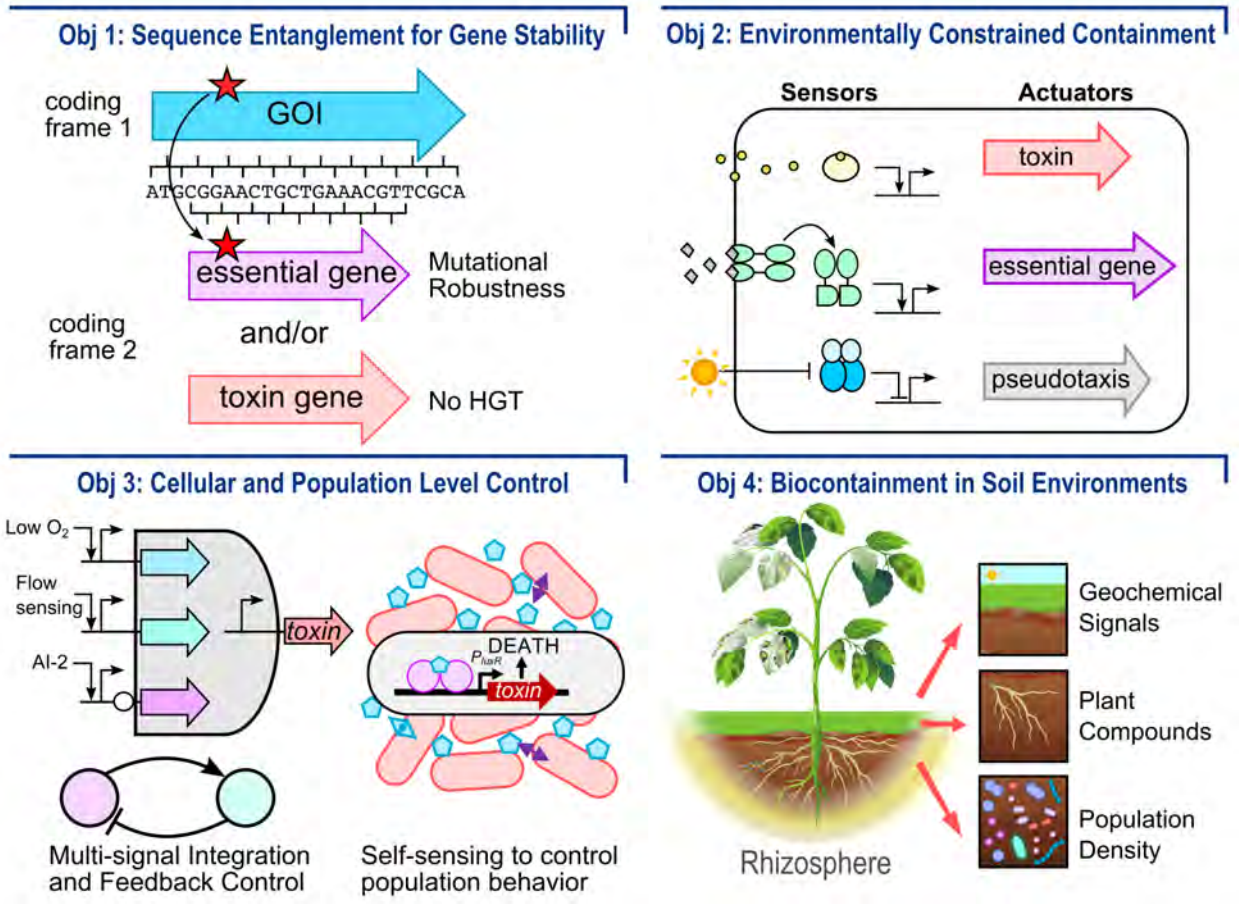
[1]Lawrence Livermore National Laboratory, [2]Columbia University, [3]University of Maryland, [4]University of Minnesota, [5]University of California, Berkeley

https://genomicscience.energy.gov/research/sfas/llnlseqcellpop.shtml

Project Goals:

Establish robust, generalizable biocontainment strategies in environmentally relevant soil microbes at the sequence, cellular, and population levels.

1) Establish sequence entanglement as a generalizable strategy to improve genetic stability and to prevent horizontal gene transfer
2) Build robust containment circuits that are resistant to mutation and responsive to *in situ* environmental cues
3) Implement cellular- and population-level containment mechanisms that are robust to environmental perturbation
4) Gain a system-level understanding of the genomic and phenotypic response of secure engineered organisms in soil and rhizosphere environments



Abstract: Genetically engineered microorganisms (GEMs) hold significant promise for establishing a sustainable bioeconomy. However, the lack of robust and generalizable biocontainment strategies hinders technology adoption and public trust. To reduce the risk of unintended ecological consequences from environmentally deployed GEMs, built-in security mechanisms are needed to ensure that GEMs function where and when needed without proliferating beyond target conditions. Here, we develop robust, generalizable biocontainment strategies in environmentally relevant soil microbes at the sequence, cellular, and population levels. Leveraging Lawrence Livermore National Laboratory's high-performance computing (HPC) and high-throughput gene-editing capabilities, we are advancing a synthetic gene entanglement concept for containment. Here, two genes are encoded within different coding frames of the same sequence space to protect engineered functions against mutational inactivation and/or to mitigate the horizontal transfer of potentially invasive genes. Building on this layer of sequence stability, we design sense-and-respond circuits that constrain the survival and function of plant-benefiting microorganisms to their target application environments. Control strategies, such as sequestration-based feedback control, multi-signal integration, pseudotaxis, and quorum sensing-based population coordination will be incorporated to increase the overall system robustness to environmental fluctuations that are expected in environmental applications. We leverage LLNL's rich experience in soil microbial ecology to evaluate the ecological effects of these containment mechanisms in soil and rhizosphere environments. Ultimately, our results will yield safeguard mechanisms that control the niche-specific function of GEMs and prevent the transfer of potentially "invasive traits" to neighboring native microorganisms, thereby enabling safer and more effective use of GEMs in environmental applications.

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-815852).

An Electrochemical Model of Rhizosphere ROS Generation to Analyze Commensal Engineering Strategies

Eric VanArsdale[1,2,3]* (Presenter email: esv5009@terpmail.umd.edu), Julianna Pitzer[1], Sally Wang[1,2,3], Gregory Payne[2,3], William Bentley[1,2,3], **Yongqin Jiao**[4]

[1]Fischell Department of Bioengineering, University of Maryland, College Park, MD; [2]Institute for Bioscience and Biotechnology Research, University of Maryland, College Park, MD; [3]Fischell Institute for Biomedical Devices, University of Maryland, College Park, MD; [4]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratories, Livermore, CA

Project website: https://genomicscience.energy.gov/research/sfas/llnlseqcellpop.shtml

Project goals: Our goal is to develop robust and generalizable cellular and population level containment mechanisms for soil microbes that are robust to environmental perturbation

A common source of plant dysbiosis caused by pathogenic signaling within the rhizosphere is overproduction of reactive oxygen species (ROS). This appears to be in part due to unregulated or pathogenic redox signaling between local microbial species and the root of the plant. One solution to this problem is to develop commensal microbial species that sense root ROS production and in turn alter the composition of the microbiome. In pursuit of this goal, we have adapted electrochemical methods to study the response of a community of bacteria to localized hydrogen peroxide stimulation as a model of root ROS production. Within these experiments, a standard gold electrode is used to synthesize hydrogen peroxide at the electrode surface through a 2-electron oxygen reduction reaction (ORR). The hydrogen peroxide is then rapidly degraded by bacterial mechanisms with distance away from the electrode. We found that within a simple electrochemical cell, only a marginal number of bacteria are stimulated by the hydrogen peroxide produced by the electrode due to spatial limitations. However, when using a "transmitter/receiver" coculture of *E. coli,* we found that the transmitter population, which produces a quorum sensing signal in response to hydrogen peroxide, could activate gene circuits within a large and homogenous population of receiver cells (see Figure 1). Using this system, we have demonstrated two methods that could be used to alter microbiome composition in response to ROS while simultaneously containing the engineered populations. First, we demonstrate how coculture signaling can be used to upregulate tyrosine synthesis, which can be used to selectively enrich auxotrophic populations. Second, we demonstrate the ability to cause the receiver population to increase its growth rate, thereby increasing its overall composition within the microbial community. Lastly, we show that these two methodologies can be combined to enrich the population of cells producing an auxotrophic nutrient. We believe these electrochemical methodologies can be useful for exploring strategies to alter rhizosphere microbial composition while also preventing the spread of engineered species.

Funding is provided by DOE/BER Genomic Science Program through Lawrence Livermore National Laboratory SFA (SCW1710), and by the NSF (ECCS 1807604, CBET 1805274, DMREF 1435957) and Semiconductor Research Corporation (SRC 2483.001) through the University of Maryland

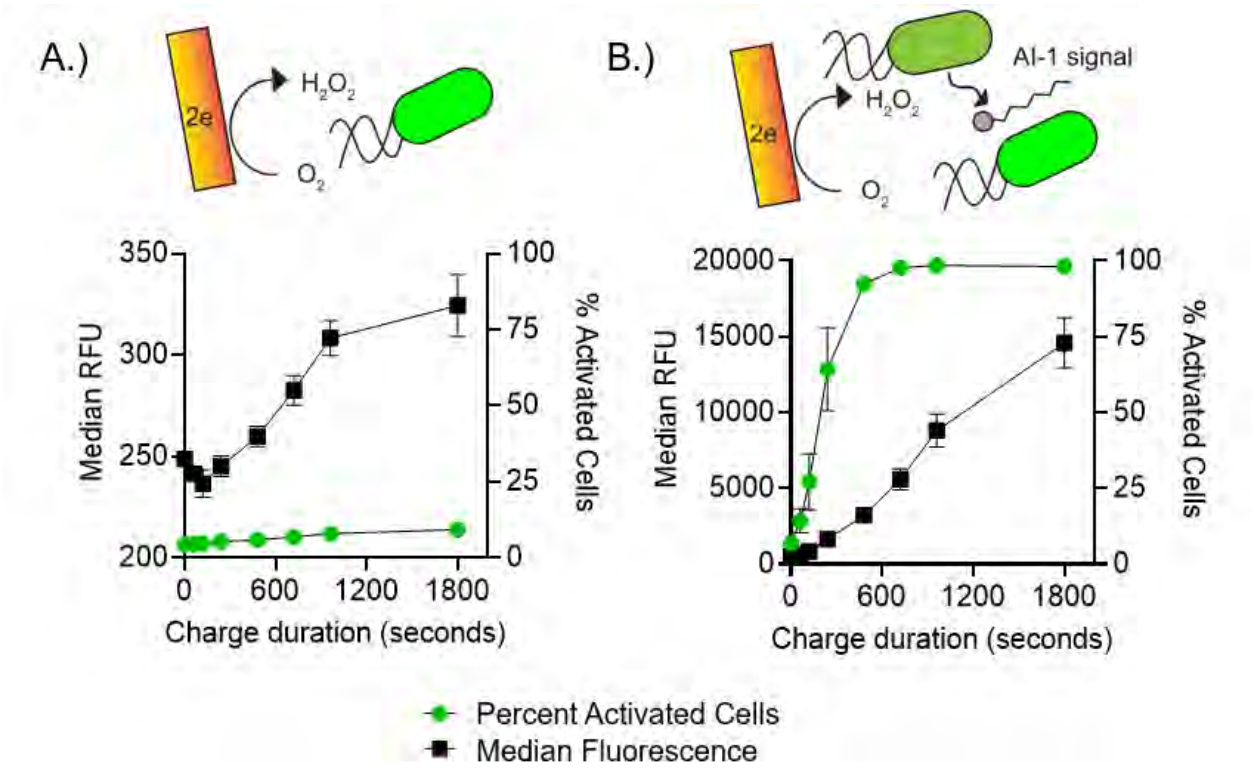


Figure 1: Comparison of Monoculture and Coculture Activation via Localized Hydrogen Peroxide Generation. In this model system we tested two separate models of root microbe activation from ROS: First, a monoculture model in which an engineered population expresses GFP in response to hydrogen peroxide (A); Second, a "transmitter/receiver" coculture model in which a small fraction of the culture is comprised of "transmitter" cells which produce a quorum sensing autoinducer-1 (AI-1) signal in response to hydrogen peroxide. The AI-1 signal in turn activates GFP expression in the second population, termed "receiver" cells (B). In each, the charge duration is the amount of time the electrode was biased to produce hydrogen peroxide. We also defined cell activation as the percent of the analyzed population that had fluorescence exceeding two standard deviations away from a null control. A.) We found that while increasing the charge duration did increase the median fluorescence of the monoculture as measured by flow cytometry, only a small portion of the population was activated. B.) Within the coculture model, however, the percent of activated cells rose rapidly with increasing charge duration indicating the quorum sensing molecule homogenized the response of the entire system.

Transforming our understanding of chloroplast-associated genes through comprehensive characterization of protein localizations and protein-protein interactions

Lianyong Wang*[1](lianyong@princeton.edu), Bibin Paulose[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 1,056 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 are currently analyzing the localization data and identifying protein-protein interactions by affinity purification-mass spectrometry of tagged proteins. 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.

Environmental and Genetic Effects on Switchgrass (*Panicum virgatum*) Biomass Composition across Diverse Environments

Mahbobeh Lesani,[1*] Li Zhang,[2] Jason Bonnette,[2] Christina Helseth,[2] David Thomas,[1] John Sanley,[2] Acer VanWallendael,[3] Steve Masterson,[4] Rob Mitchell,[4] Claire M. Curry,[1] Felix Fritschi,[5] David Lowry,[3] Edward J. Wolfrum,[6] Thomas Juenger,[2] and **Laura E. Bartley[7]*** (laura.bartley@wsu.edu)

[1] University of Oklahoma, Norman, OK, USA; [2] University of Texas, Austin, TX, USA; [3] Michigan State University, East Lansing, MI, USA; [4] USDA-ARS, Lincoln, NE, USA; [5] University of Missouri, Colombia, MO, USA; [6] National Renewable Energy Laboratory, Golden, CO, USA; [7] Washington State University, Pullman, WA, USA

Project Goals:

- **The long-term objective of this project is to develop switchgrass with consistently high cell wall composition across diverse growing environments to promote efficient and sustainable biorefining.**
- **Here, we aim to:**
    - **Understand how growth environment effects switchgrass composition**
    - **Identify switchgrass genomic loci that influence composition and that are or are not dependent on the environment**

Abstract: Biorefining of lignocellulosic material can reduce dependence on fossil sources of fuels and chemicals, thereby mitigating global climate change. Because lignocellulosic compositional traits influence efficiency of biorefining the long-term goal of this research is to identify genes for optimizing switchgrass lignocellulosic composition across diverse growth environments. Here we present the composition and quantitative trait loci for a switchgrass mapping population grown at 10 sites spanning 14 degrees of latitude in the USA. The mapping population was constructed from a cross of four grandparents, two from the upland ecotype and two from the lowland ecotype. Lignocellulosic composition (e.g., lignin, ash, nitrogen, in vitro dry matter digestibility, etc.) was predicted from near-infrared reflectance spectroscopy for ~11K samples from mid- and end-of-season 2016 harvests. Growth location affected compositional traits significantly, but not with a simple latitudinal pattern. Heritability (h^2) of composition traits was <0.5 depending on the location, with higher heritability typically at the end-of season harvest. For a subset of traits, we identified 119 quantitative trait loci (QTL), 21 of which overlapped between the mid and end-of-season, and several of which controlled multiple traits. About 40% of QTL did not show genotype by environment interactions and thus are of major interest for genes that can improve composition across diverse environments. Lowland alleles increased lignin and ash per g dry matter more than upland alleles; whereas, upland alleles increased digestibility and nitrogen. Integration of these results with gene expression and other population genetic analyses will help to identify causative

alleles for improving biorefining and may reveal more about the agronomic consequences of variation in lignocellulose towards mitigating these effects.

Funding Statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021126.

Title: Genetic signatures influencing host-microbiome interactions in switchgrass (*Panicum virgatum*)

Authors: Joseph Edwards*[1] (j_edwards@utexas.edu), Usha Saran[1], Jason Bonnette[1], Alice MacQueen[1] , Jane Grimwood[2], Jeremy Schmutz[2] , Chris Daum[3], Tijana Glavina del Rio[3] , and **Thomas Juenger[1]**

Institutions: [1]University of Texas, Austin; [2]HudsonAlpha Genome Sequencing Center, Huntsville, AL; [3]The Joint Genome Center, Berkeley, CA

Project goals:

- Uncover the role of host genotype in the assembly of plant microbiota
- Use isolated bacteria to test predictions
- Manipulate microbiota to increase sustainability

Abstract:

Plants make partnerships with soil bacteria via their roots, entering into relationships with microbiota ranging from detrimental to commensal to beneficial. While it is known that the composition of root-associated microbiota is driven by environmental factors and the genotype of the host plant (Singer et al. 2019; Edwards et al. 2018), the underlying genetic mechanisms used by the host to modulate bacterial members of the microbiota are unresolved. In this study, we use an expansive population of switchgrass natural variants (Lovell et al. 2021) planted across several common gardens across the United States to explore the genetic architecture of host-microbiome interactions. We find that genetic control of microbiota is a complex trait, i.e. many loci contribute to the abundance of different root-associated bacteria. We find a particular enrichment of significantly associated SNPs near genes involved with cell wall composition and defense response. We next used high-throughput bacterial culturing to isolate bacterial members of the switchgrass root microbiome and test for their effects on root growth when in mono-association. We find that some isolated members of the microbiome with significant associations to host genotype have large effects on root growth dynamics. Together, these efforts contribute important information to the host-microbe interaction field as well as provide interesting avenues for increased sustainability through manipulation of host genetics and microbial amendments.

References:

Edwards, Joseph A., Christian M. Santos-Medellín, Zachary S. Liechty, Bao Nguyen, Eugene Lurie, Shane Eason, Gregory Phillips, and Venkatesan Sundaresan. 2018. “Compositional Shifts in Root-Associated Bacterial and Archaeal Microbiota Track the Plant Life Cycle in Field-Grown Rice.” *PLoS Biology* 16 (2): e2003862.

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*, January. https://doi.org/10.1038/s41586-020-03127-1.

Singer, Esther, Jason Bonnette, Shawn C. Kenaley, Tanja Woyke, and Thomas E. Juenger. 2019. “Plant Compartment and Genetic Variation Drive Microbiome Composition in

Switchgrass Roots.” *Environmental Microbiology Reports* 11 (2): 185–95.

Funding statement: This research was supported by the DOE Office ofScience, Office of Biological and Environmental Research (BER), grant no. DE-SC0021126

Title: The Switchgrass Genome: Polyploidy and Introgressions Facilitate Climate Adaptation and Biomass Yield

Authors: John Lovell[1*] (jlovell@hudsonalpha.org), Alice MacQueen[2], Sujan Mamidi[1], Jason Bonnette[2], Jerry Jenkins[1], Joseph Napier[2], David Lowry[3,4], Jane Grimwood[1], Jeremy Schmutz[1,5], and **Thomas Juenger**[2]

Institutions: [1]Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; [2]Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA; [3]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, USA; [4]Department of Plant Biology, Michigan State University, East Lansing, MI, USA; [5]Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Project Goals: We built and analyzed the switchgrass genome to facilitate faster and more effective breeding for bioenergy feedstock production. In particular, we sought to: 1) define the quantitative genetic structure and molecular gene pools that can be targeted by traditional breeding, 2) find targets for genomic or marker-assisted selection through genetic mapping of climate adaptation, and 3) determine the roles that introgressions, polyploidy and other complex processes play in the evolution of switchgrass.

Abstract text: As climate and natural environments change in exceptional ways, it is increasingly critical to understand and make predictions about the fate of natural populations and productivity of agricultural systems. Plant genomes offer one mechanism to achieve this goal by presenting glimpses into the past and future of crop and wild populations. For example, historical climate variation (e.g. glacial-interglacial cycles) is a key analog for current and future environmental change, one that we explore here to dissect the genomic mechanisms of adaptation and yield improvement in the polyploid biofuel crop, switchgrass.

References/Publications

1. Lovell, J.T., MacQueen, A.H., Mamidi, S. *et al.* Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* (2021). https://doi.org/10.1038/s41586-020-03127-1

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DES-C0014156. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No DE-AC02-05CH11231.

Community Resources to Study Switchgrass Adaptation Using Genome-Wide Association

Alice MacQueen[1]* (alice.macqueen@utexas.edu), Li Zhang[1], Jason Bonette[1], John Lovell[2], Jeremy Schmutz[2], **Thomas Juenger**[1]

[1]The University of Texas at Austin; [2]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama

https://alice-macqueen.github.io/switchgrassGWAS/

Project Goals:

- **Establish software and genomic resources supporting study of common gardens of clonally replicated switchgrass genotypes across broad environmental gradients (e.g., precipitation, temperature, and soils).**
- **Identify genomic regions underlying adaptation and sustainability in switchgrass using genomewide associations.**
- **Investigate key switchgrass traits (resource use efficiency, drought tolerance, growing season phenology, freezing tolerance, tissue characteristics, and root system attributes) in climate adaptation and sustainability of switchgrass feedstock production.**

As sessile organisms, plants cannot move to escape unpredictable and changing environments. Which environments impact plants the most? How do plant genetic responses to the environment vary, and how do these responses evolve? One common hypothesis is that adaptation to specific environments, or local adaptation, occurs via tradeoffs involved in specialization: alleles with antagonistic pleiotropy increase fitness in specific environments, but have negative, pleiotropic effects in alternate environments. A contrasting hypothesis at the level of the allele is conditional neutrality, where alleles can increase fitness in specific environments without costs in alternative environments. As climates shift and climate variability increases, access to conditionally neutral alleles that improve fitness in specific stressful environments will be essential for improving crop species. Genomics-enabled research is now providing the statistical power to discover and characterize allelic variation in genes involved in adaptation.

Switchgrass (*Panicum virgatum*) is an outcrossing, polyploid C4 perennial grass that has been championed as a promising biofuel feedstock. It is a common member of most native North American prairie communities and exhibits extensive phenotypic variability and adaptation across its range, particularly in response to latitude and precipitation gradients. Here, I report on the development of genome-wide association resources for a diversity panel of switchgrass. This diversity panel includes over 700 sequenced genotypes sampled from the majority of the range of switchgrass across the eastern United States. Clones of the sequenced individuals were planted at ten field sites covering 17° of latitude (1800 km) in the central United States. Phenotyping at these common garden sites allowed us to evaluate the contributions of individual loci to traits and fitness over a wide range of climatic conditions (Lovell et al. 2021). In this poster, I present genetic analyses of phenology data collected during the 2019 growing season. In particular, I

present results from multivariate analyses of genome-wide association studies aimed at detecting genomic regions that both affect greenup or flowering in ways consistent with specific environmental cues, such as temperature and daylength. I also present some of the resources I have developed to conduct genetic and genomic studies on the switchgrass diversity panel at these common gardens, including the R software package, switchgrassGWAS.

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*, January, 1–7. https://doi.org/10.1038/s41586-020-03127-1.
2. Alice H. MacQueen. "The switchgrassGWAS R package." https://alice-macqueen.github.io/switchgrassGWAS/

This research was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research award number DE-SC0014156 and DE-SC0021126 to TEJ.

Integrated *Pseudomonas putida* Strain Design For Maximizing Biomass Conversion To Biofuels And Bioproducts.

Deepanwita Banerjee[1,2*](dbanerjee@lbl.gov), Deepika Awasthi[1,2*](dawasthi@lbl.gov), Hyun Gyu Lim[1,3], Manuel Rafael Jimenez Diaz[1,2], Thomas Eng[1,2], Blake A. Simmons[1,2], Steven W. Singer[1,2], Adam Feist[1,3,4], Aindrila Mukhopadhyay[1,2] and **Jay D. Keasling**[1,2,4,5]

[1]Joint BioEnergy Institute, 5885 Hollis street, 4th floor, Emeryville, CA 94608, USA; [2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94702, USA; [3]Department of Bioengineering, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA; [4]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark and [5]Center for Synthetic Biochemistry, Institute for Synthetic Biology, Shenzhen Institutes for Advanced Technologies, Shenzhen, China

http://jbei.org

Project Goals: Engineer novel catabolic routes to redirect carbon flow, use of functional genomics and predictive tools for systematic improvements in titer, rate, and yield (TRY) and use of adaptive laboratory evolution (ALE) to enhance host phenotype.

Plant biomass (lignocellulose) is an underutilized carbonaceous feedstock in the industrial microbial bioconversions. Though polymers of glucose, xylose and aromatic hydrocarbons form lignocellulose, only glucose bioconversion to products is the commercially established bioprocess. In the presented work we show an integrated host engineering strategy that can achieve maximum lignocellulosic carbon bioconversion to heterologous drop-in biofuels and bioproducts by engineering *Pseudomonas putida*, a non-pathogenic soil microbe. Our approach integrates A) targeted *P. putida* strain engineering to incorporate heterologous non-oxidative glycolysis to limit CO2 loss and reach theoretical maximum product yields from glucose, incorporate heterologous xylose catabolism capability and express a multi-gene heterologous pathway for methyl ketones (diesel substitute) production; B) adaptive laboratory evolution for improved growth on xylose and galactose as well as tolerance to lignin-derived aromatics and ionic liquids; and C) high throughput fitness profiling and genome-scale metabolic modeling (GSMM) to identify experimentally implementable gene deletion targets for enhanced productivity and fitness under bioreactor conditions. These host developments can help improve the titers, rates and yields (TRY) of bioproducts under industrial bioprocess parameters as modeled with a sustainable pigment, indigoidine, or in future studies with advanced biofuel candidates, methyl ketones and isopentenol.

Publications

1. Banerjee, D., Eng, T., Lau, A. K., Sasaki, Y., Wang, B., Chen, Y., … Mukhopadhyay, A. (2020). Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale. *Nature Communications*, *11*(1), 5385.
2. Lim, H. G., Fong, B., Alarcon, G., Magurudeniya, H. D., Eng, T., Szubin, R., … Feist, A. M. (2020). Generation of ionic liquid tolerant *Pseudomonas putida* KT2440 strains via adaptive laboratory evolution. *Green Chemistry*, 22(17), 5677–5690.

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.

Production Cost and Carbon Footprint of High Performance Biomass-Derived Dimethylcyclooctane as a Jet fuel Blendstock

Nawa Raj Baral,[1,2] Corinne D. Scown[1,2*] (cdscown@lbl.gov), Minliang Yang, [1,2] Benjamin G. Harvey,[3] Blake A. Simmons[1,2] and **Jay D Keasling**[1,2]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, California; [2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California;[3]Research Department, Chemistry Division, United States Navy, Naval Air Warfare Center Weapons Division (NAWCWD), China Lake, California

www.jbei.org

Project Goals: Predict the impact of JBEI research results on biofuel selling price and carbon efficiency.

Biomass-derived 1,4-Dimethylcyclooctane (DMCO) is a performance-advantaged jet fuel that can reduce greenhouse gas emissions relative to conventional Jet-A and extend the range and efficiency of aircraft with its 9.2% higher energy-density and improving combustion properties.[1] This study provides the first technoeconomic analysis and life-cycle greenhouse gas assessment for DMCO product using three different hydrogenation catalysts and two different bioconversion pathways. With current technologies and yields, the platinum-based catalyst results in the lowest production cost and carbon footprint of DMCO of $8.4/L-Jet A and 72.3 gCO2e/MJ, respectively. However, when the conversion process is fully optimized, hydrogenation with the Raney Ni-based catalyst results in the lowest cost and carbon footprint of DMCO of $1.4/L-Jet A and 19.8 gCO2e/MJ, respectively. The optimized scenario can be thought of as a theoretical floor for cost and emissions, with biomass sorghum yields exceeding >25 metric-ton/ha and near-theoretical conversion rates of sugar-to-isopentenol, isopentenol-isoprene, and isoprene-to-DMCO. While still more costly to produce that Jet-A, DMCO offers performance advantages that may facilitate early adoption in markets that place a high value energy density and other fuel properties, offering an opportunity to further optimize production through early demonstrations and scale-up and bring down the minimum selling price.

References

1. Rosenkoetter, K. E.; Kennedy, C. R.; Chirik, P. J.; Harvey, B. G. [4 + 4]-cycloaddition of isoprene for the production of high-performance bio-based jet fuel. Green Chem. 2019, 21, 5616–5623.

Funding statement.

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.

Title: An semi-automated workflow for high-throughput and quantitative proteomic analysis of metabolic engineered microorganisms

Yan Chen[1]* (yanchen1998@lbl.gov), Jennifer Gin[1], Tadeusz L. Ogorzalek[1], Nurgul Kaplan Lease[1], Paul D. Adams[1], Christopher J. Petzold[1], **Jay D. Keasling[1]**

[1] Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA, USA;

https://www.jbei.org/

Project Goals:

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.

Abstract:

Production of advanced biofuels and valuable chemicals through microbial fermentation contributes to more sustainable energy and chemical resources. Mass spectrometry (MS)–based proteomic analysis plays an important role in understanding the function of engineered microbes to aid metabolic pathway optimization through multiple DBTL cycles. As a result, there is increasing interest in large amounts of high quality proteomic data to create new and meaningful biological knowledge. Our previous efforts have greatly advanced omics sample throughput and overall data quality by establishing a rapid targeted proteomic workflow[1] and a complete automated proteomic sample preparation platform[2], yet challenges remain in automation robustness and increasing analytical sample throughput. To tackle these challenges, we established a semi-automated workflow that combines robust modular automated sample preparation protocols and fast data independent acquisition to enable high throughput and quantitative proteomics analysis of metabolically engineered microorganisms. This strategy decreased the complexity of an existing fully-automated workflow[2], reduced variance in the sample preparation process, and circumvent bottlenecks related to automation system availability. The modular automated workflow also reduced the total sample preparation time by over 50% and simplified methods development for new organism compared to the full automated sample preparation platform. The parameters in these modular automated methods are easily adjustable so that they could be transferred to labs with standard automated liquid handlers. Concurrently, we developed high-throughput shotgun proteomic data acquisition methods that achieve routine detection and quantification of over 800 proteins within 15 minutes per sample analytical time, thus providing throughput of ~100 samples/day. We are also building deep proteome libraries of bioenergy relevant organisms containing over 1,000 native proteins and ever increasing number of heterogenous pathway proteins. Preliminary testing of this integrated workflow yields quantification of >15,000 peptides from over 1,000 proteins. From these results, over 65% of these peptides achieved a CV% below 20%, and the overall median CV% is less than 15%. In the future, we will be using this semi-automated proteomic platform to comprehensively investigate the proteome changes in engineered microbes (e.g., *Pseudomonas putida*, *Rhodosporidium toruloides*).

References

- Chen Y, et al. (2019) A rapid methods development workflow for high-throughput quantitative proteomic applications. PLOS ONE 14(2): e0211582. https://doi.org/10.1371/journal.pone.0211582
- Chen Y, et al., "An automated 'cells-to-peptides' sample preparation workflow for high-throughput, quantitative proteomic assays of microbes", J. Proteome Res.2019. doi: 10.1021/acs.jproteome.9b00455.

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.

Multiple Ions in Ionic Liquid for Lignocellulosic Biomass Processing – A Boon or a Bane

Hemant Choudhary[1,2*] (hchoudhary@lbl.gov), Alexander Yao,[1] Mood Mohan,[1,2] Alberto Rodriguez,[1,2] Harsha D. Magurudeniya,[1,2] Jeffrey G. Pelton,[3] Anthe George,[1,2] Blake A. Simmons,[1,4] **Seema Singh**,[1,2] and **John M. Gladden**[1,2]

[1]Joint BioEnergy Institute, Emeryville; [2]Sandia National Laboratories, Livermore; [3]University of California, Berkeley; [4]Lawrence Berkeley National Laboratory, Berkeley

http://jbei.org

Project Goals: Affordable and scalable deconstruction with 90% yields of all major intermediates

The global energy and sustainability policies have stimulated new technology trends based on renewable energy resources such as lignocellulosic biomass (composed of cellulose, hemicellulose, and lignin) has accelerated. Nevertheless, a major constraint in the efficient utilization of lignocellulosic biomass is the embedment of crystalline cellulose within the polymer matrix of hemicellulose and lignin necessitating a pretreatment process to expose the constituents and facilitate further transformations. Significant efforts in this field has established the potential and benefits of ionic liquid (IL, organic salt that melts below 100 °C) based biomass pretreatment/processing owing to their outstanding ability to dissolve, fractionate, and convert biopolymers.

Research efforts from our group and others have established that both the cation and anion in an IL have a distinct mechanism of action and are active agents in the pretreatment. Thus, the incorporation of multiple ions with known distinct pretreatment mechanisms in an IL paves the path to develop new strategies to boost the pretreatment efficiency while reducing the cost associated with the pretreatment step. Interestingly, minimal efforts have been made to integrate the respective advantageous properties of various ions in one IL to afford a clean, viable, energy intensive, and economical biomass pretreatment method.

In this study, novel IL systems consisting of multiple ions known to be effective at biomass pretreatment were tested on woody and grassy biomass. Molecular simulations and experimental results established the synergistic advantages of combining specific individual components in these systems. For pine (woody) biomass, pretreatment with the combination of imidazolium, cholinium, acetate, and lysinate ions achieved 80% glucose and 70% xylose yields at high biomass loading. For sorghum biomass, an IL system comprising of cholinium, lysinate, and palmitate ions not only enabled a 98% glucose yield but was also found to be biocompatible in a one-pot configuration, producing the biofuel precursor bisabolene using an engineered strain of the yeast

Rhodosporidium toruloides. Additionally, the type of ions in isolation or combination also had a remarkable effect on the structure, composition, and content of lignin. The lignin content varied from ~40-75%, whereas, the weighted average molecular weight of these lignin samples was in the range of 3800-9700 Da. The HSQC analysis of these samples revealed the dominant effect of an ion over another when multiple ions were used in combination for biomass processing.

References

1. A. Yao, H. Choudhary, M. Mohan, A. Rodriguez, H. D. Magurudeniya, J. G. Pelton, A. George, B. A. Simmons, J. M. Gladden, "*Can Multiple Ions in an Ionic Liquid Improve the Biomass Pretreatment Efficacy?*" submitted for publication.
2. H. Choudhary, H. D Magurudeniya, J. M. Gladden, B. A Simmons, "*Use of a plurality of salt ionic liquids in the pretreatment of biomass*," US Provisional Patent Application 63/129,494; (Filed on December 22, 2020).

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: The development of an automated platform for the use of Polyketide Synthases for Biofuels and Sustainable Chemical Production

Alberto Nava[4] (aanava@lbl.gov), Robert Haushalter[1,3], Tyler Backman[1,3], Cameron Coates[1,2], Nurgul Kaplan Lease[1,2], Zilong Wang[1,3], Ramu Kakumanu[1,3], Bashar Amer, Edward Baidoo, **Jay D. Keasling[1,3,4]**

[1]DOE Joint BioEnergy Institute; Emeryville, CA 94608, USA; [2]DOE Agile BioFoundry; Emeryville, CA 94608, USA; [3]Lawrence Berkeley National Laboratory; Berkeley, CA 94720, USA, [4]University of California, Berkeley, 94720 CA, USA.

http://jbei.org

Project Goals:
This projects aims to establish an automated platform for the rational design and engineering of PKSs to produce biofuels and biochemicals. The application of retrobiosynthesis and data-driven, high-throughput workflow are being combined to identify targets and engineer PKS pathways for their production.

Abstract text.
Retrobiosynthesis is an approach to the design of biosynthetic pathways, which involves simulating possible biosynthesis routes in reverse, one step at a time. Modular type I polyketide synthases (PKSs), due to their collinear biosynthetic logic, have enormous potential as a retrobiosynthesis platform and have recently been engineered to produce several biofuels and bioproducts. However, the full potential of PKSs has yet to be realized for a variety of reasons including difficult cloning and heterologous expression due to the prevalence of repetitive sequences, their large size, as well as an incomplete understanding of protein-protein and substrate-protein interactions. With rapid advances in the fields of automation, machine learning, DNA synthesis, and DNA sequencing, a data-driven approach to exploring and optimizing PKS function becomes an attractive method. Here we present our work at the Joint Bioenergy Institute (JBEI) to develop an automated platform for the rational design and engineering of PKSs. We present preliminary results on the efficacy and throughput of the platform as well as preliminary characterization studies of novel chimeric PKSs. Work is ongoing to integrate features of the PKSs and their observed behavior into a model which may be used for the prediction of optimal chimeric PKSs

References
Zargar, A. *et al*. *Metab Eng*. **61**, 389-396 (2020)
Zargar, A. *et al*. *J Am Chem Soc* **142**, 9896–9901 (2020)
Yuzawa, S. *et al. Nat Commun* **9**, 4569 (2018)
Hagen, A. *et al*. *ACS Synth Biol* **5**, 21–27 (2016)
Chen, Y. *et al. J Proteome Res* **18**, 3752–3761 (2019)

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, supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy

Technologies Office, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy

Title: A mass spectrometry based high-throughput platform to study lignin modifying enzymes

Kai Deng[1]* (kdeng@lbl.gov), Nicole Ing[1], Le Thanh Mai Pham[1], Carolina A. Barcelos[1], Kenneth L. Sale[1], Paul D. Adams[1], Trent R. Northen[1], **Jay D. Keasling**[1]

[1] Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA, USA

https://www.jbei.org/

Project Goals: Short statement of goals.

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.

Abstract:

Lignocellulosic biomass has great potential to create sustainable drop-in replacements for conventional petroleum-based fuels. While significant advances have been made in utilizing the cellulose and hemicellulose components of the lignocellulosic biomass, efficient lignin depolymerization remains an enigmatic challenge. Overcoming this challenge to valorize lignin has the potential to dramatically lower the overall cost of biofuel production. However, lignin's complex and variable structure, coupled with our very limited understanding of enzymes that depolymerize lignin, make this a difficult challenge to overcome. As an important first step to rapidly accelerate our understanding of lignin active enzymes we have developed a high throughput analytical assay to assess the bond-specific activities of lignin modifying enzymes/or enzyme cocktails. This has been done by extending the nanostructure-initiator mass spectrometry (NIMS) assays we have developed to study various GH enzymes using model glycan substrates. Recently, we have shown that β-O-4 phenolic and non-phenolic dimeric model compounds are suitable for detailed analysis of laccase enzyme mixtures and peroxidases (e.g. lignin peroxidases,versatile peroxidases, and horse radish peroxidases). Building on these successes we have now synthesizing five other model compounds representing important lignin linkages (β-β, 4-O-5, β-5, 5-5', dibenzodioxocin). Together this collection of lignin dimeric model compounds covers all major lignin linkages. They can be easily integrated into our mass spectrometry-based high throughput assay to study the specificity and activity of various lignin modifying enzymes. These substrates can also be combined with our existing platform of glycan assays to create a high throughput screen for overall lignocellulose decomposition. The information obtained from our assays will help us to develop high performance enzyme cocktails to efficiently break down biomass for bioenergy production.

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.

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: Develop the fundamental understanding of cell wall biology and develop tools to facilitate bioenergy crop improvement.

Abstract: Lignocellulosic biomass will be a major sustainable feedstock for the burgeoning bioeconomy. Understanding the biosynthesis and nanoarchitecture of the plant cell wall is important foundational knowledge to enable predictive cell wall engineering. Multi-dimensional solid-state NMR spectroscopy facilitates the detailed structural elucidation of 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, we present our recent findings in wild type sorghum secondary cell walls. With 2D INADEQUATE experiments, we demonstrate that, unlike dicot and softwood plant cell walls, most grass cell wall xylan is in the three-fold screw conformation. Also, we use PDSD experiments to show that three-fold screw xylan is responsible for the most of the cellulose-xylan interactions via amorphous cellulose. Additionally, we determine that sorghum secondary walls have approximately three fold more amorphous cellulose compared to Arabidopsis, a model dicot plant. We propose a model of grass secondary cell wall with a new configuration of cellulose-xylan interaction, which will provide insights for future sorghum 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

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.

ART: A machine learning Automated Recommendation Tool for synthetic biology

Tijana Radivojevic,[1,2,3], Zak Costello,[1,2,3] Kenneth Workman,[1,3,4] William Morrell,[1,5] Mark Forrer,[1,2,5] Somtirtha Roy,[1,3] Nahan Hillson,[1,2,3] **Jay Keasling (JBEI PI),** [2,3,4,6,7,8] and Hector Garcia Martin[1,2,3,9,10*] (hgmartin@lbl.gov)

[1]DOE Agile BioFoundry, Emeryville, CA 94608, USA. [2]Biofuels and Bioproducts Division, DOE Joint BioEnergy Institute, Emeryville, CA 94608, USA. [3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Department of Bioengineering, University of California, Berkeley, CA 94720, USA., [5]Biotechnology and Bioengineering and Biomass Science and Conversion Department, Sandia National Laboratories, Livermore, California 94550, United States, [6]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs., Lyngby, Denmark. [7]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA, [8]Center for Synthetic Biochemistry, Institute for Synthetic Biology, Shenzhen Institutes of Advanced Technologies, Shenzhen, China, [9]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [10]BCAM, Basque Center for Applied Mathematics, Bilbao 48009, Spain.

http://jbei.org

Project Goals: To develop machine learning methods to effectively guide bioengineering

Synthetic biology allows us to bioengineer cells to synthesize novel valuable molecules such as renewable biofuels or anticancer drugs. However, traditional synthetic biology approaches involve ad-hoc engineering practices, which lead to long development times. Here, we present the Automated Recommendation Tool (ART[1]), a tool that leverages machine learning and probabilistic modeling techniques to guide synthetic biology in a systematic fashion, without the need for a full mechanistic understanding of the biological system. Using sampling-based optimization, ART provides a set of recommended strains to be built in the next engineering cycle, alongside probabilistic predictions of their production levels[2]. We demonstrate the capabilities of ART on simulated data sets, as well as experimental data from real metabolic engineering projects producing renewable biofuels, hoppy flavored beer without hops, fatty acids, and tryptophan. We also discuss the limitations of this approach, and the practical consequences of the underlying assumptions failing. A fundamental part of using machine learning in synthetic biology involves the availability of large amounts of high-quality training data. We show how to use the Experiment Data Depot (EDD) to store, visualize and export data in a standardized fashion.

References

1. Radivojević, Tijana, et al. "A machine learning Automated Recommendation Tool for synthetic biology." Nature communications 11.1 (2020): 1-14.
2. Zhang, Jie, et al. "Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism." Nature communications 11.1 (2020): 1-13.

This work was part of the Agile BioFoundry (http://agilebiofoundry.org) and the DOE Joint BioEnergy Institute (http://www.jbei.org), supported by the U. S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, and the Office of Science, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the

U.S. Department of Energy. NJH was also supported by the DOE Joint Genome Institute (https://jgi.doe.gov) by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DEAC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. This research is also supported by the Basque Government through the BERC 2014-2017 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323.

Development of Automated Workflows and Data Processing to Enable the Conversion of Bioenergy Crops into Biofuels and Bioproducts

Nurgul Kaplan Lease[1,3,4] (nkaplan@lbl.gov), Stephen Tan[1,3,4], Cameron Coates[3,4], Alberto Nava[5], Venkataramana Pidatala[2,4], Garima Goyal[1,3,4], Joshua McCauley[1,3,4], Shipra Gupta[1,3,4] , and Nathan J. Hillson[1,3,4]

[1]Technology and [2]Deconstruction Divisions, DOE Joint BioEnergy Institute; Emeryville, CA 94608, USA; [3]DOE Agile BioFoundry; Emeryville, CA 94608, USA; [4]Lawrence Berkeley National Laboratory; Berkeley, CA 94720, USA, [5]University of California, Berkeley, 94720 CA, USA.

Project Goals:

The development of cost-effective methods for the deconstruction of plant biomass and the subsequent microbial conversion thereof to biofuels and bioproducts is greatly enabled by robust automated workflows and rapid and flexible data acquisition and processing methodologies. The use of robotic platforms to perform labor-intensive multi-step biological tasks can increase research productivity and lower costs by reducing experimental error rates and providing more reliable and reproducible experimental data. Such biological tasks include (but are not limited to) screening ionic liquids for pretreatment efficiencies across feedstocks, screening feedstocks for sugar yields and conversion into bioproducts, and the construction and sequence validation of DNA molecules. Here, we present updates on several automated workflows under development for these types of biological tasks, which incorporate solid/liquid handling methods as well as data processing automation (via jupyter notebooks) that support the missions of the Joint BioEnergy Institute, Bioenergy Research Centers, and other DOE programs and projects.

JBEI Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

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, and was part of the Agile BioFoundry (http://agilebiofoundry.org) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy

Lignin manipulation via plant synthetic biology to produce a biodegradable polyester precursor with concurrent improvement in biomass quality for biofuel production

Chien-Yuan Lin[1,*] (chienyuanlin@lbl.gov), Aymerick Eudes[1], Khanh Vuu[1], Edward Baidoo[1], Bashar Amer[1], Patrick Shih[1], Henrik V. Scheller[1], and **Jay D. Keasling**[1]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA

http://jbei.org

Project Goals: Engineer bioenergy crops with improved biomass and sustainability traits

Lignin manipulation targeting the shikimate pathway can redirect the metabolic flux of lignin biosynthesis for bioproduct production and/or reduce biomass recalcitrance [(1)]. By expressing a bacterial 3-dehydroshikimate dehydratase (QsuB), we observed concomitant accumulation of protocatechuate (PCA), reductions of lignin content, and more than a twofold improvement in saccharification efficiency from the genetic-engineered biomass [(2)]. Dual expression of PCA decarboxylase with catechol 1,2-dioxygenase (CatA) can convert PCA to muconic acid (MA) in planta, which is a value-added bioproduct for adipic acid, terephthalic acid, and caprolactam [(3)]. 2-Pyrone-4,6-dicarboxylic acid (PDC) is a promising building block biomaterial that can serve as a starting monomer to manufacture performance-advantaged polymers and biodegradable plastics. The lack of a chemical synthesis method has hindered the large-scale PDC utilization, and metabolic engineering approaches for its biosynthesis have recently emerged. In this study, we first increased the carbon flux through the shikimate pathway for PCA production by overexpressing a feedback-insensitive 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (AroG*). Co-expression of PCA 4,5-dioxygenase (PmdA and PmdB) with 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase (PmdC) enabled the PDC biosynthesis for the first time in plant chassis while maintaining the low biomass recalcitrance under QsuB-overexpressing background. Using plant synthetic biology, we stacked all five genes (AroG*, QsuB, PmdA, PmdB, and PmdC) in a single T-DNA and successfully introduced them into the wildtype Arabidopsis. The expression of the five genes enabled PDC production at high titers (up to 3% dry weight), accompanied by a substantial reduction in lignin content and improvements in biomass saccharification efficiency. The lignin manipulation strategy for PDC biosynthesis is currently implementing in bioenergy crops to stack low-recalcitrance traits with value-added bioproduct. We hope these promising results can ultimately improve the biomass quality and value towards the sustainable development of biorefineries.

References

1. *Lin CY, Eudes A. Strategies for the production of biochemicals in bioenergy crops. Biotechnology for biofuels. 2020 Dec;13:1-25.*
2. *Eudes A, Sathitsuksanoh N, Baidoo EE, George A, Liang Y, Yang F, Singh S, Keasling JD, Simmons BA, Loqué D. Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. Plant Biotechnology Journal. 2015 Dec;13(9):1241-50.*
3. *Eudes A, Berthomieu R, Hao Z, Zhao N, Benites VT, Baidoo EE, Loqué D. Production of muconic acid in plants. Metabolic engineering. 2018 Mar 1;46:13-9.*

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.

Engineering *Rhodosporidium toruloides* for Bioproduction of Polyketide Synthase and p-Coumarate Derived Compounds

Peter B. Otoupal,[1,2]Presenting Author* (PeterOtoupal@lbl.gov), Gina M. Geiselman,[1,2] Joonhoon Kim,[1,3] Robert W. Haushalter,[1,4] Hyungyu Lim,[5] Alberto Rodriguez,[1,2] Adam M. Feist,[1,5] Aindrila Mukhopadhyay,[1,4] Jay D. Keasling,[1,5] Jon K. Magnuson[1,3] & **John M. Gladden[1,2]**

[1]DOE Joint BioEnergy Institute, USA
[2]Biomass Science and Conversion Technologies, Sandia National Laboratories, USA
[3]Chemical and Biological Processing Group, Pacific Northwest National Laboratory, USA
[4]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California,
[5]Department of Bioengineering, University of California, San Diego, CA, USA

https://www.jbei.org/research/deconstruction/fungal-biotechnology/

Project Goals/Abstract:

The oleaginous yeast *Rhodosporidium toruloides* is an ideal chassis for the valorization of lignocellulosic biomass into value-added products due to its natural capacity to co-utilize multiple carbon sources, and its ample pool of malonyl-CoA. Here we leverage this advantage by incorporating seven heterologous pathways to convert p-coumarate and/or malonyl-CoA into useful bioproducts. Pathways for two polyketides (6-methylsalicylic acid (6MSA) and triacetic acid lactone (TAL)) were introduced. While no 6MSA production was observed, a substantial amount (1.65 ± 0.06 g/L) of TAL was secreted in standard media without any optimization, leaving much room for improvement on these already significant titers. Five other pathways for converting p-coumarate, a lignin-derived precursor, were also introduced for biosynthesis of naringenin, resveratrol, curcuminoids, 4-hydroxybenzoate (4HBA), and 2-pyrone-4,6-dicarboxylic acid (PDC). We show indications of three of these pathways working, with both the production and subsequent consumption of resveratrol, conversion of protocatechuate into a toxic aldehyde intermediate (4-carboxy-2-hydroxymuconate-6-semialdehyde) between it and PCD, and a remodeling of the p-Coumarate consumption pathway resulting in extracellular accumulation of 3.17 ± 0.10 g/L 4HBA. Finally, we prelude the use of Tolerance Adaptive Laboratory Evolution (TALE) to enable robust growth of *R. toruloides* in 20 g/L p-Coumarate. These new results broaden the already substantial spectrum of biofuels and bioproducts that *R. toruloides* has been shown to produce.

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

Pseudomonas putida soil isolates that metabolize C5 sugars and lignin-derived aromatics

Mee-Rye Park[1,2] (mrpark@lbl.gov), Rahul Gauttam[1,2], Bonnie Fong[1,2], Joseph Stevens[1,2], Blake Simmons[1,2], and Steven Singer[1,2]

[1]Joint BioEnergy Institute, Emeryville, CA, 94608; [2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720

http://jbei.org

Project Goals: The Joint BioEnergy Institute (JBEI) performs fundamental research to improve the conversion of biomass to biofuels and bioproducts. A critical aspect of current research is to maximize the carbon in plant biomass that is converted by microbial hosts. To achieve this goal, we have isolated new potential hosts within members of genus *Pseudomonas* that are natively capable of utilizing C5 sugars and lignin-derived aromatics.

Pseudomonas putida is promising host for biomass conversion to fuels and chemicals because it metabolizes sugars and aromatics from plants. However, the best studied strain*, P. puitda* KT2440, lacks the native ability to metabolize C5 sugars (xylose and arabinose), the most abundant components of hemicellulose from grasses. To maximize the range of components which can be obtained from lignocellulosic hydrolysate, pentose (C5) sugars should be utilized as feedstock for value-added products. Here, we report *Pseudomonas putida* isolates obtained from soil that grow on C5 sugars as well as lignin-derived aromatics. These isolates were in two clusters closely related to *P. putida* KT2440 and were capable of growing on glucose and *p*-coumarate at rates comparable to KT2440. One set of isolates grew on xylose and the second set of isolates grew on xylose and arabinose. Genomic analysis of the isolates indicated that homologs of the Weimberg pathway for xylose oxidation (*xylD-XylX-XylA*) were present in both isolate genomes and an oxidative pathway for arabinose oxidation (*araD-araX-AraA*) was present in the one strain. Transformation protocols were established for the both strains and deletion of the periplasmic glucose dehydrogenase eliminated the ability of the strains to grow on C5 sugars, indicating both C6 and C5 sugars were oxidized by the same protein. A CRISPRi system for one strain has been established and this system of gene knockdowns, along with RNA-seq experiments will identify the genes important for xylose and arabinose metabolism. Preliminary experiments indicated that the strain can produce Type III polyketide synthases and non-ribosomal peptide synthases, suggesting it could serve as an alternative host for fuels and chemical production.

Funding statement. *This work was performed 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.*

Expression of fungal laccases in *Pichia pastoris* and characterization using lignin depolymerization

Le Thanh Mai Pham[1,2*] (thanhmaipl@lbl.gov), Kai Deng[1,2], Trent R. Northen[1,3], Steven W. Singer[1,3], Paul D. Adams[1,3], Blake A. Simmons[1,3], **Kenneth L. Sale[1,2]**
1. Joint BioEnergy Institute, Emeryville, CA 94608, USA
2. Sandia National Laboratories, Livermore, CA, 94550, USA
3. Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
http://jbei.org

Project Goals: This work demonstrates that we can heterologously express lignin-degrading enzymes (laccases in this case) that are active on lignin-like compounds and catalyze the cleavage of specific bond types critical to depolymerizing lignin.

Abstract
Enzymatic cocktails from white fungi secretomes are potentially an efficient and cost-effective approach to developing mixtures of ligninolytic enzymes for biorefinery implementation, but complex interactions among the different agents highly can lead to difficulty in condition optimization and mechanism elucidation. Instead of mixed secretion, heterologous expression of individual genes is a valuable approach to understand the structure-function relationships of these important enzymes. Herein, we report structure-based selection of three novel laccases, designated as Cer_Lc1, Cer_Lc2, and For_lac from white-rot fungi, expression in *Pichia pastoris* and characterization of their catalytic performance, thermal stability, and solvent stability. This work aimed to improve our understanding of the mechanism underlying the enzymatic degradation of lignin by monitoring and quantifying β-O-4 ether, C_α-C_β, and C_α-C_{aryl} bond cleavage in a model lignin-like dimer synthesized for use in a nanostructure-initiator mass spectrometry (NIMS) assay. We present detailed studies of the effect of a natural mediator, syringaldehyde, on bond cleavage occurrence. This study also provides a comprehensive understanding of the structure-function and the structure-stability relationship of these novel fungal laccases, which will facilitate developing this important class of enzymes for applications in the conversion of lignin to valuable products and use in biorefinery implementations.

References

Funding statement: Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Microscale thermophoresis as a powerful tool for screening glycosyltransferases involved in cell wall biosynthesis.

Wanchen Shao[1,2*] (Wanchenshao@lbl.gov), Rita Sharma[3], Pradeep Kumar Prabhakar[4], Jose Henrique Pereira[1,5], Nurgul Kaplan[1,6], Mads H. Clausen[7], Breeanna Urbanowicz[4], Paul D. Adams[1,6], Henrik V. Scheller[1,2,8] and **Jay D. Keasling**[1,6,9]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Delhi, India; [4]Complex Carbohydrate Research Center, University of Georgia, Athens, GA; [5]Molecular Biophysics and Integrated Bioimaging (MBIB) Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [6]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [7]Department of Chemistry, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark; [8]Department of Plant and Microbial Biology, University of California, Berkeley, CA; [9]Department of Chemical & Biomolecular Engineering and Department of Bioengineering, University of California, Berkeley, CA

http://jbei.org

Project Goals: Develop the fundamental understanding of cell wall biology and Develop tools to facilitate bioenergy crop improvement.

Identification and characterization of key enzymes associated with cell wall biosynthesis and modification is fundamental to gain insights into cell wall dynamics. However, the challenge is that activity assay of glycosyltransferase is very low throughput and acceptor substrates are generally not available. Here we optimized and validated microscale thermophoresis (MST) to achieve high throughput screening for glycosyltransferase substrates. MST is a powerful new method for the quantitative analysis of protein-ligand interactions with low sample consumption. The technique is based on the motion of molecules along local temperature gradients, measured by fluorescence changes. We optimized the methods to allow the determination of binding affinity by MST without purification of the target protein from the cell lysate. The application of this MST method to beta-1,4-galactosyltransferase AtGALS1 validated the capability of substrate screening. We also expanded the application to detect binding of AtGALS1 to its acceptor in presence of UDP. Furthermore, we used the method in combination with activity assays to identify potential features of the poplar ortholog PtGALS1 in substrate binding and catalytic activity. For this study we used a set of PtGALS1 mutant variants that had been designed based on the crystal structure of the enzyme as determined by X-ray crystallography. An automated pipeline using automated liquid handling technique was established to facilitate the high-throughput substrate screening of glycosyltransferases in sorghum, and until now twenty sorghum GT61s were screened with potential substrates, which will enable selection of candidates for further studies and engineering.

This work was funded by 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 DEAC0205CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

Engineering Plants with Novel Metabolic Pathways as a Production Platform for Bioproducts

Khanh M. Vuu[1,2] (kvuu@lbl.gov), Aymerick Eudes[1,2], Veronica Teixeira Benites[1,2], Bashar Amer[1,2], Edward Badoo[1,2], Dominique Loque[1,2] and **Patrick M. Shih**[1,2,3*]

[1]Joint BioEnergy Institute, Emery Station East, 5885 Hollis St, 4th Floor, Emeryville, California 94608, USA. [2] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. [3] Department of Plant Biology, University of California, Davis, Davis, California, USA.

http://jbei.org

Project Goals: To engineer bioenergy crops with improved biomass and sustainability traits.

Metabolic engineering of plant pathways poses a great challenge due to limited understanding in plant metabolism and bioengineering approaches that can be applied in plant biology[3]. Howbeit, genetic engineering in plants has gained many interests in recent years as an alternative to create sustainable resources opposed to their microbial counterparts, like yeast and bacteria. In planta, only a small percentage of biosynthetic pathways are known and, by the same token, remain a pool of untapped resources and machinery to synthesize complex metabolites. As such, we explore the potential of using plant-based production of the chemical, muconic acid (MA), an intermediate molecule that can be derived into several bioplastics. However, existing biological approaches are mostly focused on producing MA from a glucose feedstock in microbes. Here, we exploit plant primary metabolism to test and optimize various metabolic routes[1-2], enabling direct production of MA via photosynthesis. Plant-based metabolic engineering efforts may enable a more sustainable means of producing chemicals of interest and decrease our dependence on current practices that are heavily dependent on petroleum feedstocks.

References

1. Averesch, N.J.H. and Kromer, J.O. (2014). Tailoring strain construction strategies for muconic acid production in S. cerevisiae and E. coli. Metabol Eng Commun. 1, 19-28.
2. Bentley, R. and Haslam, E. (1990). The shikimate pathway–-a metabolic tree with many branches. CRC Crit. Rev. Biochem. MOI. Biol. 25, 307-384.
3. Shih, P. (2018). Towards a sustainable bio-based economy: Redirecting primary metabolism to new products with plant synthetic biology. Plant Sci. 273, 84-91.

Modifying terpenes for the development of biofuels and bioproducts

Xi Wang[1,2,*] (xiwang@lbl.gov), Hsi-Hsin Lin[1,3], Jose Henrique Pereira[1,4], Susan Tsutakawa[4], Xinyue Fang[1,2,5], Paul D Adams[1,4,6], Aindrila Mukhopadhyay[1,2], Brent H. Shanks[1,3], **Taek Soon Lee**[1,2,*] (tslee@lbl.gov), Jay D. Keasling[1,2,6,7] (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]Department of Chemical and Biological Engineering, Iowa State University, Ames, IA 50014, USA
[4] Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[5] Department of Molecular & Cell Biology, University of California, Berkeley, CA 94720, USA
[6] Department of Bioengineering, University of California, Berkeley, CA 94720, USA.
[7] Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720, USA.

http://www.jbei.org/research/divisions/biofuels-and-bioproducts/pathway-metabolic-engineering/

Project Goals: We aim to develop and demonstrate a platform for on-demand prioritization and validation of routes to biofuels and bioproducts at JBEI. Modifying terpenes show the possibility to expand the portfolio of isoprenoid biosynthesis. Combining chemical and biological catalysis shows advantages to achieve both productivity and diversity during product development. In this study, chemical and biological routes were investigated for modifying terpene molecules toward the production of biofuels and bioproducts. We also develop an enzyme fusion strategy by linking terpene synthase and P450 together to improve the productivity of terpene oxidation.

The functionalization of terpenes is a versatile route to the development of useful derivatives that can be further converted to value-added products. Combining chemical and biological catalysis shows advantages to achieve both productivity and diversity during product development. In this study, the chemical conversion of monoterpene (limonene and 1,8-cineole) to *p*-cymene was investigated. While limonene bioproduction is constrained by its toxicity, 1,8-cineole, a less toxic precursor to limonene, can be also converted to p-cymene with similar efficiency using a bifunctional metal/acid chemical catalyst. This suggests that 1,8-cineole could be a preferred hand off point from biological to chemical conversion. In a biological conversion, cytochrome P450 enzymes were investigated for the enzymatic oxidation of monoterpene. Biosynthetic pathways were further explored to develop possible derivatives, such as carvolactone, a monomer of thermoplastic polyester. Particularly, we developed an enzyme fusion strategy by linking terpene synthase and P450 together to overcome the low availability of hydrophobic and volatile terpene molecules for the P450 reaction [1]. The engineered fusion proteins of 1,8-cineole synthase and $P450_{cin}$ showed an increase of the hydroxylation of 1,8-cineole up to 5.4-fold. Structural analysis of the fusion proteins indicated a dimer formation with preferred orientations of the active sites of two domains. We also applied the enzyme fusion strategy to the oxidation of a sesquiterpene epi-

isozizaene, in which a 90-fold increase was observed in albaflavenol production. This study demonstrated a platform for the development of biofuels and bioproducts via terpene modification.

References

1. Wang, X., Pereira J. H., Tsutakawa, S., Fang, X., Adams, P. D., Mukhopadhyay, A., Lee, T. S. Efficient production of oxidized terpenoids via engineering P450 fusion proteins. *Metabolic Engineering*, 2021, 64:41–51.

Funding statement.

The material presented in this poster is part of JBEI - JointBioEnergy Institute, and is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under its contract number with Lawrence Berkeley National Laboratory DE-AC02-05CH11231.

High Throughput Bioengineering Using a Microfluidic Platform

Lauren Washburn[1,2]*(lwashburn@lbl.gov), Jess Sustarich[1,2]*(jsustarich@lbl.gov), 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 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 tiny (nL) 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.

One example platform is a device for multiplexed electroporation for automating CRISPR-based MAGE recombineering in *E. coli*.[6] 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 transform cells by on-chip electroporation. Additional reservoirs allow recovery incubation and screening on 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.

Evolution of root nodule symbiosis & engineering of symbiotic nitrogen fixation in *Populus* sp.

Thomas B. Irving[1*] (tbirving@wisc.edu), Sanhita Chakraborty[1], Sara Knaack[1], Daniel Conde[2], Matias Kirst[2], Sushmita Roy[1], and **Jean-Michel Ané**[1]

[1]University of Wisconsin-Madison, Madison, WI; [2]Univeristy of Florida, Gainesville, FL

https://nitfix.org/

Project Goal: Transfer the root nodule symbiosis from legumes to the bioenergy crops *Populus* sp.

Legumes (Fabales) and close relatives of the Fagales, Cucurbitales, and Rosales can associate efficiently with nitrogen-fixing bacteria, in symbioses that lead to the development of root nodules. Legumes, in particular, host bacteria called rhizobia in their root nodules. Studies in model legumes such as *Medicago truncatula* and *Lotus japonicus* identified that (1) rhizobia colonize legume roots intracellularly following the recruitment of the arbuscular mycorrhizal (AM) signaling pathway known as the common symbiosis pathway (CSP), and (2) nodule formation (organogenesis) evolved from the lateral root developmental pathway, and is distinguished from it by the local cytokinin accumulation that requires the transcription factor Nodule Inception (NIN).

Comparative phylogenomics suggest that the root nodule symbiosis appeared once in the last common ancestor of the Fabales, Fagales, Cucurbitales, and Rosales, and lost multiple times within this "nitrogen-fixing clade"[1]. *Populus* sp. are bioenergy crops and close relatives to the "nitrogen-fixing clade". They retain several key genes known to be required for nodule symbiosis, including *NIN*[2]. *Populus* sp. is easily transformable and represents an excellent model for synthetic biology approaches.

We demonstrated that certain species of rhizobia can activate the AM signaling pathway in *Populus* sp. We studied *Populus* sp. responses to these signals using cell biology (calcium spiking) and transcriptomic approaches (RNA-seq)

We are also working on characterizing the role of cytokinin signaling and *NIN* in *Populus* sp., generating overexpression and knock-down transgenic lines for the members of the *NIN* clade and symbiosis-related cytokinin receptors in *Populus* sp. Overexpression of some *NIN* genes altered root architecture in response to nitrate or cytokinin, and overexpression of cytokinin receptors occasionally produced nodule-like structures on *Populus* sp. roots treated with cytokinin. The initiation of lateral root organs by cytokinin is a critical feature distinguishing root nodules from lateral roots, since cytokinin inhibits the formation of the latter.

We are also characterizing the *cis*-regulatory elements of *NIN* using bioinformatic approaches such as Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) and identifying the targets of NIN and other symbiotic transcription factors using DNA affinity purification sequencing (DAP-seq). Furthermore, to understand if protein sequence changes to

NIN were required for nodulation, we are performing cross-species rescue experiments in the *M. truncatula nin-1* mutants.

In the long term, engineering a nitrogen-fixing root nodule symbiosis in *Populus* sp. would greatly enhance biomass productivity on marginal soils and the sustainability of bioenergy production. Understanding the function and mechanism of action of *Populus* sp. *NIN* and other symbiotic genes will substantially aid in this process.

References

1. A Resurrected Scenario: Single Gain and Massive Loss of Nitrogen-Fixing Nodulation; van Velzen *et al*., Trends in Plant Science, 2019
2. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis; Griesmann *et al*., Science, 2018

This project is funded by the U.S. Department of Energy grant #DE-SC0018247

Identification of the Minimal Genetic Toolkit Required for Nitrogen Fixation Using Comparative Genomics and Single-Cell Transcriptomics

Daniel Conde,[1] Kelly Balmant,[1] Paolo Triozzi,[1] Wendell J. Pereira,[1] Thomas B. Irving,[2] Lucas G. Maia,[2] Sanhita Chakraborty,[2] Heather R. Kates,[1] Sara A. Knaack,[2] Henry W. Schmidt,[1] Pamela Soltis,[1,3] Douglas Soltis,[1] Robert Guralnick,[3] Ryan Folk,[4] Sushmita Roy,[2] Jean-Michel Ané,[2] and **Matias Kirst**[1]* (mkirst@ufl.edu)

[1]University of Florida, Gainesville, FL; [2]University of Wisconsin, Madison, WI; [3]University of Florida, Florida Museum of Natural History, Gainesville, FL; Mississippi State University, Starkville, MS.

http://nitfix.org; Twitter: @nit_fix

Project Goals: The collaborative project "Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar" aims to identify the evolutionary events that enabled the symbiotic relationship between nodulating plants and nitrogen-fixing bacteria. Identifying a minimum genetic toolkit required for nodulation is a first step towards genetically engineering this trait into bioenergy crops. We are applying multiple methods to discover the genome novelties of nitrogen fixers. Here we present the outcome of two approaches: comparative genomics of nodulating and non-nodulating species and single-cell transcriptomics of nodule development in *Medicago*. Because multiple genes are likely to be required to engineer nitrogen fixation, we also describe a newly devised synthetic biology method for high-throughput parallel testing of multigene combinations for their function on nodule formation and nitrogen-fixation.

Nitrogen (N) is essential for plant growth because of its fundamental role as a component of DNA, RNA, and amino acids. Most plants cannot obtain N directly from the atmosphere but depend on its availability in the soil, in the form of nitrate, ammonium, or amino acids. Nitrogen fertilization comes at a high financial and environmental cost. In contrast, several species of four angiosperm orders (Fabales, Fagales, Rosales, and Cucurbitales) can establish a symbiotic relationship with bacteria that convert atmospheric N to ammonium. This group of species is referred to as the nitrogen-fixing clade (NFC). After bacteria infect the root, they become established in newly developed nodules. While the plant benefits by absorbing the produced ammonia, it provides nutrients to the bacteria.

A single evolutionary event may have created a predisposition for the symbiotic relationship between N-fixing bacteria and plants in the NFC. Our phylogenetic results show that many gains and losses followed this early event. In this complex evolutionary scenario, several genes and regulatory elements are likely critical for engineering N-fixation into bioenergy crops. We are applying approaches such as comparative genomics of nodulating and non-nodulating species and transcriptome analysis of single cells in nodules to identify them.

Comparative genomics – To detect genomic novelties that enable nodulation using comparative genomics, we aligned plant genomes from N-fixing and non-fixing species within the NFC and an outgroup. Next, we identified loci that display evolutionary signatures compatible with gains or losses of the trait. Based on phylogeny and analysis of coding sequences, we find strong evidence that a deletion in a Lysin Motif Receptor Like Kinases is associated with symbiosis in the NFC,

having appeared multiple times in nodulating species. A similar strategy was used to survey the regulatory regions of genes implicated in nodulation. This analysis showed that a set of conserved non-coding sequences immediately upstream of putative homologs of the *Mt*CRE1 Cytokinin Receptor are absent among species in the outgroup. This cytokinin receptor has been previously shown to be an essential part of the signaling mechanism that triggers nodule formation. The effect of these and other genes and regulators are being functionally validated to verify their molecular role in nodulation.

Single-cell transcriptomics – The cell-specific gene expression program that leads to cell division activation in the root pericycle and cortex and, posteriorly, the nodule primordia formation remains largely unknown. Uncovering the regulators of this program may determine the genetic elements critical for engineering nodule formation in non-nodulating species. We developed a high-quality nuclei isolation protocol suitable for use in the 10× Genomics Chromium system to address this question. This method was used to characterize the single-cell transcriptome of *Medicago* roots after treatment with the N-fixing bacteria *Sinorhizobium meliloti* (Fig. 1). Analysis of the single-cell transcriptome data identified cell-type-specific gene expression programs, including potential regulators of cell division initiation in the inner root cells, such as *SCARECROW*, *YUCCA 1*, and *PLETHORA 5*. Most strikingly, gene expression signatures specific to mature nodule compartments appear to emerge within few hours of treatment with *S. meliloti*. Thus, cell fate may be determined very early in this developmental process. We are now applying pseudotime, trajectory analysis to detect additional regulators for validation, as described below.

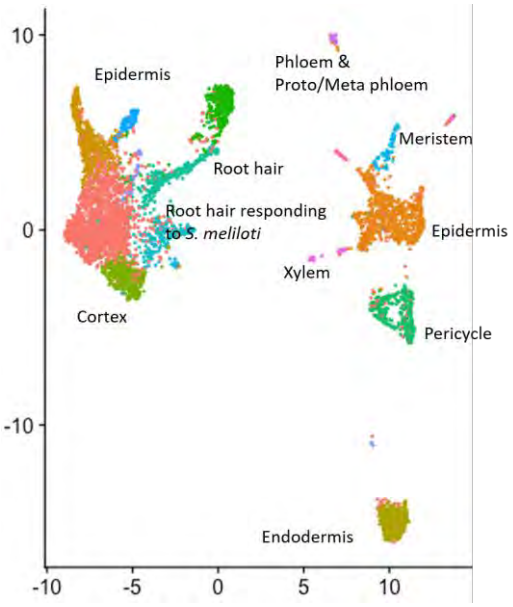


Fig. 1. Clustering of single-cell RNA-seq data generated from *Medicago* roots treated with *Sinorhizobium meliloti* for 24 hrs. Cell type were annotated according to the expression of known marker genes.

Nodule organogenesis is controlled by multiple genes that act in concert to regulate this complex developmental program. Evidence from work described above indicates that multiple genetic elements will have to be engineered to introduce nodule organogenesis in bioenergy crops. This fact motivated us to design a Golden Gate-based system to screen multigene combinations for engineering nodule-like structures in poplar. This system allows the random combination of genes and promoters up to six transcriptional units simultaneously to be tested in *Agrobacterium rhizogenes*-mediated poplar roots. Preliminary results from the application of this system indicate the possibility of engineering nodule-like structures in poplar roots.

This project is funded 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).

Leveraging temporal change in chromatin accessibility to predict regulators of N-fixing symbiosis in *Medicago* with dynamic regulatory module networks (DRMNs)

Sara Knaack[1,*] (saknaack@wisc.edu), Daniel Conde,[2] Lucas Gontijo Silva Maia,[3] Thomas Irving,[3] Kelly Balmant,[2] Matthew Crook,[4] Christopher Dervinis,[2] Heather Kates,[5] Wendell Pereira,[2] Ryan Folk,[6] Robert Guralnick,[5] Douglas Soltis,[5,7]Pamela Soltis,[5] Jean-Michel Ane,[3] Sushmita Roy[1,8]. Matias Kirst,[2]

Wisconsin Institute for Discovery, University of Wisconsin at Madison, Madison, WI; [2] School of Forest Resources and Conservation, University of Florida, Gainesville, FL; Departments of Bacteriology and Agronomy, 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

http://NitFix.org

Project Goals: We seek to identify genomic elements required for the symbiotic relationship between nodulating plants and nitrogen (N)-fixing bacteria. To identify such elements we generated parallel RNAseq and ATACseq time course data from *Medicago* plant subjected to lipo-chitooligosaccharide (LCO) (Nod factors) treatment. The gene regulatory network involved in LCO response was identified by applying a novel computational method to identify dynamically expressed (transitioning) genes, and predict key regulators of these genes. Prioritized regulators and their target genes were examined based on current literature and will be validated experimentally in the future.

Nitrogen fixation occurs naturally in a small number of plant species through a symbiotic relationship in plant root nodules colonized by N-fixing bacteria. Components of the symbiosis pathways are known, yet the gene regulatory network controlling this process is not thoroughly understood. We measured transcriptomic (with RNA-seq) and chromatin accessibility (with ATAC-seq) profiles in *Medicago* roots treated with LCOs over a 24-hour time course. LCOs are a component of the symbiotic pathway and the experiment emulates the early signaling processes in the establishment of symbiosis in *Medicago*.

To define the gene regulatory network from this parallel time course we applied a novel computational method, Dynamic Regulatory Module Networks (DRMNs). With DRMN we inferred modules of similarly-expressed genes at each time point, and per-module regulatory networks predictive of gene expression within each module. Module regulatory network edges are based on predicting gene expression from the accessibility (ATAC-seq) of gene promoters (+/- 2 kbp) and motif sites of known regulatory proteins mapped to within 10 kbp upstream to 1 kbp down-stream of a given gene transcription start site (TSS). Among the top regulators

identified in the module networks include IBM1, EDN3, MTF1, EIN3, SHY2, NSP1, and RRB9. Several (EIN3, NSP1) are known to be involved in nitrogen fixation. Our modules are furthermore enriched for root hair elongation, defense response to bacterium, chromatin organization and the MAPK cascade, recapitulating key features of symbiosis.

We leveraged the results of DRMN to define 10,176 transitioning genes (those with changing module assignment across time) and then clustered into 79 clusters. We predicted regulators of these transitioning genes using accessibility of motifs in their promoters using a structured sparsity framework, MTG-LASSO.[1] This approach leverages a group structure of similar transitioning genes to learn a statistically robust model compared to standard expression-based network inference. We identified a high confidence set of regulatory network edges which included 126 regulators (motifs) and 5,978 genes. Of the regulators identified with a large number of connections, many were identified from the module-level results (EIN3, IBM1, and MTF1), and some were particular to this analysis (including CAMTA1 and CYCRE).

To validate our regulatory network we overlapped predicted target genes of the EIN3 motif to sets of differentially expressed (DE) genes called between SKL/EIN2Δ mutant and wild type time course data from Larrainzar et al.[2] The SKL/EIN2Δ data was used for this validation because EIN2 and EIN3 function is closely related. Our inferred EIN3 targets overlap significantly (hypergeometric test $p<0.05$) with DE gene sets inferred from each time point of the Larrainzar et al.[2] dataset. Despite the differing techniques of our experiment and that of Larrainzar et al.,[2] these results demonstrate the effectiveness of our DRMN/MTG-LASSO framework for inferring regulatory networks. As future work, we will experimentally validate such predictions by functional perturbations to the regulator and measuring nodulation phenotype and downstream targets. Our dataset and predictions are a valuable resource to the plant community to study the gene regulatory programs controlling Nitrogen fixation.

References

1. Integrated Module and Gene-Specific Regulatory Inference Implicates Upstream Signaling Networks. S. Roy, et al. PLoS Comput Biol, 2013.

2. Deep Sequencing of the Medicago truncatula Root Transcriptome Reveals a Massive and Early Interaction between Nodulation Factor and Ethylene Signals. E. Larrainzar, et al. Plant Physiology, 2015.

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).

Innovations to the DayCent Biogeochemical Model to Better Simulate Carbon and Nitrogen Cycling in Bioenergy Crop Systems with Increasing Climate Variability

Danielle M Berardi,[1,2]* (danielleb@uidaho.edu), Edward Brzostek,[1,3] Melannie Hartman,[1,4] Bill Parton,[1,4] and **Tara W Hudiburg**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Department of Forest, Rangeland, and Fire Sciences, University of Idaho, Moscow; [3]Department of Biology, West Virginia University, Morgantown; and [4]Natural Resource Ecology Laboratory, Colorado State University, Fort Collins.

http://url.cabbi.bio

Project Goals: This project addresses advances in biogeochemical modeling to simulate carbon and nitrogen fluxes under varying environmental conditions so they may better quantify the extent that bioenergy crops could mitigate climate change. First, we are amending the DayCent biogeochemical model's soil module to include microbial enzyme kinetics to improve temperature and moisture response of decomposition. Second, DayCent has been updated to better simulate the unique structural and physiological attributes of perennial bioenergy grasses. Third, we are working on incorporating plant physiological responses to waterlogged soils. There is potential for flood-tolerant bioenergy crops to mitigate losses of corn and soybeans to extreme precipitation and inundation events, and we are using modeling experiments to evaluate the greenhouse gas (GHG) implications of converting flood-susceptible fields to switchgrass or miscanthus.

Soil carbon sequestration is a key component to understanding the sustainability of bioenergy feedstocks and the role they could play to mitigate climate change. Because field experiments are limited in temporal and spatial scales, using ecosystem models to simulate soil carbon fluxes in tandem with field experiments are key to understanding how policy and management decisions will affect long-term carbon sequestration and storage. Century, the monthly timestep of the DayCent biogeochemical model, has served as one of the foundational frameworks for soil carbon modeling. In recent years, there has been a call to include microbial explicit processes in soil modeling in order to better represent decomposition and soil carbon stabilization in response to warming, rewetting, and root-priming. Microbial explicit soil models have been developed in the last decade. However, they have not been widely adapted into ecosystem or Earth System scale models. Here, we are incorporating microbial process-based equations from FUN-CORPSE (a pairing of the Fixation and Uptake of Nitrogen model with the Carbon, Organisms, Rhizosphere, and Protection in the Soil Environment model) into DayCent's soil module. The new version of the model will use reverse Michaelis-Menten kinetics in the decomposition function that relates to microbial enzyme activity rather than the current decay rate-driven first order kinetic equations. To do this, the model will have explicit live microbial biomass pools in the surface and soil layers within the current DayCent model structure as well as dead microbial carbon pools that transfer directly to the passive soil carbon pool. This will allow for microbial growth and biomass to influence the rate of decomposition and represent the affinity of microbial necromass to become physically protected. Advances made with this work

can serve as an example of how to improve the representation of microbial functions for Century-based soil models in a number of ecosystem and Earth System models.

Bioenergy feedstocks such as miscanthus, switchgrass, sugarcane, and sorghum have unique structural and physiological attributes that influence carbon and nitrogen cycling. To simulate long-term dynamics of these fast-growing, high productivity bioenergy crops, we developed a new bioenergy grass PFT with additional physiological parameters for DayCent. The new PFT separates leaves and stems into separate pools rather than a single aboveground biomass pool, and it represents rhizomes as well as fine roots. This allows for leaves and stems to be parameterized with separate C:N ratios and lignin content. The new version simulates re-translocation of nitrogen from leaves to rhizomes during senescence. This stored nitrogen is then available for growth during the next growing season, allowing plant health during one growing season to affect plant growth during the next. Additionally, separation of stems and leaves can help modelers better estimate useable biomass for biofuel production.

The Renewable Fuel Standard (RFS) calls for increasing the volume of cellulosic biofuel while reducing lifecycle GHG emissions by 60%. While conversion of land used for corn-soy production to perennial, cellulosic bioenergy crops has been a controversial topic, it is becoming increasingly apparent that it will be difficult to meet GHG reduction targets only by converting uncultivated land to perennial bioenergy feedstock production. Currently, approximately 40% of corn grain is used for ethanol production and corn-soy losses have been increasing in the last decade. As much as 75% of losses are attributed to water-inundated fields. Flooding and extreme precipitation events are projected to increase due to climate change, resulting in a greater loss of corn and soy yield. Perennial bioenergy crops such as switchgrass and miscanthus are tolerant of flooding, making them a productive alternative to corn in low-lying areas that otherwise suffer recurring losses of corn and soy yield with repeat flooding events. Flooding in agricultural land typically occurs in floodplains and potholes (i.e., slightly lower-lying land that surrounding areas drain into). Figure 1 identifies counties that incurred greater than 5,000 acres of average annual losses of corn as a result of extreme precipitation and/or inundation events between 2009-2019. Using data from flooded corn, switchgrass, and miscanthus fields, we are working to improve model representation of plant physiological response to water inundation. Following model improvements, we will simulate corn-soy rotations, switchgrass, and miscanthus under current climate and increased precipitation scenarios to evaluate the GHG implications of converting corn-soy fields to perennial bioenergy crops that can withstand inundation events.

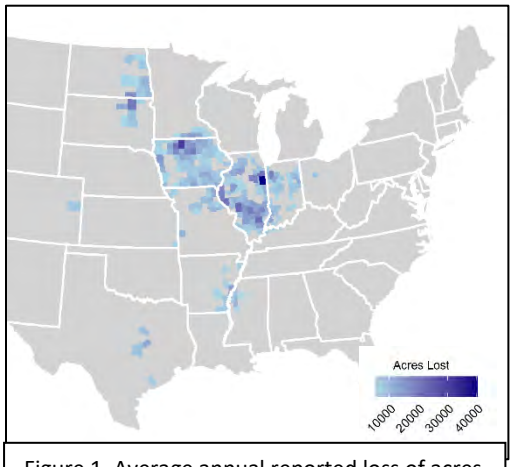


Figure 1. Average annual reported loss of acres of corn to extreme precipitation and inundation events from 2009 -2019 by county.

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.*

Engineering Vegetative Lipids in a Fast-Growing and High-Biomass Arabidopsis Line

Yingqi Cai[1,2]* (ycai@bnl.gov), Zhiyang Zhai,[1,2] Jantana Keereetaweep,[2] Hui Liu,[1,2] Changcheng Xu,[1,2] and **John Shanklin**[1,2]

[1]Center for Advanced Bioenergy and Bioproducts Innovation; [2]Biology Department, Brookhaven National Laboratory, Upton, NY

http://cabbi.bio

Project Goals: The overarching goal of our research is to enhance the production of storage lipids in plant vegetative tissues whilst minimizing negative effects on plant growth and development.

Triacylglycerols (TAGs) are storage lipids commonly found in plant seeds. As one of the most energy-rich compounds found in nature, TAGs have become an important target for renewable biofuel feedstocks. Seed-based TAGs are mostly dedicated for food and animal feed uses. Vegetative biomass, because of its high capacity for fatty acid (FA) synthesis, represents a potential renewable, sustainable, and economical platform for TAG accumulation to offset some of the increasing demand for fossil oil. While TAGs are barely detectable in plant vegetative tissues, bioengineering strategies have been developed to enhance the accumulation of storage lipids by directing carbon flux toward lipid synthesis. However, the high level of vegetative TAG accumulation in bioengineered crops is often associated with plant growth deficits (Zale et al., 2016). Previous studies identified a purple acid phosphatase (PAP2) that could increase plant growth rate, vegetative biomass, and seed yield when overexpressed in plants (Sun et al., 2012). PAP2 targets both chloroplasts and mitochondria and has been shown to elevate adenosine triphosphate (ATP) content and photosynthesis rate, most likely by dephosphorylating proteins thereby facilitating their transport into chloroplasts and mitochondria. Here, we analyzed the influence of *PAP2* overexpression on lipid metabolism and designed strategies to enhance TAG accumulation in a fast-growing and high-biomass *PAP2*-overexpression line. Lipid analysis revealed that overexpression of *PAP2* in Arabidopsis increased FA synthesis rates in siliques, leading to elevated seed oil content. On the other hand, the levels of total lipids and TAG in leaves were not altered by *PAP2* overexpression, as both FA synthesis and turnover rates were increased in leaves overexpressing *PAP2*. To enhance the accumulation of TAG in vegetative tissues of the *PAP2*-overexpression line, we crossed it with a TAG lipase mutant (*sdp1*) and a phospholipid:diacylglycerol acyltransferase (PDAT1) overexpression line to impede lipid turnover and convert membrane lipids into TAG. Our results show that these combinations produced significantly higher amounts of TAG in vegetative tissues relative to the parental lines, suggesting that PAP2 can be successfully stacked with other lipogenic factors to enhance TAG accumulation in vegetative tissues. Therefore, our work provides novel strategies to engineer storage lipids in plant vegetative tissues and potentially offsets yield drag associated with high levels of vegetative TAG accumulation.

References

1. Sun, F., Suen, P.K., Zhang, Y., Liang, C., Carrie, C., Whelan, J., Ward, J.L., Hawkins, N.D., Jiang, L., and Lim, B.L. (2012) A dual-targeted purple acid phosphatase in Arabidopsis thaliana moderates carbon metabolism and its overexpression leads to faster plant growth and higher seed yield. *New Phytologist* 194: 206-219.

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 Biotechnology Journal* 14: 661-669.

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.

Biorefinery of Lignocellulosic Carbohydrates: Production of Lipid and Ethanol through Engineered Microbial Conversion

Ming-Hsun Cheng[1,2*](minghsun@illinois.edu), Bruce S. Dien,[2,3] Yong-Su Jin,[1,2] Jong-Hyeok Shin,[1] Patricia J. Slininger,[2,3] Nasib Qureshi,[3] and **Vijay Singh[1,2]**

[1]University of Illinois Urbana-Champaign, Urbana; [2]Center for Advanced Bioenergy and Bioproducts Innovation (CABBI); [3]Bioenergy Research USDA ARS, Peoria, IL

https://cabbi.bio/

Project Goals: CABBI intends to develop efficient ways to grow bioenergy crops, innovate green bioprocessing technologies to transform biomass into valuable platform chemicals including lipids, sugars and alcohols, and market the resulting biofuels and other bioproducts.

Lignocellulose is regarded as renewable biomass for the production of biofuels and biochemicals. Glucose and xylose are the major sugars which can be recovered from cellulose and hemicellulose by hydrolysis. These cellulosic sugars can be used to produce biofuels and value-added bioproducts, such as lipids and ethanol. Bioenergy sorghum has been seen as a potential energy crop due to its high biomass productivity, drought tolerance, and wide environmental adaptability (Rooney at al., 2007, Cheng et al., 2019). In this study, lipids and ethanol were produced from bioenergy sorghum syrups using engineered yeasts. Bioenergy sorghum was hydrothermally pretreated at 50% solids loading in a continuous reactor system. Pretreated biomass was mechanically refined using a burr mill to further improve biomass accessibility for hydrolysis. Fed-batch enzymatic hydrolysis was conducted with 50% w/v solids loading to achieve 230 g/L sugar concentration. Different strains of *Rhodosporidium toruloides* were used to ferment sugars into lipids, and the highest lipid yield of 9.2 g/L was observed. The lipid yield was improved to 17.8 g/L by shifting the C/N ratio from 80 to 1700 to trigger the lipid production in the second-stage fermentation. For ethanol production, the engineered *Saccharomyces cerevisiae* SR8 δ-ADH6 was utilized to co-ferment glucose and xylose. Additionally, the effects of nutrient media (YP, YNB/Urea, and Urea), cellulosic sugar concentration, and sulfite addition were investigated to optimize the ethanol yield from sorghum syrups. The sugar concentration significantly affected the ethanol yield. The optimal ethanol yield at 73.3% was obtained from the fermentation of YNB/Urea broth consisting of 34 g glucose/L and 17 g xylose/L.

References:

1, Rooney, W.L., Blumenthal, J., Bean, B., Mullet, J.E. (2007) Designing sorghum as a dedicated bioenergy feedstock. *Biofuels, Biproducts & Biorefining* 1: 147-157.
2, Cheng, M.-H., Dien, B.S., Lee, D.K., Singh, V. (2019) Sugar production from bioenergy sorghum by using pilot scale continuous hydrothermal pretreatment combined with disk refining. *Bioresource Technology* 289: 121663.

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.*

Deep Learning of Transcriptional Regulation in *Issatchenkia orientalis*

Payam Dibaeinia[1] (dibaein2@illinois.edu), Shayan Tabe Bordbar[1], Shounak Bhogale[2], Veronika Dubinkina[3], **Sergei Maslov[3],** and **Saurabh Sinha[1,3]**

[1]Department of Computer Science, University of Illinois Urbana-Champaign, Urbana; [2]Center for Biophysics and Quantitative Biology, University of Illinois Urbana-Champaign, Urbana; [3]Department of Bioengineering and Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana

https://cabbi.bio/research/conversion-theme/

Project Goals: The project goal was to build a computational modeling tool to better understand transcriptional regulation by analyzing sequence and expression data. This tool can be used to identify key regulators of transcriptomic differences found in yeast strains that are tolerant to low pH, thus providing mechanistic insights that can be used to engineer better strains.

Understanding the "grammar" of transcriptional regulation can help us interpret the different patterns of gene expression in various experimental conditions and predict gene expression in untested conditions. Several computational models of gene expression have been developed in the past that aimed at modeling the expression driven by a promoter as a function of the sequence (*cis* elements) and the concentrations of transcription factors (TF) that bind to the promoter regions (*trans* elements). Here, we developed a Convolutional Neural Network-based Sequence-to-Expression Prediction Tool (CoNSEPT). CoNSEPT extends previous transcriptional regulation models by capturing a wide variety of non-linear plausible mechanisms that might underlie the gene regulation. A trained CoNSEPT model can be used to identify the TFs' roles and distance-dependent interactions between TFs, and to guide the future experimental design by predicting optimal *trans* conditions for a particular desired transcriptional profile. Here, we used CoNSEPT to decipher the transcriptional regulation in *Issatchenkia orientalis* (Io129 strain) using the mRNA expression profile of about 5,000 genes in 23 conditions with 22 hypothetical TFs. We showed that CoNSEPT yields a better fit to these data than a baseline linear model. Our future studies will include extending the model by employing a larger set of TFs and exploiting the trained model for suggesting new experimental conditions.

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.*

Discovering Transcription Regulation Networks in Bioenergy-relevant Yeast Species

Veronika Dubinkina[1,2*] (vd6@illinois.edu), Ping-Hung Hsieh,[1,3] **Yasuo Yoshikuni,**[1,3] **Sergei Maslov,**[1,2] and **Huimin Zhao**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI), Urbana, IL; [2]Department of Bioengineering and Carl Woese Institute of Genomic Biology (IGB), University of Illinois Urbana-Champaign, Urbana; [3]Lawrence Berkeley National Lab, Berkeley, CA

https://cabbi.bio

Project Goals: The goal of this project is to reconstruct Gene Regulatory Networks (GRN) for a set of bioenergy-relevant novel yeasts leveraging existing data on model yeast species. The resulting refined GRNs will be used to develop strategies for improving yield of pathways producing target bioenergy compounds.

Understanding gene regulatory networks (GRN) is critical for metabolic engineering of yeast species. It is known that transcription regulation undergoes rapid evolutionary rewiring leading to substantial divergence of regulatory networks across species. Thus, for poorly understood, non-model organisms, such as *Issatchenkia orientalis, Yarrowia lipolytica,* and *Rhodosporidium toruloides*, GRNs remain mostly unknown.

To facilitate the reconstruction of full GRNs for yeasts of interest we first identified all putative transcription factors (TFs) in their genomes. To do that we compiled a database of TF-specific protein domains from two different databases (DBD: Transcription factor prediction database [1] and Fungal Transcription Factor Database [2]) and performed a protein family domain search to identify all putative TF genes in new strains. We estimated that there are 306 putative TF in the *I. orientalis* genome, 336 in *Y. lipolytica*, and 398 in *R. toruloides*.

Since many transcription factor binding motifs (TFBS) and DNA-binding domains are known to be conserved, we used them to predict GRNs of novel species based on the existing data for the model yeast species *Saccharomyces cerevisiae*. We built a two-step bioinformatics pipeline for genome sequence analysis that transfers information about experimentally validated binding sites of TFs in *S. cerevisiae* to the species of interest. We used OrthoFinder along with the recently published collection of 332 yeast genomes [3] to identify putative TFs which are orthologous to some previously studied in *S. cerevisiae*. We then retrieved a set of motifs for these TFs in *S. cerevisiae* from the YEASTRACT+ database [4]. A MEME motif scan was used to identify putative targets of these TFs based on motif presence in the gene promoters of species of interest (see Table 1 for the summary of reconstructed GRNs). Notably the number of orthologous target genes for these TFs which were conserved between species is relatively small. We also analyzed existing RNA-seq data for *I. orientalis* in 12 different media conditions and partially confirmed our predicted GRNs.

We plan to collect more RNA-seq data which will help us to validate and curate reconstructed GRNs. Although we do not have motif binding information for most of the newly predicted TFs, it can be partially inferred by de novo motif discovery using clusters of co-expressed genes in

RNA-seq experiments and can be further confirmed by DAP-seq experiments. Overall, resulting GRNs will be incorporated into the integrated model of metabolic networks of novel yeasts and empower CABBI teams to predict genomic modifications that can improve production of target biofuel compounds.

Table 1. General statistics of reconstructed GRNs.

Strain	Total # of genes	# of putative TFs	# of TFs with known motifs	# targets	# links
I. orientalis SD108	4925	309	64	4361	20678
Y. lipolytica W29	7919	336	71	7320	62102
R. toruloides IFO0880	8221	398	50	7099	31295
S. cerevisiae	6725	307	118	6725	201974

References

1. Kummerfeld, Sarah K., and Teichmann, Sarah A. "DBD: a transcription factor prediction database." *Nucleic acids research* 34.suppl_1 (2006): D74-D81.
2. Park, Jongsun, et al. "FTFD: an informatics pipeline supporting phylogenomic analysis of fungal transcription factors." *Bioinformatics* 24.7 (2008): 1024-1025.
3. Shen, Xing-Xing, et al. "Tempo and mode of genome evolution in the budding yeast subphylum." *Cell* 175.6 (2018): 1533-1545.
4. Monteiro, Pedro T., et al. "YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts." *Nucleic Acids Research* 48.D1 (2020): D642-D649.

***Funding statement**: This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innova on (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 recommenda ons expressed in this publica on are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*

Towards an Efficient Multiallelic Gene Editing Platform for Highly Polyploid Sugarcane

Ayman Eid[1,2*] (amohamedeid@ufl.edu), Chakravarthi Mohan,[2] Sara Sanchez,[1,2] Duoduo Wang,[1,2] and **Fredy Altpeter[1,2]**

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation, Gainesville, FL; and [2]University of Florida, Gainesville, FL

https://cabbi.bio/research/feedstocks-theme/

Project Goals

The goal of this project is to develop strategies to achieve high levels of multiallelic targeted mutagenesis in sugarcane using CRISPR/Cas9 and other RNA guided nucleases. An efficient genome editing platform for sugarcane will support metabolic engineering approaches to convert sugarcane into oilcane. Oilcane is a genetically modified sugarcane that hyperaccumulates lipids in its vegetative biomass (Zale et al. 2016; Parajuli et al. 2020). This project goal supports 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

Site-specific nucleases (SSNs) such as CRISPR/Cas9 have revolutionized crop improvement by allowing targeted and precise gene editing. Targeted mutagenesis by CRISPR/Cas9 relies on DNA targeting by a single guide RNA molecule (sgRNA) and DNA cleavage by Cas9 nuclease. This ribonucleoprotein (RNP) complex identifies the target DNA through complementarity to the sgRNA molecule only when located upstream of a protospacer adjacent motif (PAM) that is recognized by the Cas9 moiety. Cas9 has demonstrated high gene editing levels at a canonical PAM sequence of NGG. Alternative RNA-guided nucleases have different or less stringent PAM site limitations. Reducing the PAM site limitations of Cas9 will increase the targeting possibilities, which is particularly important for complex, highly polyploid genomes like the sugarcane genome (2n=10–13x=100–130). For instance, the Cas9 variant NG (NgCas9) has been engineered to require fewer contacts with the PAM sequence and use NG instead of NGG. Following the cleavage, the DNA double-strand breaks (DSBs) are repaired by cellular repair pathways including Non-Homologues End Joining (NHEJ) or Homology Directed Repair (HDR). NHEJ is an error-prone pathway that frequently results in indels, allowing us to construct knockouts with loss of gene function. In contrast, HDR depends on a repair template which can be co-delivered with editing reagents to introduce targeted nucleotide substitutions for gain of function.

Construction of loss-of-function phenotypes by targeted mutagenesis is more challenging in highly polyploid crops than in diploid crops. The large number of homeologs and homologs in sugarcane causes functional redundancy. However, this also offers the opportunity to generate a range of phenotypes depending on the number of co-mutated copies/alleles, similar to RNAi. Alternative approaches are needed to optimize and reevaluate gene editing reagents or their delivery to address the genetic complexity and redundancy in sugarcane. Therefore, we developed an approach that allows the generation of a rapidly scorable phenotype following multiallelic mutagenesis of magnesium chelatase (MgCh), a key gene in chlorophyll

biosynthesis. Through Sanger sequencing, we were able to detect more than 50 copies of MgCh in sugarcane. The disruption of the majority of copies of this gene by CRISPR/Cas9-mediated targeted mutagenesis resulted in yellow plants that display severe depletion of green pigment already when regenerating from tissue culture. In contrast to phytoene desaturase gene (PDS) mutants which display a dwarf and albino phenotype, MgCh mutants show yellow leaf color and no obvious growth retardation. This facilitates sampling of tissues for molecular analysis and allows us to exclude analysis of somaclonal variants that may resemble the albino phenotype of PDS mutants. Sequencing analysis of yellow MgCh mutants revealed co-editing frequencies of more than 70% of the MgCh copies/alleles. Differences in different editing outcomes and phenotypes using two sgRNAs and different tissue culture treatments will be discussed.

This efficient approach to monitor events with multiallelic co-editing will accelerate further optimizations to overcome PAM site limitations and improve delivery of reagents. We are currently using this approach to compare gene-editing efficiency of NgCas9 and Cas9 using the same sgRNAs. Efficient genome editing tools will support ongoing metabolic engineering approaches to convert sugarcane into oilcane with hyperaccumulation of lipids in its vegetative biomass.

References

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

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

Funding Statement

This research 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.

Nitrogen Delivery Efficiency in the Mississippi River Basin

Theodore Hartman[1,2*](theodore@iastate.edu), Yuanyao Lee,[1,3] Madhu Khanna,[1,3] and **Andy VanLoocke**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Department of Agronomy, Iowa State University, Ames; [3]Department of Agricultural & Consumer Economics, University of Illinois Urbana-Champaign, Urbana.

https://cabbi.bio/research/sustainability-theme/

Project Goals: This research aims to calculate and analyze the spatial distribution of the nitrogen delivery efficiency for crop reporting districts within the Mississippi River Basin using the hydrology model THMB.

Nitrogen leaching in the Mississippi River Basin (MARB) is a primary driver of nitrogen (N) inputs to the Gulf of Mexico. However, while it is assumed that not all nitrogen leached from areas in the MARB makes it to the Gulf of Mexico due to in-transit losses like denitrification, the magnitude and spatial distribution of these losses are not well understood. In this work, we use the Terrestrial Hydrology Model with Biogeochemistry (THMB) to quantify how much nitrogen leached from individual crop reporting districts (CRDs) across the MARB makes it to the Gulf of Mexico. We defined this nitrogen delivery efficiency (DE) as the ratio of how much nitrogen was delivered to the Gulf of Mexico to how much was leached from a particular CRD. To track the nitrate leaching from individual CRDs, both the leaching and delivered nitrogen were normalized using values from a baseline scenario. Results from this study indicate that the average delivery efficiency ranged from 10% to 47% across the MARB (Fig. 1). Areas closer to streams and rivers had higher DE than those further away. Model sensitivity analysis indicated that DE values do not depend on the level of nitrate leaching from individual CRDs. The magnitude and spatial distribution of this DE value is important because previous work has shown that meeting the renewable fuels mandates using traditional economic drivers without considering nitrate leaching values will not necessarily result in water quality improvements. In combination with nitrate leaching output from an ecosystem model, these DE values provide a way for economic models to use estimates of the nitrogen delivered to the Gulf of Mexico in the determination of the mixture and spatial distribution of bioenergy crops needed to meet the RFS and achieve economic and ecological sustainability goals.

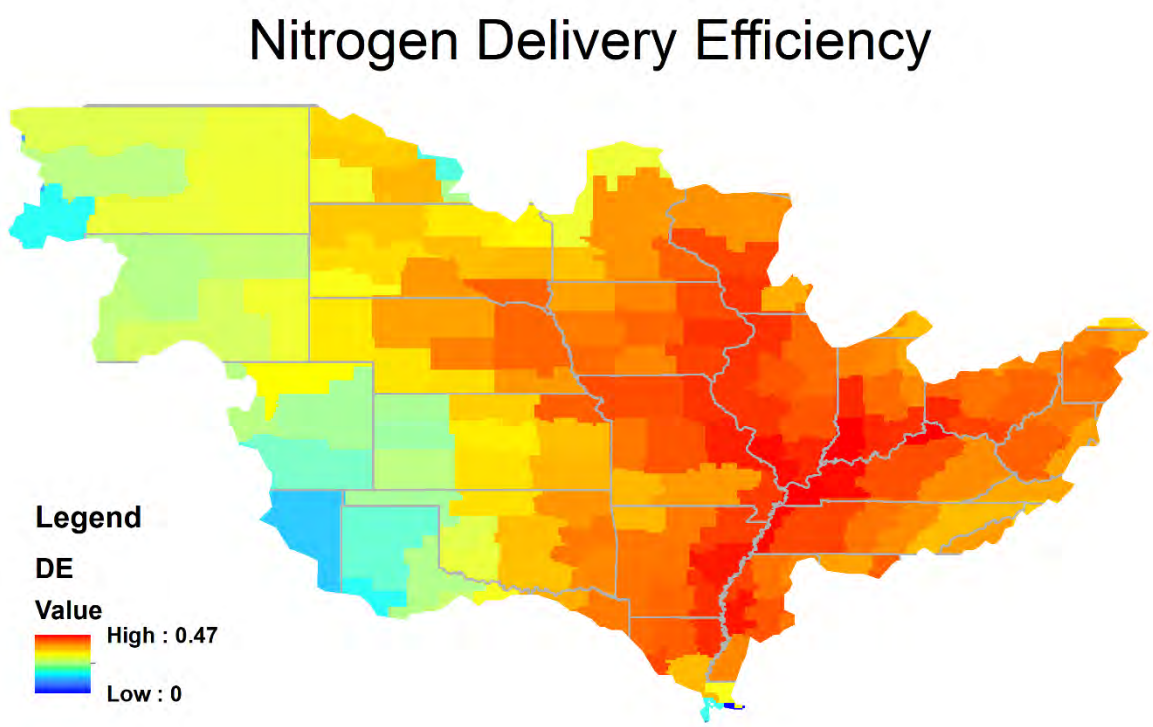


Figure 1: 30-year average delivery efficiency (DE) across the MARB

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy*

L-malic Acid Production from Xylose by Engineered *Saccharomyces cerevisiae*

Nam Kyu Kang[1,2*] (nkkang@illinois.edu), Jae Won Lee,[2,3] Donald R. Ort,[1,2,4] and **Yong-Su Jin[1,2,3]**

[1]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana; [2]DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI); [3]Department of Food Science and Human Nutrition, University of Illinois Urbana-Champaign, Urbana; [4]Department of Plant Biology and Department of Crop Sciences, University of Illinois Urbana-Champaign, Urbana

https://cabbi.bio/research/conversion-theme/

Project Goals: As *Saccharomyces cerevisiae*, a Crabtree-positive yeast, produces ethanol exclusively during glucose fermentation, the reduction of ethanol without growth defects is necessary to efficiently produce value-added products. In this study, we developed a metabolic engineering strategy which enabled a high titer production of malic acid from xylose by engineered *S. cerevisiae*.

With increasing environmental concerns and decreasing oil supply, there are growing interests in developing technologies that utilize renewable sources for the production of fuels and chemicals. Microbial conversion of lignocellulosic biomass into biofuels and chemicals can be a substitute for petroleum-based industry [1]. As hydrolysates of lignocellulosic biomass mainly contain glucose and xylose, it is necessary to develop a xylose-utilizing microorganism [2]. *Saccharomyces cerevisiae* is one of the most promising microbial strains for bioconversion of lignocellulosic biomass because there have been many efforts to make *S. cerevisiae* consume xylose by introducing the xylose isomerase and oxidoreductive pathways.

L-malic acid is widely used in the food and chemical industries. Here, we report on production of malic acid from xylose by engineered *S. cerevisiae*. To enable malic acid production in a xylose-assimilating *S. cerevisiae* with the oxidoreductase pathway, we employed the cytosolic reductive TCA (rTCA) pathway. We overexpressed *PYC1* and *PYC2,* coding for pyruvate carboxylases, a truncated *MDH3,* coding for malate dehydrogenase, and *SpMAE1,* coding for a *Schizosaccharomyces pombe* malate transporter. Additionally, *GPD1* and *GPD2,* coding for glyceraldehyde-3-phosphate dehydrogenase, were deleted to completely block the glycerol production pathway. The metabolic pathway responsible for ethanol production was partially blocked through deleting *PDC1* and *ADH1*, because complete deletion of the ethanol pathway could lead to severe growth defects due to the limited synthesis of acetyl-CoA, an important precursor of cell growth [3-5]. The resulting strain produced malic acid from both glucose and xylose, but it produced much higher titers from xylose. Interestingly, the engineered strain had higher malic acid yield from lower xylose concentrations (10 g/L), with no ethanol production, than from higher xylose concentrations (20 g/L and 40 g/L). As such, a fed-batch culture maintaining xylose concentrations below 10 g/L was conducted, and 61.2 g/L of malic acid was produced with a productivity of 0.32 g/L·h.

These results represent successful engineering of *S. cerevisiae* for the production of malic acid from xylose and therefore confirm that xylose offers the efficient production of various biofuels and chemicals by engineered *S. cerevisiae*.

References

1. Chu, S., andMajumdar, A., *Nature,* 2012. **488**(7411): p. 294-303.
2. Carroll, A. and Somerville, C., *Annual Review of Plant Biology*, 2009. **60**(1): p. 165-182.
3. Flikweert, M.T., et al., FEMS *Microbiology Letters*, 1999. **174**(1): p. 73-79.
4. Lee, Y.-G. and Seo, J.-H., *Biotechnology for Biofuels*, 2019. **12**(1): p. 204.
5. Tokuhiro, K., et al., *Applied Microbiology and Technology*, 2009. **82**(5): p. 883-90.

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.*

Testing Unifying Theories of Ozone Response in C_4 Bioenergy Grasses

Shuai Li[1,2,4*] (shuaili@illinois.edu), Christopher A. Moller,[3,4] Noah G. Mitchell,[3,4] Erik J Sacks,[1,4] DoKyoung Lee,[1,4] and **Elizabeth A. Ainsworth**[1,2,3,4]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana; [3]Global Change and Photosynthesis Research Unit, USDA ARS, Urbana, IL; and [4]University of Illinois Urbana-Champaign, Urbana

https://cabbi.bio/

Project Goals: The objectives of this study are to: i) quantify genotypic and species variation in O_3 sensitivity among C_4 bioenergy crop species, and ii) identify the factors underlying interspecific variability in sensitivity to O_3.

Abstract: Tropospheric ozone (O_3) is a damaging and widespread air pollutant that is detrimental to human and ecosystem health worldwide[1,2]. Current background O_3 concentration is estimated to decrease the yields of maize (*Zea mays*) up to 10% in the United States and reduce global crop yield (maize, soybean, rice and wheat combined) by 227 Tg annually, which portends a significant threat to future global food and energy security[3,4]. However, it is unclear how other bioenergy feedstocks, including switchgrass (*Panicum virgatum*)[5], sorghum (*Sorghum bicolor*)[6], and miscanthus (*Miscanthus* × *giganteus*), respond to O_3 stress, or whether these species share a similar O_3 sensitivity as maize.

Tropospheric O_3 negatively impacts plant growth and development at various biological processes from molecular to whole plant scales[1,2]. Although plant response to O_3 is complex, it is well-recognized that some plant species are more sensitive to O_3 than others. There are several mechanisms that potentially explain genotypic and/or species variability in O_3 sensitivity in C_3 species. For example, species-specific variation sensitivity to O_3 in C_3 tree species depends upon stomatal flux of ozone per unit leaf area, so both stomatal conductance and leaf mass per unit area (LMA) are important traits determining sensitivity[7,8]. However, there is currently a lack of unifying mechanistic explanation for variation among C_4 plants in response to O_3. Considering that leaf Kranz anatomy enables the concentration of CO_2 around Rubisco in the bundle sheath cells, limited photorespiration, and lower stomatal conductance, C_4 species may respond differently to O_3 than C_3 species.

In this study, we examined the photosynthetic response of 22 genotypes of four C_4 bioenergy species (switchgrass, sorghum, maize and miscanthus) to elevated O_3 using the unique capabilities of Free Air Concentration Enrichment (FACE) technology, which provides elevated concentrations of O_3 (100 nL L^{-1}) in open-air plots at the field scale. Because different species display different visual symptoms of O_3 damage, we proposed the reductions of photosynthetic traits (the maximum carboxylation capacity of phosphoenolpyruvate (V_{pmax}) and the maximum CO_2-saturated photosynthetic capacity (V_{max})) to O_3 as a proxy to estimate O_3 sensitivity. The

studied species displayed strong variability in V_{pmax} and V_{max} within each species. Across all species, V_{pmax} and V_{max} varied 3.8- and 2.1-fold, respectively. Elevated O_3 concentration did not alter V_{pmax} in any genotypes of switchgrass and miscanthus and reduced V_{max} in maize lines, indicating variation among C_4 species in O_3 sensitivity. O_3-induced reduction in V_{pmax} was positively associated with LMA, but negatively correlated with stomatal conductance on either an area (g_{sa}) or a mass (g_{sm}) basis. However, O_3-induced reduction in V_{max} was not correlated with LMA, but scaled negatively with g_{sa} and g_{sm}. Structural equation models provided further evidence that both V_{pmax} and V_{max} were directly related to stomatal conductance rather than to LMA. We demonstrate that there is significant variation in O_3 sensitivity among C_4 species, with maize and sorghum more sensitive to O_3 than switchgrass and miscanthus. We also demonstrate genotypes with higher stomatal conductance were more sensitive to O_3 compared to genotypes with lower stomatal conductance, and interspecific variation in O_3 sensitivity is determined by direct effects of stomatal conductance and indirect effects of LMA. Such a side-by-side field comparison study has not been conducted so far, and to our knowledge, this is the first study to provide a test of unifying theories explaining variation in O_3 sensitivity in C_4 bioenergy grasses. This information could aid in optimal placement of diverse C_4 bioenergy feedstocks across a polluted landscape.

References:

1. Ainsworth, E.A. (2017). Understanding and improving global crop response to ozone pollution. *Plant Journal* 90: 886-897.
2. Ainsworth, E.A. *et al*. (2012) The effects of tropospheric ozone on net primary productivity and implications for climate change. *Annual Review of Plant Biology* 63: 637-661.
3. McGrath, J.M. *et al.* (2015) An analysis of ozone damage to historical maize and soybean yields in the United States. *Proceedings of the National Academy of Sciences of the United States of America* 112: 14390–14295.
4. Mills, G. *et al.* (2018). Closing the global ozone yield gas: Quantification and co-benefits for multi-stress tolerance. *Global Change Biology* 24: 4869–4893.
5. Li, S. *et al.* (2019) Elevated ozone concentration reduces photosynthetic carbon gain but does not alter leaf structural traits, nutrient composition or biomass in switchgrass. *Plants* 8: 85.
6. Li, S. *et al.* (2021). Bioenergy sorghum maintains photosynthetic capacity in elevated ozone concentrations. *Plant, Cell & Environment*, in press.
7. Reich, P.B. (1987) Quantifying plant response to ozone: a unifying theory. *Tree Physiology*, 3: 63-91.
8. Feng, Z. *et al*. (2018) A unifying explanation for variation in ozone sensitivity among woody plants. *Global Change Biology* 24: 78-84.

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.*

Optimizing Measurement Methods for N_2 Fixation in *Miscanthus* × *giganteus*

Di Liang[*1,2] (diliang2@illinois.edu), Niuniu Ji[1,2], Danyang Duan[1,5], Sierra Raglin[1,5], Alonso Favela[1,6], Isaac Klimasmith[1,5], Rachel Waltermire[1,2], Sandra Simon[1,2], **Wendy Yang**[1,3,4] and **Angela Kent**[1,5]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Institute for Sustainability, Energy and Environment, University of Illinois Urbana-Champaign, Urbana; [3]Department of Plant Biology, University of Illinois Urbana-Champaign, Urbana; [4]Department of Geology, University of Illinois Urbana-Champaign, Urbana; [5]Department of Natural Resources and Environmental Sciences, University of Illinois Urbana-Champaign, Urbana; [6]Program in Ecology, Evolution, and Conservation Biology, University of Illinois Urbana-Champaign, Urbana

https://cabbi.bio/

Project Goals:

The overall goal of this project is to understand the importance of associative N_2 fixation, a microbial process that converts atmospheric N_2 into NH_4^+, in supporting miscanthus productivity. Miscanthus (*Miscanthus* × *giganteus*) is considered an ideal bioenergy crop because of its high yield versus low energy inputs. Many studies have reported high N use efficiency associated with miscanthus (Cadoux et al., 2012), with low or no N fertilization effects observed (Christian et al., 2008). Further, although associative N_2 fixation has been observed in miscanthus (Keymer & Kent, 2014), the contribution of N_2 fixation to the miscanthus N budget at the ecosystem level is still unknown. To determine if N_2 fixation could be a substantial source of N during miscanthus development, we conducted a year-long field study to investigate the "hotspots" and "hot moments" of N_2 fixation. Our results will help to advance the understanding of environmental sustainability and N economy of miscanthus.

Abstract:

Understanding the potential contribution of N_2 fixation to available N for miscanthus requires reliable methods of estimating N_2 fixation rates. Currently, the acetylene reduction assay (ARA) and $^{15}N_2$ incorporation method are commonly used (Smercina et al., 2019). ARA depends on nitrogenase, the enzyme involved in N_2 fixation, to break the triple bond of acetylene instead of N_2, such that ethylene could be measured by a gas chromatograph (GC) with a flame ionization detector (FID) (Hardy et al., 1968). In comparison, the $^{15}N_2$ incorporation method is based on the differences of ^{15}N concentrations in samples that are subjected to either ^{15}N-labeled or ^{15}N natural abundance reference gas during lab incubation, such that N_2 fixation rates can be calculated directly (Gupta et al., 2014). Although both ARA and $^{15}N_2$ incorporation have

their own advantages and disadvantages, it is still unknown which method works best for measuring N_2 fixation in bioenergy crops.

Existing studies on miscanthus have mostly focused on measuring N_2 fixation using only one aforementioned method (Davis et al., 2010). The correlations between ARA and $^{15}N_2$ incorporation, also known as the conversion factors, are poorly understood, especially among different miscanthus tissues. To address this knowledge gap, we used both ARA and the $^{15}N_2$ incorporation method to measure N_2 fixation in leaves, stems, rhizomes, roots, bulk soils, and rhizosphere soils of mature miscanthus grown on marginal soil. Results from both methods confirmed that rhizosphere soils had the highest N_2 fixation rates, followed by roots and bulk soils. In comparison, the aboveground miscanthus tissues exhibited little to no N_2 fixation capacities. Additionally, we found significantly different conversion factors among miscanthus tissues and soils.

References

Cadoux, S., Riche, A.B., Yates, N.E., & Machet, J.M. (2012). Nutrient requirements of *Miscanthus x giganteus*: Conclusions from a review of published studies. *Biomass and Bioenergy, 38*, 14-22. doi:https://doi.org/10.1016/j.biombioe.2011.01.015

Christian, D. G., Riche, A. B., & Yates, N. E. (2008). Growth, yield and mineral content of *Miscanthus×giganteus* grown as a biofuel for 14 successive harvests. *Industrial Crops and Products, 28*(3), 320-327. doi:10.1016/j.indcrop.2008.02.009

Davis, S.C., Parton, W.J., Dohleman, F.G., Smith, C.M., Grosso, S.D., Kent, A.D., & DeLucia, E.H. (2010). Comparative biogeochemical cycles of bioenergy crops reveal nitrogen-fixation and low greenhouse gas emissions in a *Miscanthus × giganteus* agro-ecosystem. *Ecosystems, 13*(1), 144-156. doi:10.1007/s10021-009-9306-9

Gupta, V.V.S.R., Kroker, S.J., Hicks, M., Davoren, C.W., Descheemaeker, K., & Llewellyn, R. (2014). Nitrogen cycling in summer active perennial grass systems in South Australia: non-symbiotic nitrogen fixation. *Crop and Pasture Science, 65*(10). doi:10.1071/cp14109

Hardy, R.W.F., Holsten, R.D., Jackson, E.K., & Burns, R.C. (1968). The Acetylene-Ethylene Assay for N_2 Fixation: Laboratory and Field Evaluation. *Plant Physiology, 43*(8), 1185-1207. doi:10.1104/pp.43.8.1185

Keymer, D.P., & Kent, A.D. (2014). Contribution of nitrogen fixation to first year *Miscanthus × giganteus*. *GCB Bioenergy, 6*(5), 577-586. doi:10.1111/gcbb.12095

Smercina, D.N., Evans, S.E., Friesen, M.L., & Tiemann, L.K. (2019). Optimization of the $^{15}N_2$ incorporation and acetylene reduction methods for free-living nitrogen fixation. *Plant and Soil, 445*(1-2), 595-611. doi:10.1007/s11104-019-04307-3

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.

Engineering of the Non-Model Yeast *Issatchenkia orientalis* to Produce Organic Acids

Teresa Martin[*1] (fraterma@illinois.edu), Zia Fatma,[1*] Vinh Tran,[1,2*] and **Huimin Zhao[1,2,3]**

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI), Urbana, IL; [2]Department of Chemical and Biomolecular Engineering, University of Illinois Urbana-Champaign, Urbana; [3]Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois Urbana-Champaign, Urbana

https://cabbi.bio/

Project Goals: The goal of this research is to engineer a non-model yeast, *Issatchenkia orientalis,* to produce organic acids in high titers. 3-Hydroxypropionic acid is one of the key sugar-based building block chemicals that can be produced in biorefineries. The ability to produce 3-hydroxypropionic acid from glucose or other renewable carbon sources would provide a biosustainable alternative to acrylic acid production from fossil resources.

The nonconventional yeast *Issatchenkia orientalis* is a potential platform microorganism for production of organic acids thanks to its unusual ability to grow in highly acidic conditions. However, the lack of efficient genetic tools, including a stable episomal plasmid and precise genome editing tool, prevented metabolic engineering of this organism. We previously developed this genetic toolbox to efficiently engineer this non-model yeast to produce value-added chemicals.[1,2] Here we present the production of 3-hydroxypropionic acid (3HP) from *I. orientalis.* There are multiple pathways to produce 3HP; however, only two are generally used, the malonyl Co-A pathway and the β-alanine pathway. We have generated both pathways on a plasmid to determine preliminary production quantities of 3HP in shake flasks as well as the dependence of pH on the production. Our highest production of 3HP thus far is ~1.5 g/L from 50g/L glucose as the carbon source over 5 days fermentation from the β-alanine pathway. The malonyl Co-A pathway yielded only ~700 mg/L. Further experiments will be done utilizing the β-alanine pathway to further improve production by improving the expression of the beta alanine pyruvate amino transferase as well as genome engineering of the base strain.

References

1. Tran, V., Cao, M., Fatma, Z., Song, X., Zhao, H. *mSphere*, 2019, e00345-19.
2. Cao, M., Fatma, Z., Song, X., Hsieh, P.H., Tran, V., Lyon, W.L., Sayadi, M., Shao, Z., Yoshikuni, Y., Zhao, H. *Metabolic Engineering*, 2020, 59:87-97.

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.*

The Genome of the Perennial Biomass Grass *Miscanthus sinensis*

Therese Mitros[1,4*] (tmitros@berkeley.edu), Kankshita Swaminathan,[2,4] Brandon T. James,[2,4] Mohammad B. Belaffif, [2,4] Adam M. Session,[1] Guohong Albert Wu,[3] and **Daniel S. Rokhsar**[1,3,4]

[1]University of California, Berkeley; [2]HudsonAlpha Biotechnology Institute, Huntsville, AL; [3]Joint Genome Institute, Berkeley, CA; and [4]DOE Center for Advanced Bioenergy and Bioproducts Innovation
https://cabbi.bio/research/feedstocks-theme/

Project Goals: Despite C4 perennial crops having tremendous potential for maximizing agricultural yields and minimizing environmental impacts, genetic improvement is limited due to the complexity of their genomes. The objectives of this study are to:

1. **Provide a chromosome-scale reference and annotation of the *Miscanthus* genome**
2. **Conduct transcriptomic profiling to find tissue-specific genes and nitrogen remobilization pathways**
3. **Population sampling to track admixtures and hybridizations, and inform future breeding efforts**
4. **Continue to improve genomic resources, as well as develop methods and datasets to understand the biology of the grass stem and its modifications, for the bioenergy grasses of interest, including miscanthus, sorghum, and Saccharum**

Miscanthus is a perennial wild grass of global importance for papermaking, roofing, decorative plantings, and as an emerging highly productive temperate biomass crop. We published a chromosome-scale assembly of the *Miscanthus sinensis* genome, providing a resource for *Miscanthus* that links its chromosomes to the related diploid *Sorghum* and complex polyploid sugarcanes[1]. We establish the paleo-allotetraploid nature of miscanthus by the identification of sub-genome-specific repeats. Analysis of *M. sinensis* and *M. sacchariflorus* populations demonstrates extensive interspecific admixture and hybridization, and documents the origin of the highly productive triploid bioenergy crop *M.* × *giganteus*. Transcriptional profiling of leaves, stems, and rhizomes over multiple seasons provides insight into rhizome development and nutrient recycling, processes critical for sustainable biomass accumulation in a perennial temperate grass. An improved PacBio HiFi assembly provides a more contiguous assembly with better-resolved repeat content.

References

Mitros, T., Session, A.M., James, B.T. *et al.* Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nat Commun* **11,** 5442 (2020). https://doi.org/10.1038/s41467-020-18923-6

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.

Genetic Designs Targeting Accumulation of Vegetative Lipids, Multiplexing of Editing Reagents for "In Context Promoter Bashing" and Gaining Insight on Roots Exudates Impact on Shaping the Root Microbiome in Sorghum (*Sorghum bicolor* (L.) Moench

Truyen Quach[1,2*] (tquach2@unl.edu), Ki Youl Park,[1,2] Timothy J. Nicodemus,[1] Chinedu Nwafor,[1] Shirley Sato,[1,2] Zhengxiang Ge,[1,2] Tieling Zhang,[1,2] Tara Nazarenus,[1,2] Ming Guo,[1] James Schnable,[1] Bin Yu,[1] David Holding,[1] **Kankshita Swaminathan,[2,3] Edgar B. Cahoon,[1,2]** and **Tom Elmo Clemente[1,2]**

[1]Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln; [2]DOE Center for Advanced Bioenergy and Bioproducts Innovation; and [3]Hudson Alpha Institute for Biotechnology, Huntsville, AL

https://cabbi.bio/research/feedstocks-theme/

Project Goals

The goals of this project are to: introduce novel genetic variation into sorghum as a means to synthesize value-added, lipid-based co-products, to ascertain the role of the sorghum root exudate, sorgoleone, on shaping the microbiome about the rhizosphere and its relationship to inhibition of nitrification in soil, and to test a multi-plex Cas9-based delivery of editing reagents for "in context promoter bashing" as a means of perturbation of expression for specific gene calls associated with the PEPCK C4 photosynthetic pathway.

Abstract

Sorghum (*Sorghum bicolor* (L.) Moench) has high biomass potential, and sweet genotypes can accumulate sugar reserves of up to 15% in stalks. These phenotypic attributes, coupled with its capacity to perform C4 photosynthesis and its relatively low agronomic input requirements, make sorghum an ideal feedstock for the bioeconomy. A transdisciplinary team of researchers has been assembled to take a holistic approach to addressing selected input and output traits to improve the sustainability of this feedstock for the production of vegetative lipid co-products in stalks, to gain insight on the role of root exudates, namely sorgoleone, in shaping the rhizosphere microbiome, and to implement genome editing platforms for "in context promoter bashing" in the crop. A priority output trait being targeted is the shifting of metabolic flux in parenchyma cells of stalks to alter carbon reserves from sucrose to lipids within the cells. To address these target output traits, genetic designs are being assembled employing a modular assembly approach in a build-test-learn process through the introduction of novel genetic variation implementing the tools of synthetic biology, in both grain and sweet genotypes of sorghum, as a means to assess impact on sucrose flux post-flowering on lipid accumulation in stalks. We target three processes of lipid (triacylglycerol) production and storage, regarded as push, pull, and protection, using key variants of enzymes (WRINKLED1, DGAT, and OLEOSIN) regulating these processes. To gain insight on the role of sorgoleone on modulating the root microbiome, sorghum events were generated that significantly reduce the synthesis of sorgoleone, and the putative sorgoleone biosynthetic pathway has been introduced into maize, leading to the synthesis of this novel exudate in corn. These novel sorghum and maize biologicals position our team to address what impact this exudate has on shaping the soil microbiome and its relationship

to biological inhibition of nitrification (BNI). Lastly, we designed and introduced a multi-plex guide approach into sorghum targeting promoter regions of selected gene calls in the PEPCK C4 photosynthesis pathway, referred to as "in context promoter bashing", as a way to alter expression of the respective downstream gene calls. Data gathered from the phenotypic characterizations of these biologicals will form the basis of future genetic designs to improve both quantity and functional quality of vegetative lipid accumulation in sorghum, revealing what, if any, role the sorghum exudate sorgoleone has on the BNI. This information can be exploited as an indirect means to improve nitrogen use efficiency and mitigate nitrogen run-off. Lastly, this program is expanding the genetic toolkit through deliberate edits in regulator regions as a way to perturb gene expression in C4 feedstocks.

References

Dayan, F. E., Rimando, A. M., Pan, Z., Baerson, S. R., Gimsing, A. L., & Duke, S. O. (2010). Sorgoleone. *Phytochemistry*, *71*(10), 1032-1039.

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.

Responsiveness of Miscanthus and Switchgrass Yields to Stand Age and Nitrogen Fertilization: A Meta-regression Analysis

B.P. Sharma[1,2*](bpsharma@illinois.edu), N. Zhang,[1,2] D.K. Lee.[1,2] E.A. Heaton,[1,3] E.H. DeLucia,[1,2] E.J. Sacks,[1,2] I.B. Kantola,[2] N.N. Boersma,[1,3] S.P. Long,[1,2] T.B. Voigt,[1,2] and **M. Khanna[1,2]**

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation, [2]University of Illinois Urbana-Champaign, Urbana; and [3]Iowa State University, Ames

Project Goals

This study examines the productivity response of bioenergy crops to various agronomic, climatic, and management factors. We estimate production functions for miscanthus and switchgrass in generalizing the combined response of N fertilization and stand age on harvestable biomass controlling for soil productivity, climatic factors, and experimental variabilities. One implication of the predicted yield response is to guide the existing biophysical (mechanistic) growth models for improved yield modeling through validation. Furthermore, the yield estimates will serve as a benchmark for regional bioenergy assessment models which heavily rely on relative profitability of these crops for mapping land use change.

https://cabbi.bio/

Abstract

Motivation: Ambitious cellulosic biofuel mandates in the U.S. require continuously updated information on productivity response of dedicated energy crops such as miscanthus and switchgrass to numerous climatic, agronomic and management factors. Research has shown that these crops perform relatively better in marginal growing conditions compared to row crops. Studies have also provided evidence of ecosystem benefits of these crops, including increased soil carbon, reduced N_2O emissions, and reduced NO_3 leaching, compared to row crops. These crops are considered as sustainable energy sources mainly because of their high productivity and nutrient use efficiency, including the ability to recycle nutrients annually. However, depending on soil and climate, input requirements of these crops are expected to vary spatially and temporally. Among other factors, optimal N use is of much significance since excessive application not only leads to economic inefficiency but may become counterproductive by releasing N_2O emissions and NO_3 leaching, which have well known GHG implications. As perennials, these crops have establishment, growth, and senescent phases which all differ in harvestable biomass, which again depend on climatic and management factors. Thus, growth consideration in terms of productivity and optimal rotation have economic significance because of an extended waiting period to start realizing profits. The primary objective of this study is to develop an empirical model and estimate using a suite of meta-analysis econometric techniques to explain the productivity response of perennial energy crops to climatic, agronomic, and management factors including N and stand age, simultaneously capturing the study and location-specific experimental variabilities as well as their differences across harvest years, effectively capturing study and location-specific temporal dependencies in yield measurements.

Data and Methods: The datasets in the analyses include a total of 1,884 yield observations of miscanthus obtained from 27 locations in 14 states, and 2,903 yield observations

of switchgrass obtained from 23 locations in 15 states from the U.S. For each plot-level observation, dry biomass yield (Mg ha^{-1}), location, latitude and longitude, planting year, growing year, plot size (m^2), planting method, planting density (ha^{-1} or kg ha^{-1}), harvest year, N application rate (kg ha^{-1}), and cultivar (switchgrass) are directly extracted from texts and tables of the published articles or datasets obtained through personal communications. The data we have collected has clustered structure with yield measurements taken at various levels, i.e. study, location, and year. Furthermore, each study and location has repeated observations for a specific year attributed to harvests from multiple replications. Thus, an econometric technique within the multi-level meta-regression context is applied to correctly address the cross-sectional (locations) and temporal (years) dependencies among repeated observations from the same study in estimating production functions.

Results: Using appropriate functional forms and interactions for N and age, we found that the yield of miscanthus varies over N rates with substantial differences between the matured and older stands (Fig. 1). For example, without N, mean predicted yields of 6th and 10th year stands are 22.92 and 9.41 Mg ha^{-1}, respectively; yields increased to 28.37 and 19.10 Mg ha^{-1}, respectively, with 225 kg N ha^{-1}, with further increments to 30.82 and 25.79 Mg ha^{-1}, respectively, with 450 kg N ha^{-1}. The increased yield response to N of older age stands support some of the earlier findings that the decline in productivity, to some extent, can be offset with additional N fertilization. However, the differences in switchgrass yields with increasing N rates for the matured and older plants are quite identical. Specifically, plants at various stages of maturity showed an increasing response to N up to a specific level and then followed a declining trend with further increase (Fig. 2). For example, without N, mean predicted yields of 6th and 10th year stands are 8.39 and 5.81 Mg ha^{-1}, respectively; yields increased to 14.69 and 12.99 Mg ha^{-1}, respectively, with 200 kg N ha^{-1}, and then slightly decreased to 13.88 and 12.39 Mg ha^{-1}, respectively, with 250 kg N ha^{-1}. A general conclusion that can be drawn from the predicted yield response of age and N interaction, controlling for study, location, climatic, and management factors is that the yield maximizing N rate varies for miscanthus whereas it remains fairly identical for switchgrass over the maturity stages.

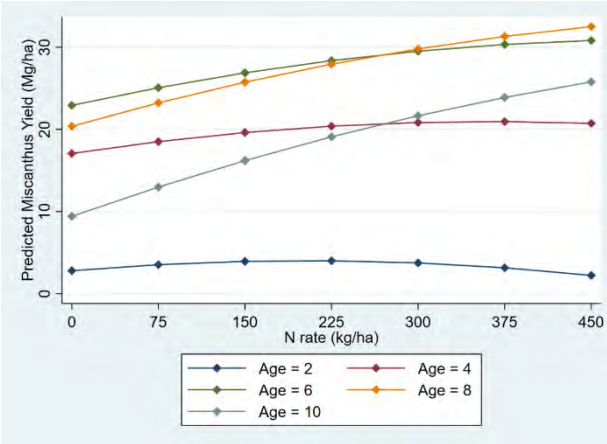


Fig 1. Predicted miscanthus yield

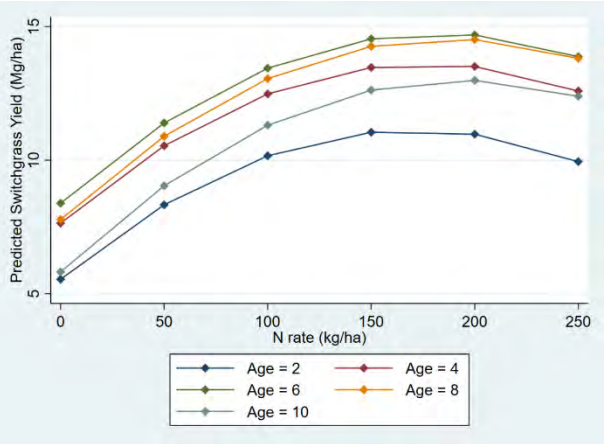


Fig 2. Predicted switchgrass yield

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.

Optimizing Miscanthus Regeneration and Transformation

Anthony Trieu[1,2*] (ttrieu@hudsonalpha.org), Rebekah Wood,[1] Brandon James,[1,2] Mohammad Belaffif,[1,2] Shilpa Manjunatha,[1,2] Rebecca Billingsley,[2,4] Subarna Sharma,[2,4] Anjali Arpan,[2,4] Erik Sacks,[2,3] **Nancy Reichert,**[2,4] and **Kankshita Swaminathan**[1,2]

[1]HudsonAlpha Institute for Biotechnology, Huntsville, AL; [2]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [3]University of Illinois Urbana-Champaign, Urbana; and [4]Mississippi State University, Mississippi State

https://cabbi.bio/research/feedstocks-theme/

Project Goals: One of the missions of the Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) is to develop efficient ways to understand, grow, and sustainably increase the value of bioenergy crops. The goal of our work is to develop transformation methods to evaluate gene function and engineer miscanthus for traits of interest.

Miscanthus, an important bioenergy crop for biomass production, is a C4 grass native to Asia. Conventional breeding of miscanthus for genetic improvement and analysis of gene function is difficult because it displays self-incompatibility, heterogeneity of offspring, and relatively long life cycles. We are therefore exploring methods for direct genetic modifications to explore the function of genes from genomic studies and ultimately tailor miscanthus with traits of interest. Toward this goal, we have developed particle bombardment and *Agrobacterium*-mediated transformation methods for miscanthus. We have screened a number of genotypes from the miscanthus collection at the University of Illinois, as well as commercially available lines. From these, we have selected a few *M. sinensis*, *M. sacchariflorus*, and *M. x giganteus* lines that perform well in tissue culture and are transformable. Currently, we are optimizing transformation methods for these selected lines using highly embryonic calli induced from immature inflorescences and seeds. We have successfully transformed miscanthus with constructs carrying different selection markers and trait genes. We are in the process of analyzing T0 transformants from these experiments.

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.*

Quantifying Root Carbon Rhizodeposition from Bioenergy Cropping Systems in the Midwest United States

Adam C. von Haden[1,2]* (avh@illinois.edu), **Wendy H. Yang,**[1,2,3,4] and **Evan H. DeLucia**[1,2,3]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Institute for Sustainability, Energy, and Environment, University of Illinois Urbana-Champaign, Urbana; [3]Department of Plant Biology, University of Illinois Urbana-Champaign, Urbana; [4]Department of Geology, University of Illinois Urbana-Champaign, Urbana.

https://cabbi.bio/research/sustainability-theme/

Project Goals: Quantifying and predicting ecosystem-level carbon (C) cycling processes are critical steps toward assessing the overall sustainability of bioenergy and bioproduct feedstocks. While aboveground C cycle components are routinely evaluated *in situ*, belowground C components remain physically and technically challenging to study and therefore remain poorly understood. Rhizodeposition of water-soluble, low-molecular weight C compounds via root exudation, sloughing, and mucilage may represent a substantial transfer of C from plants to soil, but this flux is seldom represented in ecosystem C budgets and models. Moreover, rhizodeposition C may provide a labile energy source for soil microbes, which provides an important link between cycling of C and other critical soil nutrients such as nitrogen. Our goal was to quantify C rhizodeposition in key bioenergy cropping systems to reduce uncertainty in empirical C budgets and improve ecosystem-scale predictive models.

The potential for bioenergy cropping systems to mitigate net carbon (C) emissions through soil C sequestration represents a major sustainability benefit over fossil fuels. Most empirical research has indicated that perennial biofuel cropping systems enhance soil C sequestration rates over their annual counterparts. However, the mechanisms underpinning this phenomenon are often speculative, and therefore predictive ecosystem models may not capture these dynamics accurately. Although root rhizodeposition of soluble C compounds may account for a substantial portion of net primary productivity in bioenergy cropping systems, it is not well-represented in many C budgets and ecosystem models due to the paucity of *in situ* empirical studies. We collected water-soluble root rhizodeposition C from miscanthus, switchgrass, bioenergy sorghum, and maize cropping systems near the peak of two growing seasons in Central Illinois. In addition, we collected root cores to scale specific root rhizodeposition rates to the ecosystem level. On most dates, bioenergy sorghum had significantly higher specific root rhizodeposition rates than miscanthus, and bioenergy sorghum often had higher specific root rhizodeposition rates than maize and switchgrass. Bioenergy sorghum that was fertilized at a typical nitrogen (N) rate trended toward lower specific root rhizodeposition rates than unfertilized bioenergy sorghum, but this pattern was not statistically significant. After scaling by root biomass, average ecosystem-level rhizodeposition was approximately 450 mg C m^{-2} d^{-1}, with no consistent differences observed among the bioenergy cropping systems. Our preliminary estimate indicates that root rhizodeposition accounts for approximately 2% to 7% of annual net primary productivity in both perennial and annual cropping systems. Thus, although the root rhizodeposition C flux is not likely the primary input for soil C sequestration, it nonetheless may

play a substantial role in nutrient cycling and therefore warrants explicit attention in ecosystem C budgets and models.

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.*

Improving Photosynthetic Efficiency of C4 Energy Crops: A Dynamic Modeling Analysis

Yu Wang[1,2*] (yuwangcn@illinois.edu), Kher Xing Chan,[1,2] and **Stephen P. Long** [1,2]

[1]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana;
[2]DOE Center for Advanced Bioenergy and Bioproducts Innovation

https://cabbi.bio/research/feedstocks-theme/

Project Goals:
To meet the increasing societal need for energy, one of the missions of the Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) is developing efficient ways to increase the biomass productivity of bioenergy crops and improve the efficiency of conversion from biomass into valuable chemicals.

The goal of our work is to engineer, or identify, more photosynthetically efficient sorghum and sugarcane germplasm. A dynamic model from leaf metabolism to crop canopy could help us identify potential targets for improving energy, water, and nitrogen use efficiency, and increase the biomass productivity of bioenergy crops in various environmental conditions.

1. **Develop a dynamic metabolic model for general C4 plants.**
2. **Parameterize the model using measured gas exchange data of bioenergy crops, such as sugarcane and sorghum, and predict targets for increasing photosynthesis under non-steady state conditions.**
3. **Develop a canopy model for C4 bioenergy crops and identify targets for increasing canopy photosynthesis and biomass accumulation.**

Photosynthesis is the ultimate source of bioenergy and bioproducts, and its efficiency is closely related to the productivity of crops. Increasing crop photosynthetic efficiency is one means to meet the pressure of increasing food, biofuels, and bioproducts without the need to bring more land into agriculture. The most productive C4 biofuel crops, such as corn, sugarcane, and sorghum, utilize NADP-ME-type C4 photosynthesis. Despite high productivities, these crops still fall well short of the theoretical maximum energy conversion efficiency. Understanding the basis of these inefficiencies is key to improving the productivity of these C4 crops.

Attempts to improve photosynthesis have focused on efficiency under constant high light, steady-state conditions. However, in the field, leaves are rarely in steady state and are affected by frequent light fluctuations. This has led to a growing awareness of the need to address photosynthetic efficiency in fluctuating light.

We measured the rates of CO_2 uptake and stomatal conductance of sorghum and sugarcane under fluctuating light regimes. The measured gas exchange data was combined with a new dynamic C4 photosynthesis model to infer the limiting factors for each crop species under non-steady

state conditions. Our modeling identified Rubisco activase, the PPDK regulatory protein, and rate of stomatal opening as key limitations under non-steady state conditions, although this may vary between species. Our work identified feasible targets for improving photosynthetic efficiency of bioenergy crops, which are now being bioengineered.

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.*

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

https://cabbi.bio/

Project Goals: The goal of this project is to metabolic engineering of *Yarrowia lipolytica* to enhance accumulation of 3-acetyl-1,2-diacyl-sn-glycerol. The engineered Yarrowia strains can be a useful base strain to screen highly active TAG synthase, which can be leveraged in the sustainability teams to pursue high TAG production in plant lines.

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 modify *Y. lipolytica* to accumulate acTAG while maintaining a high carbon flux toward TAG synthesis. 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 that deleted three acyltransferase, two lipases, and one dehydrogenase. This engineered strain is incapable of producing lcTAG and forming lipid bodies due to removal of three TAG synthase genes. In the second strategy, we evaluated activities of 28 bioprospected acTAG synthase homologs in an engineered *Y. lipolytica* with two deleted acyltransferases and two deleted lipases. We integrated each acTAG synthase homolog randomly in the strain and quantified cellular lipid content using a nile-red staining fluorescence assay. Our preliminary results show that 5 acTAG synthase homologs showed activities greater than the literature enzyme *Ea*DAcT from *Euonymus alatus*. To this end, we engineered a *Y. lipolytica* strain which lacks native TAG synthesis and identified five promising acTAG synthase candidates. We believe these efforts will be beneficial toward enhancing acTAG production in *Y. lipolytica*.

References

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.

Comparison of Actively Growing (RNA) and Potentially Active (DNA) Soil Microbial Communities in *Miscanthus x giganteus*

Jihoon Yang[1,2*] (jhyang@iastate.edu), Jaejin Lee,[1,2] and **Adina Howe**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation, [2]Department of Agricultural and Biosystems Engineering, Iowa State University, Ames.

Project Goals: This study provides insights into plant-soil-microbe interactions that are associated with fertilization management and stand age in miscanthus. Actively growing microbial communities (RNA) are used to identify key taxa and functions that contribute to nitrogen cycling. These results are directly compared to observations from DNA-based characterizations, where varying patterns have been observed. This effort will provide clearer insights into how to sustainably manage nitrogen needs in CABBI feedstocks based on microbial community characterization.

https://cabbi.bio/research/sustainability-theme/

Introduction

Miscanthus x giganteus is a promising bioenergy crop because of its ability to produce a large amount of biomass. Plant, soil, and microbe interactions have been shown to influence its productivity under varying fertilization rates and stand ages. Previously, the characterization of microbial communities has been limited to DNA-based analyses which provide information on community membership but not activity. In this study, we use RNA-based approaches to expand our understanding of the impacts of the actively growing soil microbial community on the sustainable production of *M. giganteus*.

Research Approach

Soil samples (n=403) from plots receiving 0, 200, and 400 kg N ha^{-1} were obtained from fields containing two-, three-, and four-year-old *M. giganteus*. Paired DNA and RNA extractions were performed using the MagAttract PowerSoil DNA EP kit (Qiagen) and Rneasy PowerSoil Total RNA kit (Qiagen), respectively. Amplicon sequencing was performed with 150 bp paired-end libraries for the 16S rRNA gene of both extracted DNA and RNA (Illumina Miseq platform, Argonne National Laboratory). Taxa associated with DNA and RNA communities were identified using the DADA2 package and annotated with the closest representative in the RDP classifier. Statistical comparisons of enriched taxa under fertilization and differences between RNA and DNA communities were made with PERMANOVA and PCoA using the Bray-Curtis dissimilarity matrix.

Results

RNA and DNA membership were observed to be significantly different in our miscanthus soils (Figure 1A), highlighting the need to understand and differentiate actively growing communities in response to management strategies. The difference between RNA- and DNA-microbial communities was largest in the more mature miscanthus stands, with significant differences observed in three- and four-year-old miscanthus (Figure 1B). The effect of nitrogen fertilization on actively growing communities was most pronounced in the four-year-old miscanthus stands (Figure 1C). In these communities, nitrogen fertilization generally reduced the proportion of taxa

associated with nitrogen fixation and nitrification. Additionally, taxa identified as unique to actively growing microbiomes in three-year and four-year-old miscanthus include the nitrifying bacteria including *Nitrosomonas*, *Nitrosospira*, and *Sporosarcina* (Figure 1D).

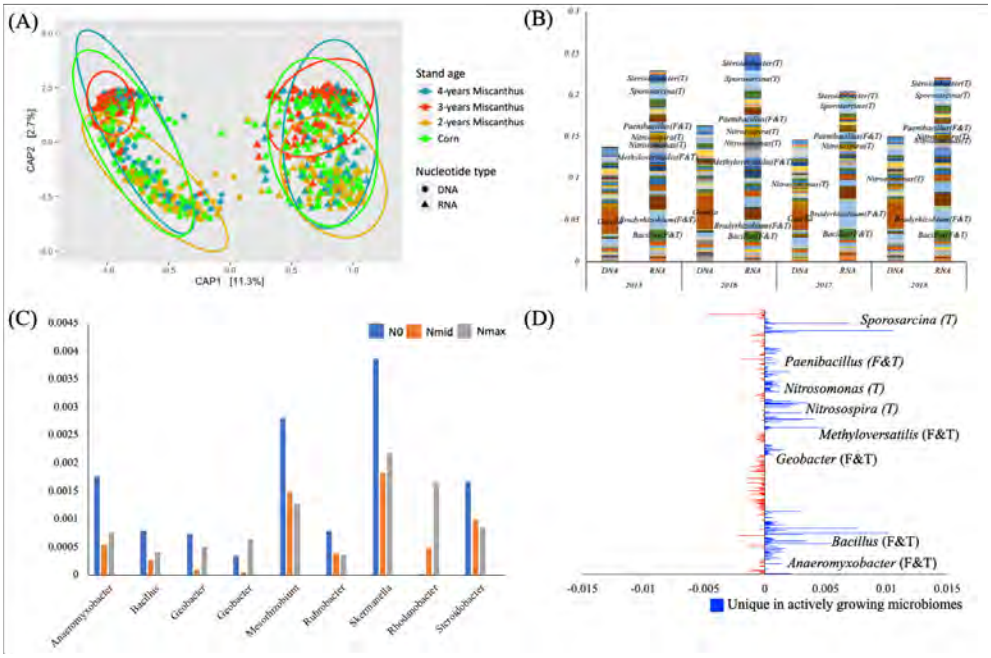


Figure 1. (A) microbial membership difference between actively growing and potentially active microbiomes, (B) abundance of nitrogen cycling bacteria with different stand ages of miscanthus, (C) differences in the nitrogen-fixing and nitrifying bacteria with different concentration of fertilizer, and (D) unique bacteria in the actively growing microbiomes.

Summary

The potentially active (DNA) and actively growing (RNA) microbiomes in miscanthus are different. Nitrogen-fixing and nitrifying bacteria were 1.5 times more abundant in the actively growing soil microbial communities compared to DNA-based estimates, suggesting that these communities can provide contrasting insights into responses to N-cycling. The actively growing microbiome of mature miscanthus was associated with less turnover, in contrast to younger miscanthus where the actively growing community had a more variable impact of fertilization.

Future Work

Seasonal patterns of microbial response to fertilizer between RNA and DNA approaches will be examined, with the hypothesis that more mature miscanthus will have a more predictable response in the actively growing membership. The role of actively growing microbiomes will be further explored through the measurement of N-cycling genes in RNA extracts.

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 author(s) and do not necessarily reflect the views of the U.S. Department of Energy.*

Metagenomic binning of high latitude (79°N), mineral permafrost active layer

Katie Sipes, Tatiana Vishnivetskaya, Sarahi Garcia, Karen Lloyd

Svalbard active layer permafrost is at risk of permanent thaw as the climate warms. In order to study as many organisms as possible in this diminishing environment we sequenced five locations in the Bayelva monitoring site in Ny Ålesund Svalbard (79°N). Using the novel technology of Novaseq (Illumina), we produced 1 Tb of metagenomic data from 56 active layer soil samples, rivalling the amount of data produced by the TARA Oceans project. We assembled these metagenomes separately and binned them to make metagenomic assembled genomes. We present preliminary information on over 1,000 taxonomically identified MAGs in this active layer. As this investigation continues, we will employ the super-computing facilities at Oak Ridge National Laboratory to investigate the metabolic annotations and functional pathways of these organisms that are at the forefront of the effects of climate change.

Contribution Evenness: A functional redundancy metric for microbially-mediated biogeochemical rates and processes

Taylor Royalty and Andrew D. Steen

The concept of functional redundancy has received considerable attention in both the macroecology and microbial ecology literature. As a result, multiple metrics of functional redundancy have been proposed. These vary in how they weight trait levels, species abundance, functional richness, and species richness. Trait-based models are also increasingly popular ways to represent the function of microbial ecosystems. Here we present a new functional redundancy measurement tailored to be useful for community-aggregated parameters, which are summed ecosystem processes contributed by species, such as rates of chemical transformations. We focus in particular on community-aggregated parameters mediated by microbial communities. We call this approach contribution evenness (CE). CE measures how evenly species contribute to a community-aggregated parameter based on species contribution levels and abundances. As CE is an evenness measurement, quantitative representation of functional redundancy spans between 0 and 1, where 0 corresponds to a single species contributing and 1 corresponds to all species contributing equally to a community-aggregated parameter. CE is sensitive to the resilience of an ecosystem function to species extinction—an ecological consequence of functional redundancy. To illustrate how CE can be used, we analyzed the functional redundancy of eight nitrogen-transforming pathways using 2,631 metagenome-assembled genomes from 47 marine sites. We found that, on average, NH_4^+ assimilation pathway was the most functionally redundant (0.44 ± 0.08) while dissimilatory nitrate reduction had the lowest functional redundancy (0.005 ± 0.005). As demonstrated here, CE is a metric that is sensitive to trait resilience and allows for comparative analysis of functional redundancy for different traits.

Microbial organic matter degradation in the active layer of Svalbard permafrost

Katie Sipes, Tatiana Vishnivetskaya, Reagan Paul, Karen Lloyd

The active layer of permafrost in Svalbard, Norway (79°N) is deepening at rate of about 1 vertical centimeter per year in most locations around the Bayelva River in the Leirhaugen glacier plain. We used a multifaceted approach to study the organisms that live in these expanding active layers including metagenomes, culturing, and soil geochemistry. *We hypothesized that the organisms that were obtained in culture would be found in the metagenomes and the metagenomes would give insight on the metabolisms of these organisms.* Further, the soil geochemistry would suggest the types of energy sources that the organisms use. Our metagenome data yielded 5 metagenome assembled genomes (MAGs) from two active layer cores. Our culture isolation efforts yielded 10 *Psudomonas* spp strains. Related isolates have been previously found in soil samples from Therperuvian Andean Plateau, Thuringian Basin, Council Alaska, Livingston Island Antarctica and in Western Spitsbergen. We saw evidence for microbial heterotrophy in one of the sites due to the carbon isotope values becoming richer in ^{13}C with depth. We also found enzymatic degradation activity of a wide range of small peptides and carbohydrates. This multifaceted approach lays the groundwork to understand how the active layer will become a hot spot for microbial activity as the climate continues to warm.

Optimizing Lipogenic Factors for Vegetative Oil Accumulation

Sanket Anaokar[1*] (anaokars@bnl.gov), Yuanheng Cai[2], Yingqi Cai[1] and **John Shanklin[1, 2]**

[1]Brookhaven National Laboratory, Upton, NY; [2]Stony Brook University, Stony Brook, NY

https://rogue.illinois.edu/

Project Goals

1. Increasing oil accumulation and in mature stems of energycane and *Miscanthus*

2. Identifying overexpression/downregulation gene combinations to optimize TAG accumulation

3. Dissecting mechanisms underlying the strength of lipogenic factors

Abstract text

Plant vegetative tissues account for a most of plant biomass, so it is of great interest to engineer plants with high oil content in vegetative tissues. However, plants normally only accumulate high levels of oils in the seeds. Our work therefore focuses on translating knowledge gained from seed oil accumulation to vegetative tissues. Oil accumulation in plants generally falls into a 'push-pull-protect model in which push factors increase fatty acid synthesis, pull factors efficiently channel fatty acids into TAG, and protect factors minimize TAG degradation. Increased oil accumulation requires the up-regulation of genes involved in lipid synthesis and down-regulation of those involved in lipid degradation. We are investigating modifications to well-described lipogenic factors in addition to additional potential lipogenic factors to boost TAG accumulation.

The Push factor is acetyl Co-A carboxylase (ACCase), the initial and commonly rate-limiting step in FA synthesis, that produces malonyl-CoA for fatty acid biosynthesis. In plants, there are two types of ACCase, namely heteromeric ACCase (HetAcc) and homomeric ACCase (HoACC). HetACC is composed of four subunits and located in the plastid, whereas HoAcc is a cytosolic protein comprising a single large polypeptide. Although HetACC is predominantly involved in *de novo* fatty acid synthesis in the plastid, its activity is strongly feed-back inhibited by 18:1-ACP, PII and BADC. Therefore, we re-targeted the cytosolic HoACC into the plastid by fusing the N-terminal signal peptide of Rubisco small subunit fused to GFP with HoAcc (RGHoACC). Confocal image analyses showed that RGHoACC is correctly targeted into the plastids. Further analysis showed that RGHoACC transient over-expression in tobacco leaves resulted in more than 20% total fatty acid accumulation and a > 7-fold increase in TAG accumulation relative to EV control.

The Pull factor is diacylglycerol acyl transferase (DGAT), that catalyzes the final step 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 accumulated more than a 6-fold increase in TAG relative to the control leaves.

For the Protection of TAG in vegetative tissue we engineered an improved version of the *sesame* oleosin. Based on published reports, it has been suggested that oleosin is marked for degradation via conjugation with ubiquitin. Lysine, cysteine, serine, threonine or the N-terminus a target protein are the most common target residues for Ubiquitin ligation. In an attempt to stabilize the *sesame* oleosin, we choose to convert all six-lysine residue to arginine residues (KR) to maintain the positive charge distribution. Dr. N. Roberts previously reported a variant of *sesame* oleosin, in which he replaced six positions in the amino acid sequence for cysteines. In our current study, we choose to eliminate the amino acids at those six locations (CysDel) from the oleosin peptide creating a shorter version of Ole1. Our results indicate that both lysine-to-arginine conversions (KR) and amino acid deletions (CysDel) improved their ability to protect TAG in leaves of *N. benthamiana.* Upon combining the Ole1_CysDel and KR mutations, we created an oleosin that supported a 2-fold improvement of TAG accumulation in *N. Benthamiana* leaves. As the K-to-R replacements (KR) were applied to all the lysines, and amino acid deletions (CysDel) were applied to all six locations, we predicted some of the modifications would have a negative effect on TAG accumulation. By analyzing dropout mutations at each position, we determined the contribution of each mutation. We combined this information to create Ole1_5 Mod, containing mutations at five of the 11 locations. The Ole1_5 Mod variant further improved TAG accumulation in leaves of *N. Benthamiana* relative to all other *sesame* oleosin variants.

In summary, our strategy was first to individually optimize the Push, Pull and Protect factors described above. We subsequently combined the optimized factors with respect to promoting plant vegetative oil 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 empty vector controls, reaching a final TAG content of ~4% (w/w) over a three-day period.

Funding statement

The work of Sanket Anokar, 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).

Metabolic Engineering of Energycane for Hyperaccumulation of Triacylglycerol and Improved Biomass Production

Viet Dang Cao[1*] (caodangviet@ufl.edu), Guangbin Luo[1], Baskaran Kannan[1], Hui Liu[2], John Shanklin[2], Stephen P. Long[3], and **Fredy Altpeter[1]**

[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

Renewable Oil Generated with Ultra-productive Energycanes—or ROGUE—is engineering the two most productive American crops—energycane and Miscanthus—to produce a sustainable supply of biodiesel, biojet fuel, and bioproducts.

Project goals are to:
1) Engineer energycane to produce an abundance of lipids in the form of triacylglycerol which can be converted into biodiesel, biojet fuel, and bioproducts.
2) Altered expression of flowering genes, pyruvate Pi dikinase and proteins involved in chloroplast division in energycane to enhance biomass yield and cold tolerance.

Abstract

Metabolic engineering to achive hyper-accumulation of lipids [e.g. triacylglycerol (TAG)] in the vegetative tissues of high biomass crops is a promising strategy to improve lipid yields for biofuel production. Energycane is an ideal feedstock for this approach due to its superior biomass production and persistence.

In this study, a multigene expression construct was used to elevate the production of free fatty acids, catalyze their conversion into TAG and prevent TAG hydrolysis. This construct was transferred into energycane callus, using the biolistic particle delivery system. Presence of transgenes in the regenerated plants were confirmed by PCR. A combination of TLC and GC-MS analysis revealed that the TAG contents in transgenic leaf tissue was elevated more than 18-fold compared to wild-type energycane. Currently, we are propagating the highest TAG producing energucane plants for field trials which are scheduled to begin under USDA-APHIS permit in April 2021 and will allow analysis of lipid and biomass production.

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 is expected to 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. Often flowering also leads to dehydration of the stalk tissues, which negatively affects stalk density, and also compromises sugar extraction in conventional sugarcane or lipid extraction in metabolically engineered lipid cane. Therefore, we recently generated transgenic energycane plants harboring a construct for RNAi mediated 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 biosafety of the engineered crop. Transgenic plants are currently being propagated in the greenhouse in preparation of field testing and will be characterized for target gene suppression during photo inductive period as well as for flowering and biomass production.

Genetic improvement of photosynthetic efficiency could potentially be achieved by developing a photosynthetically more effective canopy. To evaluate the effect of chloroplast size on light penetration into the canopy and biomass production, we intend to modify the expression of the cytoskeletal Filamenting temperature-sensitive Z (FtsZ) protein, which is critical for chloroplast division. Overexpression and RNAi constructs of FtsZ were introduced into energycane callus and regenerated through somatic embryogenesis. Pyruvate orthophosphate dikinase (PPDK) has been proposed as rate limiting enzyme in C4 photosynthesis. It regenerates the substrate phosphoenol pyruvate (PEP) for the initial carbon-fixation step. C4 plants are also severely limited by low temperature, possibly because PPDK is highly cold-labile and partially dissociates below 14 ºC. Therefore, we decided to explore the over-expression of Miscanthus x giganteus PPDK in energycane. MxgPPDK with its native regulatory sequences were introduced into energycane callus by biolistic gene transfer. The regenerated plants will be evaluated for the effect of PPDK overexpression on photosynthetic efficiency, cold tolerance and biomass accumulation.

Funding Statement

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018254.

Development of Sustainable Transformation of *Miscanthus* × *giganteus* to Improve Photosynthesis

Kher Xing Chan[1]* (cindyckx@illinois.edu), Karolina Sobańska[2], Cameron Kern[1], Benjamin Haas[1], Steven J. Burgess[1], Fredy Altpeter[3], and **Stephen P. Long[1]**

[1]University of Illinois at Urbana-Champaign, Urbana, IL; [2]Institute of Plant Genetic, Poznań, PL; [3]University of Florida, Gainesville, FL

https://rogue.illinois.edu/

Project Goals

Renewable Oil Generated with Ultra-productive Energycanes—or ROGUE—is engineering the two most productive American bioenergy crops—energycane and *Miscanthus*—to produce a sustainable supply of biodiesel, biojet fuel, and bioproducts.

The main objectives of this work are:

1) To improve the conversion of sunlight into plant biomass/metabolites through photosynthesis without the need for increased quantities of either water, or fertilizer.
2) To transfer ROGUE technologies from the lab bench to crops through an efficient pipeline.

Abstract

Improving photosynthetic efficiency of bioenergy crops, such as *Miscanthus* × *giganteus* and energycane, would provide sustainable sources of energy and bioproducts to achieve zero net greenhouse gas emission in future. These highly productive C_4 crops utilize the NADP-ME type C_4 photosynthesis. In this type of photosynthesis, the amount of pyruvate orthophosphate dikinase (PPDK) and rubisco limit the regeneration of phospho*enol*pyruvate (PEP) (Long *et al.*, 2013, Wang *et al.*, 2008, Naidu *et al.*, 2003). Research also demonstrated improvement in photosynthetic efficiency under fluctuating light when photoprotection response time is accelerated by overexpression of genes involved in non-photochemical quenching (NPQ): zeaxanthin epoxidase (ZEP), violaxanthin de-epoxidase (VDE) and Photosystem II subunit S (PsbS) (Kromdijk *et al.*, 2016). We propose that the overexpression of PPDK will alleviate rate limitation in C_4 photosynthesis, while increasing expression of NPQ-related genes can improve photosynthetic efficiency in *Miscanthus*. Although *M.* × *giganteus* is an ideal candidate of bioenergy crop with minimal invasive potential, the propagation of this feedstock is limited by the sterility of the plant due to its triploid genome (Boersma & Heaton, 2014). Biolistic transformation of *Miscanthus* is generally performed on calli induced from immature inflorescences collected during the summer (Sobańska *et al.*, 2019). However, this is a limitation to develop a continuous transformation system and has been a bottleneck of our effort to understand the effect of overexpressing the photosynthetic genes. In this study, we used calli induced from immature inflorescence and stem

meristemic tissue of *M.* × *giganteus* for biolistic transformation and performed a proof-of-concept experiment using fluorescent protein construct to check the efficiency of the systems. Preliminary data suggested that the transformed calli from stem meristemic tissue could be transformed and regenerated. These two embryogenic callus induction systems would provide a continuous source of calli for transformation to study the effect of photosynthetic genes, such as PPDK, VDE, ZEP and PsbS in *Miscanthus*.

References / Publications

Long SP, Spence AK (2013) Toward Cool C_4 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 C_4 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 C_4 photosynthesis in *Miscanthus* × *giganteus*: Adaptation in Amounts and Sequence of C_4 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.

Boersma NN, Heaton EA (2014) Propagation method affects *Miscanthus* × *giganteus* developmental morphology. *Ind Crops Prod* 57(M):59–68.

Sobańska K, et al. (2019) Optimised expression cassettes of hpt marker gene for biolistic transformation of *Miscanthus sacchariflorus*. *Biomass and Bioenergy* 127:105255.

Funding Statement

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018254.

Improving Photosynthetic Efficiency of C_4 Bioenergy Crops in Fluctuating Lights

Moonsub Lee[1*] (mlee128@illinois.edu), Fredy Altpeter[2], Ryan Boyd[1], and **Donald Ort[1]**

[1]University of Illinois at Urbana-Champaign, Urbana, IL; [2]University of Florida, Gainesville, FL

https://rogue.illinois.edu

Project Goals

Renewable Oil Generated with Ultra-productive Energycanes—or ROGUE—is engineering the two most productive American bioenergy crops—energycane and *Miscanthus*—to produce a sustainable supply of biodiesel, biojet fuel, and bioproducts.

Project goals are to:

1) Engineer energycane and *Miscanthus* to produce an abundance of natural oil that can be converted into biodiesel, biojet fuel, and bioproducts.
2) Improve how plants convert sunlight into plant matter through photosynthesis without more water or fertilizer.

Abstract

As demand for energy increases, bioenergy crops have gained considerable attention as an alternative energy source. To improve yield, previous studies have focused on breeding, soil fertility, and harvest management. However, yield potential depends on a crops ability to intercept light and the efficiency of converting light into biomass through the process of photosynthesis. It is known that fluctuations in light intensity, potentially caused by sun-leaf angle, wind or intermittent clouds, can decrease photosynthetic efficiency. Previous modeling of photosynthesis suggested that the discoordination of metabolic cycles in chloroplasts is a key factor reducing photosynthetic efficiency caused by fluctuating light. Therefore, we hypothesized that increased metabolite pool size may minimize decreases in photosynthetic efficiency by providing a buffer when cycles become asynchronous during fluctuating light. We further hypothesized that

increased chloroplast volume could increase metabolite pool sizes. For that reason, we are altering chloroplast volume by engineering chloroplast division genes in two high performing bioenergy crops sugarcane and energycane. Currently, we are confirming changes to chloroplast volume of our transgenic lines in greenhouse conditions. We will select the best performing transgenic lines for field trials.

References / Publications

Slattery, R.A., Walker, B.J., Weber, A.P., and Ort, D.R. (2018). The impacts of fluctuating light on crop performance. Plant physiology, 176, 990-1003.

Zhu, X.G., Long, S.P., and Ort, D.R. (2010). Improving photosynthetic efficiency for greater yield. Annual review of plant biology, 61, 235-261.

Funding Statement

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018254.

Extending the Use of Time-domain ^{1}H-NMR for Rapid and Non-invasive Quantification and Characterization of *In-situ* Lipids in Transgenic Bioenergy Crops

Shraddha Maitra[1*] (smaitra@illinois.edu), Bruce Dien[2], Stephen P. Long[3], and **Vijay Singh[1]**

[1, 3]University of Illinois at Urbana-Champaign, Urbana, IL; [2]National Center for Agricultural Utilization Research, United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Peoria, 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. On successful completion of the project, modified bioenergy crops would generate approximately 15 times more oil per acre than seed crops like soybean. The present work demonstrates the development of a non-invasive technique for rapid and chemical-free quantification and characterization of the total in-situ lipids in transgenic bioenergy crops.

Abstract

Biofuels provide green alternatives to the fossil-fuel. However, for biodiesel, we are still dependent on oilseeds, which also constitute part of human food. Bioenergy crops with high triacylglyceride content have the potential to replace oilseeds. To this end, Andrianov et al (2010), Sanjaya et al (2013), and Zale et al (2016) have successfully reported enhanced TAG accumulation in *Nicotiana tabacum*, *Arabidopsis thaliana*, and sugarcane respectively, (Andrianov et al., 2010; Sanjaya et al., 2013; Zale et al., 2016). For energycane, research efforts are underway (ROGUE project) to accumulate TAG molecules. The analysis of *in-situ* lipid contents during developmental stages of transgenic lines and subsequent to each bioprocessing protocol requires tedious sample preparation and extraction with an organic solvent. Therefore, time-domain ^{1}H-NMR spectroscopy has been successfully adapted for the quantification and characterization of *in-situ* lipid in bioenergy crops that eliminates the steps involved in sample preparation and solvent extraction. Td ^{1}H-NMR was

calibrated to quantify *in-situ* lipids. The measured values with NMR spectroscopy were validated by comparing them with the values obtained from the conventional solvent extraction method. The cross-referenced values were not significantly different. Moreover, the relaxation time distribution spectrum was analyzed to characterize i*n-situ* lipids into bound and free form. It also helped in investigating the suitability of pretreatment protocols for transgenic bioenergy crops by facilitating the resolution of the effect of pretreatment protocols on the local proton-population of the biomass sample upon pretreatment. Biomass pretreated with a two-staged pretreatment protocol showed promising results for recovery of both fermentable sugar and oil and can be directly analyzed by NMR without neutralization steps, unlike dilute acid and alkaline pretreated biomass.

References / Publications

Andrianov, V., Borisjuk, N., Pogrebnyak, N., Brinker, A., Dixon, J., Spitsin, S., Flynn, J., Matyszczuk, P., Andryszak, K., Laurelli, M., Golovkin, M., Koprowski, H., 2010. Tobacco as a production platform for biofuel: Overexpression of Arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. Plant Biotechnol. J. 8, 277–287. https://doi.org/10.1111/j.1467-7652.2009.00458.x

Sanjaya, Miller, R., Durrett, T.P., Kosma, D.K., Lydic, T.A., Muthan, B., Koo, A.J.K., Bukhman, Y. V., Reid, G.E., Howe, G.A., Ohlrogge, J., Benning, C., 2013. Altered Lipid Composition and Enhanced Nutritional Value of Arabidopsis Leaves following Introduction of an Algal Diacylglycerol Acyltransferase 2. Plant Cell 25, 677–693. https://doi.org/10.1105/tpc.112.104752

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

Funding Statement

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018254.

Deciphering genetic and physiological mechanisms that enhance nitrogen use efficiency and seed oil accumulation in camelina

Chaofu Lu [1,*] (clu@montana.edu), Chengci Chen [1], Luca Comai [2], Jed Eberly [1], Andreas Fischer,[1] Jennifer Lachowiec [1], Trent Northen [4], Timothy Paulitz [3], William Schillinger [3], Jorg Schwender [5] , John Shanklin [5], Susannah Tringe [4], and Qing Yan [1]

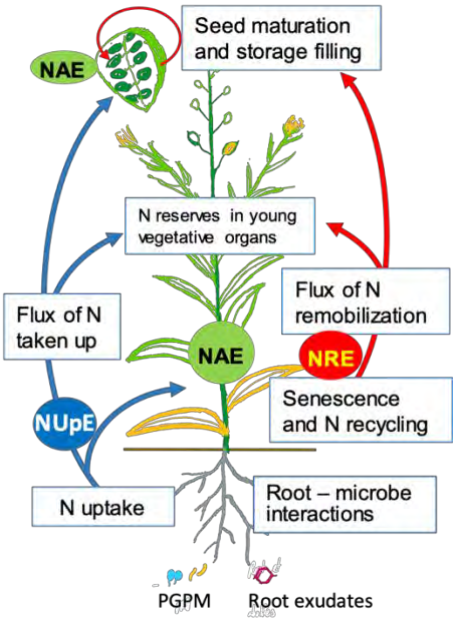
[1] Montana State University, Bozeman, MT; [2] University of California, Davis, CA; [3] Washington State University, Pullman, WA; [4] Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Brookhaven National Laboratory, Upton, NY.

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.

Camelina (*Camelina sativa*) is a promising non-food oilseed crop that provides a biofuel feedstock especially in the Northwest of the United States. Integration of camelina, a broadleaf cruciferous crop, could also improve the regions' cereal-based cropping systems and boost rural economies. This project addresses two critical challenges in camelina biology that hinder its great potential: 1) Enhancing nitrogen utilization efficiency (NUE) for economic profitability and reducing the negative environmental impact associated with nitrogen fertilization; 2) Boosting oil yield for its productivity competitiveness to other major oilseed crops such as canola.

Our overall goal is to obtain a systems-level understanding of genetic and physiological mechanisms that may be used to enhance NUE and oil accumulation in camelina; specifically how camelina plants, in partnership with beneficial soil microbes, acquire N to promote vegetative growth and biomass accumulation, then during reproductive growth remobilize N assimilates from senescing tissues to seeds, and finally utilize the C and N assimilates for the biosynthesis of seed storage products, particularly oil. We will answer some key questions in these processes based on our unique expertise and recent discoveries, which together with our other related research in camelina, have resulted in abundant genomic resources and established tools. Importantly, camelina is a closely related species to Arabidopsis. Our research will therefore benefit significantly from the knowledge already obtained in this latter model plant. We will achieve the following **specific objectives**:

1) Identify and characterize genes and gene networks in camelina that play key roles in regulating nitrogen use efficiency (NUE) and seed oil accumulation.
2) Define the camelina rhizobiomes, investigate their dynamics in composition and abundance responding to camelina genotypes and N conditions; and to isolate beneficial bacteria and test their functions in promoting camelina growth.
3) Characterize key physiological mechanisms in camelina that enhance the efficiencies of N uptake (NUpE) from soil and promote efficient N remobilization (NRE) from source to sink (developing seed) tissues, and those that enhance sink capacity (seed size and oil synthesis activity) to boost oilseed yield.



The goal will be achieved by the genomics driven approaches, including quantitative molecular genetics, comparative transcriptomics, proteomics, and metabolomics, as well as biochemical studies and metabolic flux analyses in phenotypically diverse camelina lines that are isolated from natural variation and mutants or created by transgenics.

Anticipated Outcomes and Potential Impact

This highly integrative project will provide a systems-level understanding of fundamental mechanisms in camelina that regulate nitrogen uptake, assimilation, remobilization, and seed development and oil accumulation. This knowledge is necessary for the development of next-generation high-oil-yielding camelina varieties with minimum nitrogen fertilization for sustainable bioenergy production. Besides the direct outcome of several natural and engineered camelina lines with advanced traits like large seed and high oil content, results from this research will fill knowledge gaps in many key areas of plant development, physiology and plant-microbe interactions. The basic science advancement will benefit the development of other related brassica oilseed species for sustainable production. In addition, this project will provide training opportunities for young scientists including graduate students and postdocs in genomics, biochemistry and biotechnology.

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.

Co-consumption of mixed sugars through the division of labor (DOL) in a synthetic *Saccharomyces cerevisiae* consortium

Jonghyeok Shin (Shin9114@illinois.edu)[1*], **Yong-Su Jin**[1], and **Ting Lu**[1].
[1]University of Illinois at Urbana−Champaign, Urbana

Project Goals: Short statement of goals. (Limit to 1000 characters)
For bioconversion of multiple sugars in lignocellulosic hydrolysates into value-added products, there are two utilization strategies: single 'superbugs' (SS) capable of utilizing all sugars, and multiple 'division of labor' (DOL) strains using each sugar. Our working hypotheses are that DOL might be favored over SS because of heavy metabolic cost—metabolic burdens and toxicity of a product—for utilizing multiple sugars and that DOL is more adaptive than SS to changing sugar compositions. To test the hypotheses, we constructed a set of engineered *Saccharomyces cerevisiae* (*S. cerevisiae*) strains to implement both strategies and built a multiscale mathematical model to quantitatively elucidate the general rules for designing optimal mixed sugar utilization.

Abstract
Abundant and inexpensive agricultural residues contain mixed sugars—mostly glucose and xylose—which can be utilized by wild-type or engineered microbes for the production of biofuels and chemicals. Previous studies have shown that microbes often exhibit a sequential utilization of the mixed sugars—preferential consumption of glucose over xylose. This resulted in impaired yield and productivity of a target molecule. Here, we dissected glucose and xylose consumption by constructing a yeast consortium with glucose utilizing strain and a xylose utilizing strain. The glucose-utilizing strain (Y_{G1}) was constructed by deleting endogenous hexose transporters (*HXT1-7* and *GAL2*) and introducing a heterologous glucose-specific transporter into *S. cerevisiae*. The xylose-utilizing strain (Y_{X1}) was constructed by deleting endogenous hexose transporters (*HXT1-7* and *GAL2*) and introducing a heterologous xylose-specific transporter into an engineered *S. cerevisiae* expressing xylose reductase (*XYL1*), xylitol dehydrogenase (*XYL2*), and xylulokinase (*XYL3*).

We observed that the consortium consisting of Y_{G1} and Y_{X1} could consume glucose and xylose simultaneously when equal amounts of glucose and xylose are present in a culture medium. However, we observed potential issues in achieving simultaneous consumption of glucose and xylose at various concentrations which are often resulted from different feedstocks. First, while the major hexose transporters were deleted, xylose consumption was still inhibited by glucose in Y_{X1}. In order to prevent glucose repression and eliminate consumption in Y_{X1}, we deleted hexokinases (*HXK1-2*) and glucokinase (*GLK1*) which are involved in glucose repression and metabolism. As a result, Y_{X2} (*HXK1-2* and *GLK1* deleted Y_{X1}) grew well on xylose in the presence of glucose and did not show glucose consumption at all. Second, Y_{G1} and Y_{X2} showed different specific glucose and xylose uptake rates, leading to difficulties in constructing an optimal consortium. If the consumption rates of glucose and xylose are not equally controlled, one of the strains could dominate the consortium. In order to synchronize sugar consumption rates of Y_{G1} and Y_{X2}, a less efficient glucose-transporter was introduced to Y_{G1} to build Y_{G2} and the copy numbers of the xylose-specific transporter in Y_{X2} were doubled to build Y_{X3}. By replacing the glucose sugar transporter and increasing the copy number of the xylose transporter, the glucose and xylose uptake rates of Y_{G2} and Y_{X3} were adjusted to a comparable level. Lastly, as monitoring the populations of Y_{G2} and Y_{X3} during cultivation was difficult, fluorescence proteins (GFP and RFP) were expressed in Y_{G2} and Y_{X3} to enable real-time monitoring of the populations in the consortium. The resulting

Y_{G2f} and Y_{X3f} were employed for studying the division of labor during the fermentation of glucose and xylose. Finally, we co-cultivated Y_{G2R} and Y_{X3R} in mixed sugar conditions of glucose and xylose. As expected, glucose and xylose were simultaneously consumed by a consortium of Y_{G2f} and Y_{X3f} strains. Next, we compared the consumption rates of glucose and xylose by the consortium containing Y_{G2f} and Y_{X3f} and a single strain SR8 which can consume glucose and xylose. The overall sugar consumption rate by the SR8 was higher than that by the DOL consortium under the tested conditions. This was mainly because of the inhibition on xylose utilization by ethanol at high concentrations, suggesting that simultaneous consumption might not be a prominent solution for the production of ethanol from a mixture of glucose and xylose. To improve ethanol production from a mixture of glucose and xylose by a DOL consortium, we switched the order of glucose and xylose utilization. As xylose consumption was severely inhibited by the produced ethanol from glucose consumption, we reasoned that utilization of xylose ahead of glucose consumption might increase overall sugar consumption and ethanol production rate from a mixture of glucose and xylose by a DOL consortium. By controlling the timing of DOL between Y_{G3f} and Y_{X3f}, we observed that early xylose consumption by only Y_{X3f} could avoid ethanol inhibition when Y_{G3f} (D452-2 expressing GFP) was inoculated later. These results suggest that we need to think of the timing of DOL for maximizing the overall sugar consumption and product formation rates.

In summary, we confirmed that fine-tuning consumption rates of glucose and xylose by Y_G and Y_X can be utilized to optimize the simultaneous consumption of glucose and xylose by a DOL consortium in response to different sugar concentrations in cellulosic hydrolysates. Nonetheless, the overall sugar consumption and ethanol production rates by SS were higher than those by DOL. As the performance of DOL was impacted by severe inhibition on xylose fermentation by ethanol produced from glucose fermentation, we controlled the timing of DOL by inoculating a xylose-utilizing strain only and supplementing a glucose-utilizing strain later. As a result, the performance of DOL was comparable to that of SS. Our results demonstrate that the division of labor in ecosystems and controlling the timing of DOL can be applied for mixed sugar fermentation for the efficient production of value-added products.

Probing the Stability of a Synthetic Symbiotic Microbial Consortium

Yongping Xin[1,2,*] (yongping@illinois.edu), **Yong-Su Jin**[1,3], and **Ting Lu**[1,2]

[1]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL; [2]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL. [3]Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, IL

Project Goals:

To systematically quantify the robustness of an engineered mutualism between *Lactococcus lactis* and *Saccharomyces cerevisiae* under various perturbations, including initial population variations, chemical, physical and nutritional fluctuations and additional species invasion, and to identify key driving forces that shape ecosystem stability.

Abstract:

Microbial communities are constantly exposed to disturbances such as climate change and nutrition fluctuation in natural habitats, which shape their composition structure, population dynamics and ecological functioning. Understanding their resistance and resilience (i.e., stability) in disturbed environments is important for designing management strategies and creating artificial ecosystems. However, it remains difficult to quantitatively measure and predict due to the complexity of native ecologies. One alternative way to gain insights is to learn from well-defined, synthetic microbial consortia. Here we develop an engineered consortium of *Lactococcus lactis* and *Saccharomyces cerevisiae* that feed each other and use it to investigate the stability of symbiosis under a set of physical, chemical and biological perturbations. Our results show that the consortium is stable and converge to a single state for a wide range of initial population ratios. It also withstands upon two hours of modest pulse perturbations of pH, temperature, antibiotic and nutrition; however, under strong pulse or press perturbations, the ecosystem often transitions to an alternative stable state or even collapses depending on perturbation strength. Similarly, we observe the same characteristics of the consortium during invasion whereby the outcome is governed by specific types of perturbations introduced by foreign species. Learning from the findings, we show that modest supplement of cross-feeding metabolites confers long-term survival of community

populations during periodic dilution which crashes the ecosystem. We further extend the same cross-feeding mutualism to cases with different microbial species, demonstrating that the observed ecosystem stability arises primarily from underlying social interactions but not species. Together, our experiments provide a systematic understanding of the stability of a microbial symbiosis under pulse and press disturbances, providing insights into the organization of cooperative communities and applications of these ecosystems for biotechnological purposes.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0019185.

Constructing the Nitrogen Flux Maps (NFMs) of Plants

Hiroshi Maeda[1*](maeda2@wisc.edu), Kaan Koper[1], Rika Judd[1], Hugh Salamon[2], Yasuo Yoshikuni[2], Markus De Raad[2], Trent Northen[2], Shogo Hataya[3], Taichi Takasuka[3], Sebastian Huss[4], Zoran Nikoloski[4]

[1]University of Wisconsin-Madison, WI; [2]Lawrence Berkeley National Laboratory (LBNL), Oakland, CA; [3]Hokkaido University, Sapporo, Japan; [4]University of Potsdam & Max Planck Institute of Molecular Plant Physiology, Germany

Website: 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 limited in plants. Thus, N use efficiency (NUE) directly impacts overall yield and performance of bioenergy and agricultural crops. Improved NUE can also reduce the use of N fertilizers and environmental issues caused by N eutrophication. Despite the critical roles N play in both plant productivity and environmental health, unlike extensively-studied carbon (C) flux map of plant metabolism, little is known about how assimilated N flows through the metabolic network, namely the "**N flux map (NFM)**". **Aminotransferases (ATs)** play pivotal roles in interconnecting different branches of N metabolic pathways, but the multi-substrate specificities of ATs remain largely uncharacterized due to their poor sequence-function relationships and tedious aminotransferase activity assays[1-3]. The main objectives of this project are **to construct plant NFMs from plant genomes and to determine system-level functionality of AT enzymes and plant N metabolic network**. The obtained NFMs will provide a novel framework to advance basic understanding of plant N metabolism and facilitate rational engineering of plants having high productivity with limited N input.

To address this grand challenge, our project makes use of rapidly growing numbers of plant genomes, high-throughput functional characterization platforms, and computational modeling to deduce both biochemical and systems level functionality of ATs and NFMs. Over the past year we have made progress on each aim. In **Aim 1** to construct the framework of NFMs, we generated initial NFMs, the N atomic map of C3 plants, based on models that include primary as well as genome-scale metabolic pathways (i.e. AraCore and PlantSEED models, respectively)[4,5], which will be further expanded to include more N compound-containing metabolic pathways. In **Aim 2**, we have successfully set up high-throughput AT enzyme assay platforms using nanostructure-initiator mass spectrometry (NIMS)[6-8], and have characterized broad substrate specificity of tyrosine and tryptophan ATs from Arabidopsis. The detailed kinetic analyses of these enzymes further confirmed the findings from the NIMS assay. Additional ATs are currently being generated by various means—wheat germ *in vitro*[8,9], *E. coli*, and *Nicotiana benthamiana* expression systems—to screen their substrate specificities. These biochemical data will be used to further

refine the NFMs and also mapped onto AT phylogeny and sequence similarity network to infer AT functions from other species. In **Aim 3** for validating the NFM and using it in reaction flux estimation, ^{15}N-labeled precursor feeding has been set up using Arabidopsis hydroponic growth system, and the time-dependent incorporation of ^{15}N labeling was observed in major N-containing compounds (e.g. amino acids) were obtained. After further optimization (e.g. to determine metabolic steady state), the kinetic ^{15}N labeling data will be integrated into the N atomic map (from Aim 1) and be used to determine which reactions carry flux in specific conditions by using non-stationary metabolic flux analysis (MFA) at a genome-scale level.

The resulting NFMs will serve as a novel framework to i) elucidate how N flows through the plant metabolic network in a quantitative manner, ii) simulate how plant metabolism responds to different N availability at a systems level, and iii) identify potential targets for improving N use efficiency. We will also establish **open source public databases and pipelines** in DOE Systems Biology Knowledgebase (KBase) for other researchers to be able to predict AT functions and construct NFMs from any given plant genomes.

References

1. Maeda H., Yoo H. & Dudareva N. Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. *Nat. Chem. Biol.* **7,** 19–21 (2011).
2. Wang M., Toda K., Block A. & Maeda HA. TAT1 and TAT2 tyrosine aminotransferases have both distinct and shared functions in tyrosine metabolism and degradation in Arabidopsis thaliana. *J. Biol. Chem.* **294,** 3563–3576 (2019).
3. Wang, M. & Maeda, H. A. Aromatic amino acid aminotransferases in plants. *Phytochem Rev* 1–29 (2017).
4. Arnold, A. & Nikoloski, Z. Bottom-up Metabolic Reconstruction of Arabidopsis and Its Application to Determining the Metabolic Costs of Enzyme Production. *Plant Physiol.* **165,** 1380–1391 (2014).
5. Seaver SMD, Gerdes S, Frelin O, Lerma-Ortiz C, Bradbury LMT, et al. 2014. High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. Proc Natl Acad Sci U S A. 111(26):9645–50
6. Northen, T. R., Yanes, O., Northen, M. T., Marrinucci, D., Uritboonthai, W., Apon, J., Golledge, S. L., Nordström, A. & Siuzdak, G. Clathrate nanostructures for mass spectrometry. *Nature* **449,** 1033–1036 (2007).
7. de Raad M., de Rond T., Rübel O., Keasling J. D., Northen T. R. & Bowen, B. P. OpenMSI Arrayed Analysis Toolkit: Analyzing Spatially Defined Samples Using Mass Spectrometry Imaging. *Anal. Chem.* **89,** 5818–5823 (2017).
8. Deng, K., Takasuka, T. E., Heins, R., Cheng, X., Bergeman, L. F., Shi, J., Aschenbrener, R., Deutsch, S., Singh, S., Sale, K. L., Simmons, B. A., Adams, P. D., Singh, A. K., Fox, B. G. & Northen, T. R. Rapid kinetic characterization of glycosyl hydrolases based on oxime derivatization and nanostructure-initiator mass spectrometry (NIMS). *ACS Chem. Biol.* **9,** 1470–1479 (2014).
9. Takasuka, T. E., Walker, J. A., Bergeman, L. F., Vander Meulen, K. A., Makino, S., Elsen, N. L. & Fox, B. G. Cell-free translation of biofuel enzymes. *Methods Mol. Biol.* **1118,** 71–95 (2014).

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*

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 three new reference quality *Lemnaceae* genome assemblies of diploid *L. gibba*, diploid *W. australiana*, and allopolyploid *L. minor* clones using single molecule long-reads and Hi-C contact maps. Comparisons of the chromosome-scale assemblies reveal that the 21 chromosomes of *L. gibba* are highly colinear with the two subgenomes of *L. minor*, while *W. australiana* has 20 chromosomes like *S. polyrhiza*, yet with significant architectural differences. *S. polyrhiza*, *L. gibba*, and the B subgenome of *L. minor* all encode around 18,000 genes – significantly fewer than terrestrial monocots such as rice and Brachypodium, and comparable to the *Chlamydomonas reinhardtii*. Strikingly, *W. australiana*

has undergone a reduction to only about 14,000 coding genes. Methylome and small RNA sequencing revealed dramatic differences between the three genera consistent with known pathways of RNA directed DNA methylation. Orthologous gene analysis across the *Lemnaceae*, 10 other monocots and 7 non-monocots, revealed variations that likely account for these contrasts, as well as for reduced morphology, clonal reproduction, and aquatic habit.

The Birchler Lab has completed the construction of a transgene stacking system with alternating transformation vectors that enable iterative recombination into a locus determined by a previously integrated landing pad. This novel design is compatible with consecutive transformation of *Lemnaceae* under strictly clonal propagation. Lines bearing the landing pad have been regenerated and are being screened for single insertions. In addition, antibodies against centromeric histone H3 have been raised for four species to visualize centromere organization and identify centromeric repeat sequences across the duckweeds via CUT&RUN.

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 minor* line Lm8627. This resulted in <1% of DW of TAG along with large reduction in growth rate along and significant developmental abnormalities. We revisited our TAG producing strategy and constructed a CFP-N terminally tagged version of the Arabidopsis WRI1 under the control of inducible estradiol inducible XVE promoter. This was co-expressed with a Sesame Oleosin 1 gene variant, (ROGUE Biostems Design) in which its degradation signals had been minimized to optimize its TAG protective function and a very strong mammalian DGAT2 (CABBI Energy Center funding) with both genes under strong constitutive promoters to create Lm8627-33 transgenics. Growth of Lm8627-33 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. Thus, our inducible WRI1 strategy resulted in up to 124-fold increase in TAG in line Lm8627-33-6 four days after induction with little to no reduction in growth rate and no developmental abnormalities observed.

Funding for this project is provided by the DOE Office of Biological & Environmental Research (DE-SC0018244).

Using Gene Editing and an Accumulated Bioproduct as a Reporter for Genotypic and Phenotypic Heterogeneity in Growth-vs-Production for *Methylobacterium extorquens* Conversion of Aromatics to Butanol

Christopher J. Marx[1,PI] (cmarx@uidaho.edu), Tomislav Ticak,[1] Shahla Nemati,[1] Akorede L. Seriki,[1] Triana N. Dalia[2], Nkrumah A. Grant,[1] Isaiah D. Jordan,[1] Eric L. Bruger,[1] Andrew T. Johnson,[1] Garrett T. Woelfl,[1] Mahsan Karimi,[1] Monica J. Pedroni,[1] Sergey Stolyar,[1] Ankur B. Dalia,[2] and Andreas E. Vasdekis,[1]

[1]University of Idaho, Moscow, ID; [2]Indiana University, Bloomington, IN

https://marxlab.org/doe-biosystems-project/

Project Goals: With a unique capacity to assay growth and production – for either a tremendous number of genotypes in a mixture, or for individual cells – we will provide an unprecedented view of the critical tradeoff between growth and production. This will be used to guide development of *M. extorquens* as a novel platform for conversion of methoxylated aromatics to butanol. We will accomplish this work through the following aims:

1. Engineer/evolve improved use of methoxylated aromatics in M. extorquens

2. Explore growth-vs-production tradeoffs for genetic and phenotypic variation in PHB production

3. Combine improvements in substrate use and production capacity

4. Exchange PHB synthesis for butanol synthesis to test best genotypes

Abstract. Over the past year we have made progress on several fronts in order to move towards achieving our project's goals. First, we have developed a novel "Hi-PIE" method to introduce of genome-edited gene clusters from *Vibrio* into *Methylobacterium*. Second, we have developed the capacity to visualize PHB in unstained cells of *M. extorquens*. Third, we have developed a chromosomal system for tracking gene expression of multiple loci in single cells of *M. extorquens*. Fourth, we have completed analysis of the novel pathway for aromatic utilization that we have found in *M. extorquens* and we have evolved strains to effectively utilize vanillic acid. Fifth, and perhaps most importantly, we have uncovered that utilization of methoxylated aromatics depends critically upon the ability to handle formaldehyde toxicity, and that this capacity involves a substantial degree of phenotypic heterogeneity. All of these steps forward move us toward our ultimate goal to develop *M. extorquens* for conversion of methoxylated aromatics to butanol.

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-SC0019436.

Systems metabolic engineering of *Novosphingobium aromaticivorans* for lignin valorization

Marco N. Allemann (allemannmn@ornl.gov),[1] Gerald N. Presley,[1] Richard J. Giannone,[1] David C. Garcia,[1] Leah H. Burdick,[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 guaiacyl (G), syringyl (S), and *p*-hydroxyphenyl (H) 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.

We are identifying new aromatic degradation pathways to increase the catabolic potential of *N. aromaticivorans* F199, using a combination of barcoded transposon insertion sequencing, proteomics, and *in vitro* biochemistry. In several cases, *N. aromaticivorans* F199 is known to assimilate an aromatic compound but the pathway requires additional characterization, as we demonstrated with the aromatic monomer syringate.[1] Additionally, F199 can catabolize previously undescribed compounds, such as the β-1 linked dimer 1,2-diguaiacylpropane-1,3-diol (DGPD). We have recently identified a novel enzyme, LsdE, required for DGPD catabolism.[2] However, there are multiple aromatic compounds for which *N. aromaticivorans* F199 lacks the necessary catabolic pathway. We have previously isolated additional Sphingomonads that metabolize several of these compounds and are currently investigating the relevant pathways for transfer to F199. Of note, one of these isolates is remarkably similar at the genetic level to *Sphingobium* sp. SYK-6, a well characterized model for bacterial lignin degradation.

To further enable the genetic exploration of lignin catabolic pathways and to integrate these into novel hosts, we are developing new genetic tools for *N. aromaticivorans* F199. A broad host range plasmid that replicates in the strain has been identified. To better understand the effect of host genetic variation on pathway function, we are also adapting a newly-developed technique of bacterial quantitative trait locus (QTL) mapping. We have acquired and sequenced additional strains of *N. aromaticivorans* and demonstrated intraspecific recombination using genome shuffling. By combining novel pathway discovery with this developing genetic toolset, we can engineer *N. aromaticivorans* F199 to more efficiently valorize lignin.

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. Cecil, J. H., Garcia, D. C., Giannone, R. J. & Michener, J. K. Rapid, parallel identification of pathways for catabolism of lignin-derived aromatic compounds in Novosphingobium aromaticivorans. *Appl. Environ. Microbiol.* AEM.01185-18 (2018). doi:10.1128/AEM.01185-18
2. Presley, G. N.*, Werner, A. Z.*, Katahira, R., Garcia, D. C., Haugen S. J., Ramirez, K. J., Giannone, R. J., Beckham G. T., and Michener, J. K. Pathway discovery and engineering for cleavage of a ß-1 lignin-derived biaryl compound. *In review*.

Scalable Computational Tools For Inference Of Protein Annotation And Metabolic Models In Microbial Communities

Janaka N. Edirisinghe[1,*] (janakaed@anl.gov), Michael Shaffer[2,*] (michael.t.shaffer@colostate.edu), Mikayla A. Borton[2], Evan Stene[3], Lucia S. Guatney[3,4], Farnoush Banaei-Kashani[4], Kelly C. Wrighton[2], Christopher Henry[1], and **Christopher S. Miller**[3]

[1]Argonne National Laboratory, Argonne, IL; [2]Colorado State University, Ft. Collins, CO; [3]University of Colorado Denver, Denver, CO; and [4]University of Colorado Anschutz Medical Campus, Aurora, CO

Project Goals:
Advances in 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 of microbial communities of DOE relevance. We are developing a suite of improved computational tools for protein annotation, and integrating these tools with simultaneous advances in inference of community-level metabolic models that incorporate complex interactions between environment and microbes, and among microbial community members. By integrating these tools into the DOE Systems Biology Knowledgebase (KBase)[1], we aim to make these tools accessible to a broad user base of scientists.

Abstract:
In order to infer, understand, and model microbial and ecosystem traits and processes of relevance to biogeochemical cycling, protein and metabolite function need to be encoded, inferred, and studied in the context of community-level metabolic models. Our approach to inference of improved models relies on developing new computational tools for the microbial sciences community, in three main areas: 1) improved homology-based and non-homology-based protein annotations, 2) building an iterative cycle of gap-filling organism-level and metabolic models with improved protein annotations, and informing probabilistic protein annotations based on metabolic models, and 3) integrating improved protein annotations related to exchange metabolisms with community-level flux balance metabolic models, in order infer ecosystem-level processes and community-level interactions.

To improve the inference of protein annotations, we have recently developed a computational pipeline, Distilled and Refined Annotation of Metabolism (DRAM) that integrates annotations across sensitive sequence-homology searches from a variety of both broad and specialized databases[2]. The comprehensive approach yields an increased number of protein annotations, especially for difficult to annotate taxa with poor genomic sampling. DRAM is also unique in the synthesis it performs, returning annotations in metabolism-centric outputs and visual outputs that allow for quick comparisons of key metabolisms for expert curation. We have recently integrated DRAM into KBase as an additional annotation option, which allows for integration with a growing number of state-of-the-field Flux Balance Analysis (FBA) models in KBase.

DRAM is also being extended to automate non-homology methods common to expert curation, including using conserved gene neighborhood information across metagenome-assembled genomes for inference of protein function. Other planned non-homology methods for incorporation into a probabilistic annotation framework include co-expression across metatranscriptome, co-occurance across multiple samples, and co-occurrence with metabolite networks when metabolomics data is available.

Many of these non-homology methods require read-mapping-based counting across multiple samples. To make these approaches scalable, we are adopting a training approach for learning-enhanced read mapping. Using deep learning techniques and models borrowed from Natural Language Processing (NLP) applied to genome sequences (such as BERT[3]), we generate vectorized representations of assemblies. These same vectorizations can be generated from reads using the trained model and compared to assembly vectors as a form of mapping. The efficiency of the learned model is evaluated by its ability to match two related sequences (for example kmers from the same read, or adjacent sequences) as well as the agreement of learning-enhanced mapping and traditional read mapping tools.

The Argonne team has made significant progress towards enabling modeling tools to use improved annotations from DRAM and other algorithms. Specifically, an API was developed in KBase to support consistent reading and writing of ontology terms from KBase genomes. The KBase metabolic model reconstruction tool was adapted to permit users to build models from a variety of annotation ontologies, including EC, GO, KO, and RAST (in collaboration with the Stuart SFA). Additionally, the KBase model gapfilling tool is being adapted to permit the use of evidence from alternative annotation sources. This tool is available in prototype form in a Jupyter notebook. Finally, a new metagenome model reconstruction algorithm was added to KBase, permitting the reconstruction of metagenome models from annotated metagenome assemblies (AMA). The annotation ontology API was connected to AMA objects so alternative annotations (like DRAM) could eventually be performed on these objects as well. All of this lays the groundwork for the use of metabolic models on the levels of Metagenome-Assembled Genomes (MAGs) and entire metagenomes, to integrate multiple sources of annotation evidence and propose the most consistent metabolic representation of biological systems possible. This work builds towards our goal of community-level models for the inference of ecosystem-level processes and community-level interactions.

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;36:566–9. Available from: http://www.nature.com/articles/nbt.4163
2. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res*. 2020;48:8883–900. Available from: https://academic.oup.com/nar/article/48/16/8883/58847382.
3. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv Preprint*. 2018; Available from: http://arxiv.org/abs/1810.048053.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021350 and Systems Biology Knowledgebase (KBase) was supported by Award Numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

Discovering peptidic natural products by integrating computational mass spectrometry and genome mining

Liu Cao,[1*] Yi-Yuan Lee,[1] Bahar Behsaz,[1] and **Hosein Mohimani**[1]

[1]Computational Biology Department, Carnegie Mellon University, Pittsburgh

http://mohimanilab.cbd.cmu.edu/

Project Goals

Peptidic natural products (PNP) are a major source of signal molecules and drug leads [1]. The existing techniques for PNP discovery require isolation of bioactive molecules and structure elucidation, which are time consuming and expensive. Recent advances in high-throughput mass spectrometry (MS) and next generation sequencing has resulted in large MS/genomic datasets, which are gold mines for PNP discovery. However, currently there is no efficient algorithm to mine these datasets. We have developed computational tools to integrate MS/genomic data for automated discovery of PNPs from environmental isolates/communities.

HypoNPAtlas is a database of hypothetical natural products that is readily searchable against MS [2]. Seq2ripp predicts the structure of ribosomally synthesized and post-translationally modified peptides from microbial genome [2]. NRPminer integrates MS/genomic data to discover non-ribosomal peptides (NRP) [3]. MolDiscovery is a probabilistic model that efficiently searches small molecules MS [4]. These tools has enabled discovery of various novel PNP from public datasets. One of the NRP showed anti-parasite activity [3].

References

[1] Gordon M. Cragg, and David J. Newman. “Natural products: a continuing source of novel drug leads.” Biochimica et Biophysica Acta (BBA)-General Subjects 1830.6 (2013): 3670-3695.

[2] Yi-Yuan Lee, Haodong Liu, Neel Mittal, Liu Cao, Alexey Gurevich, and Hosein Mohimani. “hypoNPAtlas: an atlas of hypothetical natural product for mass spectrometry database search”. In preparation.

[3] Bahar Behsaz, Edna Bode, Alexey Gurevich, Yan-Ni Shi, Florian Grundmann, Deepa Archarya, Andrés Mauricio Caraballo-Rodríguez, Amina Bouslimani, Morgan Panitchpakdi, Annabell Linck, Changhui Guan, Julia Oh, Pieter C. Dorrestein, Helge B. Bode, Pavel A. Pevzner, and Hosein Mohimani. “Integrating Genomics and Metabolomics for Scalable Non-Ribosomal Peptide Discovery”. Provisionally accepted by Nature Communications.

[4] Liu Cao, Mustafa Guler, Azat Tagirdzhanov, Yiyuan Lee, Alexey Gurevich, Hosein Mohimani. “MolDiscovery: Learning Mass SpectrometryFragmentation of Small Molecules”. Submitted to Nature Communications, under revision.

Spatially resolved proteomic analysis to reveal controls on rhizosphere microbial recruitment and track carbon exchange between plants and microorganisms

James Moran[1,*] (james.moran@pnnl.gov), Vivian Lin[1], Allison Thompson[1], Ying Zhu[1], Samuel Purvine[1], Nikola Tolic[1], Anna Lipton[1], Elias Zegeye[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. High diversity and microbial activity within rhizosphere (especially relative to bulk soil) reflect the tight inter-Kingdom metabolic coupling present in this zone where interactions are linked to increased nutrient mobilization, enhanced pathogen resistance, and improved drought tolerance. Soil, as a precursor to rhizosphere development, is extremely heterogeneous with large spatiotemporal changes in biological and geochemical conditions over short distances. Rhizosphere displays comparable variability with rates and composition of root exudation, recruitment of microbial members, and overall growth or constriction reflecting shifting environmental conditions, seasonal cycles, and plant growth stage. Efforts to both improve biomass production on marginal soils and increase subsurface C storage are motivating a need to better understand and engineer specific rhizosphere interactions with a suite of resulting spatially resolved techniques being developed to better interrogate this system. *Here, we describe and demonstrate a spatially specific, non-destructive proteomic technique that enables evaluation of microorganisms present and metabolically integrated with a rhizosphere.* For instance, we use this approach to identify microbial recruitment associated with *de novo* rhizosphere microbial community development by comparing nascent rhizosphere associated with root tip to communities aligned with more mature root segments.

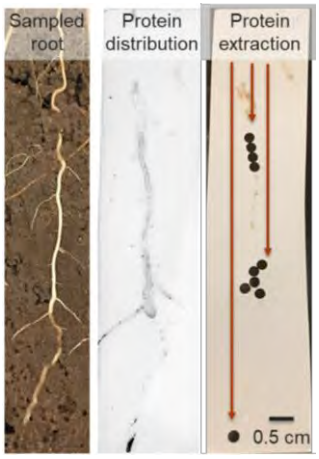


Figure 1: Sample selection method involved direct blotting of rhizosphere with a membrane for spatially resolved, nondestructive protein extraction followed by staining to map protein distribution and excision of manually selected regions for analysis.

Our method employs a membrane blotting technique for nondestructive removal of mobile proteins from rhizosphere

(Figure 1). The proteins are largely immobilized when bound to the membrane, enabling retention of their two-dimensional distribution in the sample and can then be imaged/mapped using a fluorescent, general protein stain (SYPRO Ruby). Membrane sections correlating to sample areas of interest (e.g., high protein abundance, specific root morphology, alignment with rhizosphere geochemical gradient, etc.) are manually excised from the membrane and bound proteins are digested and purified in preparation for proteomic analysis using Orbitrap mass spectrometry. Importantly, the approach is nondestructive in that, while mobile proteins are removed from the system, the host plant, rhizosphere, and soil are not removed, making both timeseries investigations and analysis using additional tools feasible (e.g., imaging, stable isotope analysis, nutrient quantification, etc.). The demonstrations we performed all used switchgrass (*Panicum virgatum*, Cave-in-rock) cultured in rhizoboxes containing soil harvested from the Kellogg Biological Station (Hickory Corners, MI).

We used comparison of proteomic assessment of rhizosphere at the tip of a growing root with that associated with a more mature rhizosphere as an initial demonstration of this technique. The relatively rapid growth rate of root tips can result in them spatially outpacing microbial colonization. Yet, due to multiple interactions between growing root tips and soil, these areas are also hotspots for deposition of plant derived C which can help fuel establishment of rhizosphere microbial communities. Our analysis (Figure 2) demonstrated dominance of switchgrass associated proteins at the root tip with observation of a relatively smaller composition of bacterial proteins. In contrast, rhizosphere associated with the mid root had overall increase alpha diversity and a dominance of bacterial taxa at the expense of switchgrass related proteins.

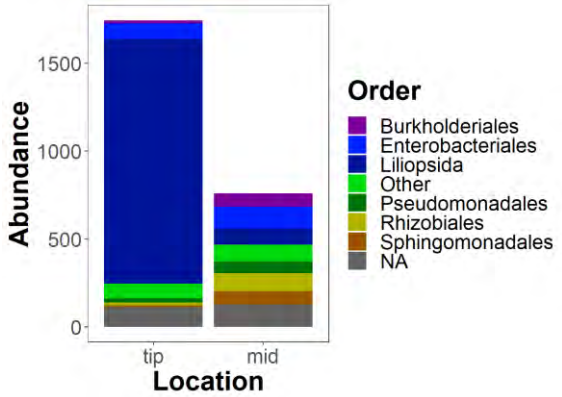


Figure 2: Proteins associated with switchgrass dominated samples from the root tip while microbial diversity in the mid root samples is indicative of establishment of the rhizosphere community.

We further used this approach to track differential protein signatures associated with changes in P availability in a series of split root rhizoboxes where roots from individual plants could grow into both P replete and deplete conditions. We incorporated a $^{13}CO_2$ stable isotope tracer into this experiment to better evaluate the exchange of plant derived carbon to the microbial community. Surprisingly, there was no large-scale preferential incorporation of ^{13}C under either set of conditions; roughly twelve percent of observed proteins were identified as containing tracer in each condition. This is consistent with bulk scale measurements suggesting commensurate levels of root exudation in each condition. However, all observed ^{13}C labeled switchgrass associated phosphatases were in P replete environments. Thus, while these results do not suggest a wholesale shift in release of plant root exudates based on P availability, we did observe indications of a spatial physiological response in switchgrass. In contrast, we identified a larger number of bacterial phosphate transporters in P deplete conditions. While switchgrass had access to ample P in one portion of its root network, the bacterial members in the P deplete environment had limited P availability which may account for the observed increased frequency of associated transporters.

Microbial Osmotolerance Mechanisms in Hydraulically Fractured Shale Elucidated Through Metagenomics Analysis

Jishnu Adhikari[1*] (Jishnu.Adhikari@unh.edu), Fabrizio Colosimo[1], **Paula J. Mouser[1]**

[1]Department of Civil and Environmental Engineering, University of New Hampshire, Durham, New Hampshire 03824, United States

Project Goals:
The process of extracting natural gas from deep shale requires the injection of large volumes of fluid and proppants. These fluids interact with the shale matrix and return to the surface with distinct biogeochemical changes. As the salinity of produced fluids approaches equilibrium with that of the shale formation a low diversity microbial community of halotolerant fermenters and methanogens is established. The goal of this project is to improve our understanding of the growth and adaptations of the microbial community in the hydraulically fractured wells. We have identified three specific objectives to understand the in situ physiologies and kinetic rates governing biogeochemical reactions in these microbial communities: (1) characterize variables influencing growth parameters and membrane features of shale taxa, (2) characterize interactions between shale matrices and microorganisms, and (3) elucidate engineered and environmental processes driving biogeochemical signatures at field scale.

Abstract
Previous research has shown salinity is a major driver of microbial community composition and function in hydraulically fractured shales, with produced fluids from shale wells increasing to brine level salinities during the first six months of natural gas production. In this study, we investigated the mechanisms for osmotolerance in bacterial metagenomes from hydraulically fractured shale wells in the Appalachian Basin (Marcellus Shale and Utica-Point Pleasant Formation). With increasing salinity and decreased diversity, we identified three osmotolerance adaptations in the microbial community. Genes for Na^+/K^+ transport increase in relative abundance with time of production as do genes for osmolyte import/biosynthesis. This indicates that the microbial community uses a combination of a salt-in strategy, where cells maintain osmotic balance by K^+ importation coupled to Na^+ export, and a compatible-solute strategy where cells gain osmotolerance via importation/synthesis of organic osmolytes such as glycine, betaine and ectoine for osmotolerance. A second finding in the metagenome sequences is the increase in gene abundance for sodium-dependent respiration with increasing salinity. A higher Na^+ gradient outside the cell results in a sodium-tight membrane, and enables sodium-dependent ATP synthesis in some taxa. This research provides an improved understanding of genomic osmoprotectant capabilities, including membrane associated osmoprotectant proteins in deep shale microbial communities during high salinity gradients, with implications on halotolerant bacterial energetics in this and other hypersaline environments.

Funding statement.
This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DESC0019444.

'Omics Analyses of the Hydraulically Fractured Shale Isolate *Halanaerobium* Highlights Membrane Modifications that Underpin Adaptation Under Deep Subsurface Biogeochemical Drivers

Fabrizio Colosimo,[1*] (Fabrizio.Colosimo@unh.edu), Allison R. Wong,[2] Elizabeth K. Eder,[2] Heather M. Olson,[2] Samuel O. Purvine,[2] Jennifer E. Kyle,[2] Rosalie K. Chu[2] David W. Hoyt,[2] and **Paula J. Mouser[1]**

[1]University of New Hampshire, Durham, NH[2] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

Project Goals:

The injection of fluids and proppants to fracture the deep shale introduces microbial cells and substrates to low-permeability rocks. Microorganisms in hydraulically fractured wells govern biogeochemical reactions and often produce acids and sulfides, leading to corrosion and gas souring, and form biofilms, resulting in clogging and fouling events. The overarching goal of this research is to advance our comprehension of the microbial diversity and function in non-sterile hydraulically fractured wells. Our current understanding of microbial growth within fractured hydrocarbon-bearing rock is based primarily on genomic information, we identified three specific objectives that will shed light on in situ physiologies and kinetic rates, governing biogeochemical reactions: (1) characterize variables influencing growth parameters and membrane features of shale taxa, (2) characterize interactions between shale matrices and microorganisms, and (3) elucidate engineered and environmental processes driving biogeochemical signatures at field scale.

Abstract

The Gram-positive *Halanaerobium* spp. is a dominant bacterial genus across geographically distinct fractured shale formations, which are increasingly used for natural gas extraction. These bacteria encounter harsh physicochemical conditions in the deep terrestrial biosphere, including high temperatures, brine-level salinities, anoxia, and elevated pressures. Microbial membranes act as the first line of defense against these environmental stressors, and maintaining membrane functionality during changing environmental conditions requires careful regulation of intact lipid composition and membrane-embedded proteins. Here we characterized the physiology of the model organism *Halanaerobium congolense* WG10, grown for the first time under continuous culture (chemostat) conditions. The anaerobe *H. congolense* WG10 was cultivated at 20% salinity under three growth rates (hydraulic retention times (HRTs) of 48, 24, and 19.2 hrs) and two temperatures (25ºC and 40ºC) under complete control of system pH and redox conditions using a 1-L Sartorius Biostat® Q-plus system. We applied an integrative 'omics approach to characterize metabolomic, proteomic, and lipidomic features (MPLEx analysis) and quantify metabolite production (using 1H-NMR and GC-FID) under steady state growth rates.

'Omics analysis of metabolites, proteins and lipids for steady state cells revealed that *Halanaerobium* alters its cell membrane to maintain fluidity in the lipid bilayer while maximizing growth rate.

Halanaerobium modulates the ratio of phospholipid headgroups in response to changes in temperature, and increases cardiolipin and phosphatidylethanolamine polar lipid abundance with increasing carbon availability. We also observed higher abundance of neutrally charged simple glycerol lipids at lower temperature and growth rates, while glycerophospholipids with larger polar heads and zwitterionic lipids prevail at warmer temperatures and faster growth rates. The observed accumulation of cardiolipin and phosphatidylethanolamine at higher temperature and growth rates is likely to results in the destabilization *of H. congolense* WG10 membrane, and the formation of localized, reversible HII phases in its membrane. On the other hand, at sub-optimal temperature for growth, *H. congolense* WG10 responds with the stabilization of the cytoplasmic membrane by a better compaction/higher packing of the phospholipid species within the bilayer leaflets.

Proteomics analysis identified a total of 2,227 out of 2,800 predicted protein-coding genes. Among these, 356 proteins were found to be significantly higher in abundance in one or more treatments (Student's t test, $p < 0.05$), including known stress regulators involved in cellular envelope homeostasis such as cold shock proteins (CspA), a typA, bipA GTP binding protein involved in stress response, and a nucleotide-binding universal stress protein (UspA). We also identified lipid-A synthesis proteins at lower temperature or high growth rate, a lipopolysaccharide endotoxin uncommonly found in Gram-positive bacteria.

Both proteomic and metabolic data supported significant activity for the utilization of 1,3-propanediol, especially in warmer temperatures (40ºC) and longer HRTs (24 and 48 hrs). Proteins associated with the methylglyoxal bypass pathway (e.g. glyoxalase) and two subunits of the propanediol dehydratase (PduD, PduE), which catalyzes the formation of propionaldehyde from 1,3-propanediol, were important during lower temperature growth (25ºC), suggesting this pathway is activated under cold stress. The propanediol dehydratase is a cobamide-dependent enzyme that has been shown to also dehydrate ethylene glycol (and other ethoxylate-based surfactants) to acetaldehyde. The formation of both propionaldehyde and acetaldehyde was confirmed with GC-FID analysis. In addition to aldehydes, we identified ketones (acetone), volatile fatty acids (acetate, lactate, formate), alcohols (ethanol, propanol), and amino acids (alanine, valine) metabolites via 1H-NMR and/or GC-FID.

Collectively, our 'omics continuous culture approach sheds new light on the metabolism and membrane features of the halotolerant bacterium *Halanaerobium* under biogeochemical drivers relevant to engineered shale, with implications on membrane charge, permeability, and metabolism.

Funding statement.
This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DESC0019444.

The Plasminogen-Apple-Nematode (PAN) domain is a key negative regulator of Jasmonic Acid and Ethylene-mediated defense responses in plants.

Kuntal De[1], Debjani Pal[1], Kai Feng[1], Timothy B. Yates[1], Carly Shanks[1], Kuntal De[1], Debjani Pal[1], Jing Hou[1], Sara Jawdy[1], Lee Gunter[1], Steven Lebreux[1], Jin Zhang[1], Kate Stuart[1], Stephen P. DiFazio[2], Lawrence Smart[3], **Wellington Muchero[1]* (mucherow@ornl.gov)**

[1] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2] Department of Biology, West Virginia University, Morgantown, WV; [3] Horticulture department, Cornell University, Geneva, NY

Project goals: This project seeks to elucidate the molecular basis of negative host defense regulation during establishment of host-microbe interactions in the genus *Salix*

Negative regulation of Jasmonic Acid and Ethylene mediated defense responses by G-LecRLKs in plants during host-cell invasion by microbes has been reported in *Populus* during infection with the fungal pathogen *Sphaerulina musiva*[1], parasitism of *Arabidopsis* by nematodes[2], and engineering of *Arabidopsis* into a host of the fungal symbiont *Laccaria bicolor*[3]. G-LecRLKs are the only Plasminogen-Apple-Nematode (PAN) domain carrying proteins in the plant kingdom and have been reported to mediate self-incompatibility during pollination as well[4]. Across eukaryotes, immunosuppression is an essential biological phenomenon for gamete fertilization, cell growth and proliferation during organismal development. Here, we propose that the PAN domain, comprised of a core of highly conserved amino acid residues, is a unifying feature that is found in association with proteins involved in immunosuppression across highly divergent organisms. We present biochemical and molecular genetic evidence that mutating conserved amino acid residues restores defense signaling in both *Arabidopsis* and tobacco overexpressing *Salix*-derived G-LecRLKs.

References:

1. Muchero, W., Sondreli, K. L., Chen, J. G., Urbanowicz, B. R., Zhang, J., Singan, V., ... & Yang, J. Y. Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant–pathogen interactions in a tree. *Proceedings of the National Academy of Sciences*, *115*(45), 11573-11578 (2018).
2. Kaloshian, I., & Teixeira, M. A. *U.S. Patent Application No. 15/799,191* (2018).
3. Labbé, J.*, Muchero, W*., Czarnecki, O., Wang, J., Wang, X., Bryan, A. C., ... & Wang, D. Mediation of plant–mycorrhizal interaction by a lectin receptor-like kinase. *Nature Plants*, *5*(7), 676 (2019).
4. Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A., & Hinata, K. The S receptor kinase determines self-incompatibility in Brassica stigma. *Nature*, *403*(6772), 913 (2000).

Acknowledgement: This research was supported by the United States Department of Energy's Office of Science Early Career Research Program under the Biological and Environmental Research office.

Multifactorial Nutrient-Dependent Proteomics Elucidates Lipid Accumulation and Regulation of Photosynthesis

Tim L. Jeffers*[1] (tim_jeffers@berkeley.edu), Sam Purvine[2], Carrie Nicora[2], Sean D. Gallaher[3], Ryan McCombs[1], Sabeeha S. Merchant[1,4], Mary Lipton[2], **Krishna K. Niyogi**[1,4,5], Melissa S. Roth[1]

[1]Department of Plant and Microbial Biology, University of California, Berkeley; [2]Pacific Northwest National Laboratory; [3]UCLA DOE Institute for Genomics and Proteomics, University of California, Los Angeles; [4]Lawrence Berkeley National Laboratory; [5]Howard Hughes Medical Institute, University of California, Berkeley

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. Enabled by cutting-edge synthetic biology and genome-editing tools, we will integrate the systems data in a predictive model that will guide us in redesigning and engineering the metabolism of *C. zofingiensis*. Here, we focus on elucidating nutrient-dependent effects 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 enable bioengineering of microalgae to maximize production of biofuels and bioproducts.

Nutrient regulation of photosynthesis and lipids can be leveraged for enhancing biofuel production. The oleaginous green alga *Chromochloris zofingiensis* has a reversible nutrient-dependent switch for photosynthesis and lipid accumulation (1) that is dependent on hexokinase (HXK1) (2). Here we show that this photosynthetic switch is dependent on iron and that replete iron supplementation can activate photosynthesis in the presence of photosynthesis-repressing glucose. To elucidate regulation of trophic states and the accumulation of triacylglycerols (TAGs), we conducted a full combinatorial proteomic analysis of the wild type vs. two independent *hxk1* mutant strains grown with and without glucose (+Glc) and in iron-replete (+Fe) and iron-limiting (-Fe) conditions (n = 3-4, 47 total samples). Isobarically labelled proteomes were analyzed via mass spectrometry, detecting peptides corresponding to ~80% of protein-coding gene models, with 52% being detected in all 12 conditions. A targeted analysis revealed 116 proteins that are highly induced by TAG-accumulating conditions (WT+Glc), capturing most *de novo* fatty acid biosynthesis enzymes, lipid droplet components, and uncharacterized proteins that are novel candidates for engineering increased TAG accumulation in autotrophic cells (**Figure 1**). We also found 782

proteins specifically responsive to the heterotrophic state (WT-Fe+Glc), including ~88% of the photosynthetic electron transport chain (ETC) subunits, which were depleted when photosynthesis was switched off. The proteomic data are consistent with a physiological explanation of iron's role in regulating photosynthesis, showing prioritization of the Fe-rich respiratory ETC over the photosynthetic ETC and revealing a putative allocation of Fe to the mitochondria by organellar transporters. This systems-level factorial experiment confirms the power of using *C. zofingiensis* for gene discovery related to nutrient physiology, photosynthesis, and bioengineering targets for metabolic redesign for sustainable algal 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.

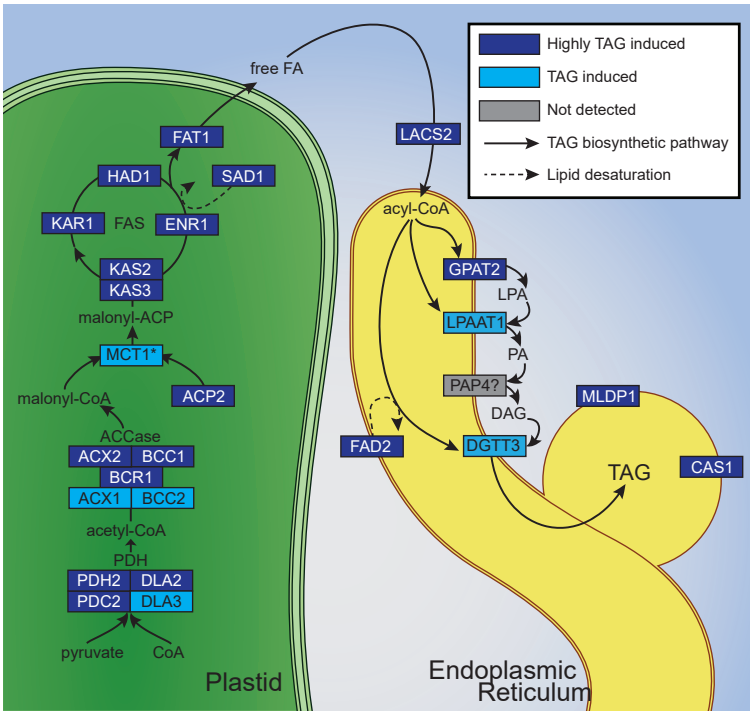


Figure 1. Known players involved in TAG accumulation are enriched in WT+Glc proteomes. A concerted upregulation of the *de novo* fatty acid biosynthesis proteins occurs in the plastid in response to +Glc and is dependent on HXK1. Enzymes of the Kennedy pathway and lipid droplet biogenesis that act in the endoplasmic reticulum to deposit storage lipids are also induced in the TAG-accumulating conditions. This figure also includes highly induced players in lipid desaturation.

Metabolic Model of *Chromochloris zofingiensis*, an Emerging Model Green Alga for Sustainable Fuel Production

Michelle F. Meagher (mmeagher@mymail.mines.edu)[1*], Yuntao Hu[2,3], Melissa S. Roth[3], Alex J. Metcalf[1], **Krishna Niyogi[3]** ,Trent R. Northen[2,4], and Nanette R. Boyle[1]

[1]Colorado School of Mines, Golden CO, [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]University of California, Berkeley, CA; [4]Joint Genome Institute, Berkeley, CA

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.* The Boyle group is focused on the development of computational models of cellular metabolism in this organism to study carbon partitioning on a genome wide scale. The Northen group is focused on the exometabolomics and metabolomics analysis of the *C. zofingiensis* to help us understand the algal nutrient preference and metabolism during trophic transitions.

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.

To create a genome-scale metabolic model of this organism, we implemented a new computational algorithm, Rapid Annotation of Photosynthetic Systems[2] (RAPS). This model was used to perform flux balance analysis (FBA) of this organism with experimentally determined constraints for photoautotrophic and heterotrophic growth on glucose. The results of these FBA studies show highly variant metabolic flux distributions for these different growth conditions. Using experimentally determined constraints of nutrient uptake, it was found that glucose consumption

was higher than that required to match experimental growth rate data. Model simulations predicted the excretion of lactate, among other products, as an outlet for the carbon imbalance. We were able to validate this assumption with metabolomics analysis.

To gain 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. Our results showed that *C. zofingiensis* secreted more diverse exometabolites during its heterotrophic growth stages when supplemented with glucose. Although *C. zofingiensis* has a clear nutrient (carbon/nitrogen source) preference order, it will recapture its secreted metabolites when the exogenous glucose is limited. 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. We are now incorporating the exometabolomic and metabolomic data to improve our understanding of algal metabolism and nutrient demand. The metabolic model that has been developed, used in combination with this extensive data set, has great potential to elucidate dramatic metabolic shifts within the organism and 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. 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

Structural Gene Organization in *Chromochloris zofingiensis* Can Drive Advancements in Synthetic Biology

Jeffery M. Moseley,[1*] (jlmoseley@berkeley.edu), Sean D. Gallaher,[2*] Sean R. McCorkle,[3] Crysten E. Blaby-Haas,[3] and **Sabeeha S. Merchant**[1]

[1]Quantitative Biosciences Institute, Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, [2]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90035, [3]Biology Department, Brookhaven National Laboratory, Upton, NY 11973

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 large-scale multi-'omics systems analysis to understand the genomic basis for energy metabolism partitioning as a consequence of carbon source. We are integrating the systems data in a predictive model that will guide the redesign and engineering of metabolism in *C. zofingiensis.* Toward our goal, we are also leveraging fundamental discoveries in green algal genomics and transcriptomics to build new synthetic biology tools for the design and synthesis of engineered bioproduction pathways.

In contrast to prokaryotes, where functionally cooperating proteins are often encoded by operons and co-transcribed, such structural organization had been thought to be rare in most eukaryotes. However, as the number of sequenced eukaryotic genomes and transcriptomes has increased, and the function of those encoded proteins has been revealed, non-random gene organization (i.e. physical clustering of pathway members and co-regulated genes) is emerging as a characteristic of eukaryotic genomes. Current methods for identifying functionally cooperative gene neighborhoods in eukaryotes depend on the availability of functional annotations, which limits our ability to identify clustered functional gene units in algal genomes. Over half of algal proteins are of unknown function, and many of the remaining genes may be mis-annotated, or have vague functional assignments. A related challenge is the quality of structural annotations that are needed to predict coding regions and serve as the input for downstream comparative genomic analyses.

To address these concerns in the oleaginous green alga, *Chromochloris zofingiensis,* we recently embarked on an effort to re-annotate the genome using single-molecule, long-read sequencing of whole transcripts on the PacBio platform (Iso-Seq)[1]. Unexpectedly, this effort revealed hundreds of examples where two, three, or more genes were exclusively co-transcribed on polycistronic transcripts. In a survey of seven green algal species representing the breadth of the chlorophyte lineage, we observed that many of these polycistronic operons were evolutionarily conserved among species that had diverged hundreds of millions of years ago. Informed by these discoveries,

we designed synthetic bicistronic constructs that were capable of co-expressing pairs of proteins, including reporter and selectable proteins, in vitro. These findings provide an opportunity to build synthetic biology tools for designing optimized algal strains. Here, we present our progress on using highly conserved polycistronic mRNAs as scaffolding for algal engineering.

C. zofingiensis has a robust cell wall that complicates efforts at engineering the species to express transgenes. Given that many of the polycistronic loci we identified in *C. zofingiensis* were widely conserved in Chlorophytes, we used known polycistronic loci from that species to identify candidates in a more genetically tractable species: the oleaginous Trebouxiophyte alga, *Auxenochlorella protothecoides* (strain Cp0710). In a preliminary survey we found four potential polycistronic genes that are conserved between *A. protothecoides*, *Chlamydomonas reinhardtii* and *C. zofingiensis*. Iso-Seq analysis of *A. protothecoides* transcription has not been reported, but mapping of RNA-seq reads to these four loci was consistent with the expression of polycistronic transcripts. We have demonstrated transformation and gene targeting by homologous recombination in another *A. protothecoides* strain, UTEX 250, and we employed this platform to test whether inter-ORF sequences from the putative *A. protothecoides* polycistronic genes could be used for in vivo transgene expression. An artificial polycistronic gene using the *A. protothecoides* TOM22-SDHAF3 homolog inter-ORF sequence to link synthetic ORFs encoding Arabidopsis THIC (hydroxymethyl pyrimidine synthase) and Chlamydomonas BKT1 (beta-carotene ketolase) was targeted to the lycopene cyclase epsilon (LCYE) locus in *A. protothecoides* UTEX 250. Transformants were selected on the basis of gain-of-function thiamine prototrophy, requiring THIC activity. The orange color of the transformants, resulting from synthesis of red keto-carotenoids, indicated that BKT1 was functional as well. Future experiments will explore the use of artificial polycistronic genes to alter fatty acid and lipid composition.

[1] Gallaher, S.D. *et al.*, Widespread polycistronic gene expression in green algae. *Proc. Natl. Acad. Sci. U.S.A.,* in press (2021).

This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), award number DE-SC0018301 and a FICUS JGI/EMSL project, award number 503551.

The Filamentous Fungus *Trichoderma atroviride* as a Model System for Understanding Fungal Genetics, the Plant-Fungal Symbiosis, and Interactions With Diverse Bacteria

Catharine A. Adams[1,2] (catadams@berkely.edu), José Manuel Villalobos Escobedo[1,2], Ran Shi[1], Hsiao-Han Lin[2], Trenton K. Owens[2], Jens Heller[2], Dawn M. Chiniquy[2], Jenny C. Mortimer[2,3], **N. Louise Glass[1,2] (lglass@berkeley.edu),** Adam M. Deutschbauer[2], and Trent R. Northen[2]

[1]University of California, Berkeley; [2]Lawrence Berkeley National Laboratory, Berkeley; [3]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.

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. However, much remains unexplored regarding the gene function of different *Trichoderma* species, their association with plants, and their interactions with other rhizosphere bacteria and fungi. Here we take three separate but complementary approaches to understand *Trichoderma* and its ecological role in the rhizosphere.

Traditional methods of assigning gene function are tedious and slow. To expedite this process, we are leveraging Random Barcode Transposon-site Sequencing (RB-TnSeq) technology to identify gene function in *T. atroviride* IMI in a massively parallel fashion. We currently have a ~140 million uniquely bar-coded plasmid library in *Agrobacterium tumefaci*ens with selection for transformation into *T. atroviride.* We anticipate that we will have an insertional *T. atroviride* library in 2021 for utilization in fitness experiments with different stressors, carbon and nitrogen sources, as well as interactions with other microbes and plants. Our ultimate goal is to develop RB-TnSeq libraries for different species of *Trichoderma* and other rhizosphere-associated fungi to investigate the function of genes within these organisms in shaping rhizosphere communities and plant interactions.

We are also investigating how *Trichoderma* associates with the model temperate grass *Brachypodium distachyon*, in the EcoFAB. The EcoFAB, or fabricated ecosystem, is an easily

constructed chamber that allows for overt control of the microbial community. In addition, we will use a variant of the EcoFAB specifically designed to separate plant roots from a given nutrient source, the so called MycoFAB, which enables the exploration of how fungi access nutrients to trade with their plant hosts and microbial neighbors. By preventing roots from accessing essential nutrient sources, such as insoluble phosphates, plants can be reliably inoculated with fungi such as *Trichoderma*, or even arbuscular mycorrhizal fungi, which are known to access these nutrients. For EcoFAB experiments, seeds of *B. distachyon* are germinated on filter paper for approximately 4 days, and the roots are subsequently dipped in a suspension of fungal conidia, and transferred into the EcoFAB, which is housed in a Magenta box to retain axenic conditions. Our preliminary results indicate that substrate strongly influences the directionality of the association: *T. atroviride* had the most detrimental effect on *B. distachyon* when grown in a hydroponic system, but when grown in sand, the fungus trended toward having a positive impact on the plant. Next steps include exploring how seedling age, watering regime, and nutrient availability shift the plant-fungal symbiosis. In addition, we will evaluate interactions with *B*. *distachyon* and our synthetic bacterial community (syncom) with wild type and available *T. atroviride* mutants predicted to disrupt plant-fungal interactions.

Lastly, we are exploring how TnSeq can be used to study microbial species interactions. Previous studies have demonstrated that *Trichoderma* species can exert strong antagonistic effects on a diverse range of microbes. Preliminary data using spent media experiments from *Trichoderma* growth and subsequent exposure to bacterial TnSeq libraries showed both a decrease and enrichment for specific bacterial metabolic pathways. We will compare the effect of *T. atroviride* IMI and selected *T. atroviride* mutants on rhizosphere plant growth promoting bacteria (PGPB), plant pathogens, and non-plant associated bacteria, to determine whether interactions are broadly inhibitory or more specifically inhibitory against a particular species.

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: TEAMS: Advancing microbiome science through high-throughput, automated EcoFAB experiments

Authors: Peter Andeer,[1*] (pfandeer@lbl.gov), Arunima Bhattacharjee,[3] John Vogel,[2] Kirsten Hofmockel,[3] and **Trent Northen[1,2]**

Institutions: [1]Lawrence Berkeley National Laboratory, [2]Joint Genome Institute, [3]Pacific Northwest National Laboratory

Website: https://ecofab-teams.lbl.gov/

Project Goals: Short statement of goals. The vital role soil and plant microbiomes play in ecosystems is increasingly apparent, yet the lack of standardized and reproducible experimental systems represents a major challenge for microbiome science. Fabricated ecosystems (EcoFABs) are sterile devices that provide unique capabilities in the control and measurement of simplified microbial communities with tremendous potential to advance a mechanistic understanding of soil and plant microbiomes. The Trial Ecosystems for the Advancement of Microbiome Science (TEAMS) project is creating, validating, and disseminating EcoFAB technologies, complete with standardized model microbial communities, that are tailored for users of DOE's unique resources and user facilities such as EMSL, JGI, KBase, and NMDC, as well as other DOE stakeholders. EcoFAB technology is being distributed as EcoFAB kits that have been validated through multi-laboratory reproducibility studies, through development the EcoBOT which automates EcoFAB experiments for the highest degree of standardization, and online resources. Together, these standardized and reproducible experimental capabilities will help advance mechanistic microbiome science.

Abstract text: We have developed a new EcoFAB device, 'EcoFAB 2.0' to facilitate widespread EcoFAB studies including those performed on secure, automated platforms. EcoFAB 2.0 is largely constructed out of injection molded polycarbonate and has the same footprint as a 96 well plate that makes it compatible with robotics platforms and many instruments. Like the original EcoFAB devices[1], the EcoFAB 2.0 has a root/soil growth area that allows for visualization of the plant roots and rhizosphere including microscopic analyses to track fine changes in root morphology and microbial colonization. Injection molding provides improved reproducibility (vs. 3D printing), enables use of higher performance materials, and can easily produce parts for thousands of devices. In addition to these features, EcoFAB 2.0 also has a built-in plant chamber with external sampling ports so the plant remains in a sterile environment during imaging and collection of plant exudates. EcoFAB 2.0 and a customized EcoFAB developed at PNNL to provide microenvironments will be used in a series of multi-laboratory microbiome ring-trials to assess the reproducibility of observations made when studying defined microbial communities. The TEAMS project is also developing the EcoBOT, an automated platform for high-throughput, standardized EcoFAB studies. The EcoBOT integrates a liquid handling unit with a plant growth chamber that will house over 150 EcoFAB devices, an

epifluorescence, inverted microscope and a hyperspectral imager. The EcoBOT will accelerate mechanistic plant microbiome studies by enabling highly standardized EcoFAB experiments.

References/Publications

1. Gao J, Sasse J, Lewald KM, et al. Ecosystem Fabrication (EcoFAB) Protocols for The Construction of Laboratory Ecosystems Designed to Study Plant-microbe Interactions. J Vis Exp. 2018;(134):57170. Published 2018 Apr 10. doi:10.3791/57170

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.

Title: Generating a Stable, Reproducible Rhizosphere Microbial Community

Authors: Joanna Coker[1]* (jkcoker@health.ucsd.edu), Kateryna Zhalnina[2], Clarisse Marotz[1], **Karsten Zengler[1]**, Trent Northen[2]

Institutions: [1]University of California, San Diego, [2]Lawrence Berkeley National Laboratory

Website: https://ecofab-teams.lbl.gov/

Project Goals: Short statement of goals. The vital role soil and plant microbiomes play in ecosystems is increasingly apparent, yet the lack of standardized and reproducible experimental systems represents a major challenge for microbiome science. Fabricated ecosystems (EcoFABs) are devices that provide unique capabilities in the control and measurement of simplified microbial communities with tremendous potential to advance a mechanistic understanding of soil and plant microbiomes. The Trial Ecosystems for the Advancement of Microbiome Science (TEAMS) project is creating, validating, and disseminating EcoFAB technologies, complete with standardized model microbial communities, that are tailored for users of DOE's unique resources and user facilities, such as EMSL, JGI, KBase, and NMDC, as well as other DOE stakeholders. EcoFAB technology is being distributed as EcoFAB kits that have been validated through multi-laboratory reproducibility studies, through development of the EcoBOT, which automates EcoFAB experiments for the highest degree of standardization, as well as through online resources. Together, these standardized and reproducible experimental capabilities will help advance microbiome sciences.

Abstract text: Microbial soil communities form commensal relationships with plants to promote the growth of both parties. Optimization of plant-microbe interactions to advance sustainable agriculture is an important field in agricultural research. However, investigation in this field is hindered by a lack of model microbial community systems. Here, a model community of 17 soil organisms, isolated from the switchgrass rhizosphere, has been developed and optimized for use with EcoFAB devices. EcoFAB devices allow reproducible research in model plant systems, with precise control of environmental conditions and easy measurement of plant metrics. This model rhizosphere community grows reproducibly *in vitro* between replicates and experiments. Highest community alpha diversity is achieved with low-nutrient media and starting composition ratios adjusted for the growth rate of individual organisms. Controlled printing of cells at picoliter scale enables exact cell number ratios from the start. Community cryopreservation with glycerol allows robust regrowth of the community from frozen stocks, allowing for dissemination of communities. Our results demonstrate the generation of a stable rhizosphere microbial community that can be used with EcoFAB devices and shared between research groups for maximum reproducibility. This community will be used as part of an international ring-trial experiment examining the reproducibility of plant host selection of this model rhizosphere community within standardized EcoFAB devices.

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.

Techniques for *in situ* DNA delivery and targeted editing within microbial communities

Benjamin M. Rubin[1], **Spencer Diamond[1] (sdiamond@berkeley.edu**), Brady F. Cress[1], Matthew Nethery[2], Claudio Hidalgo[2], Alexander Crits-Christoph[1], Harida Shivram[1], Trenton K. Owens[3], Netravathi Krishnappa[1], Rodolphe Barrangou[2], Jillian F. Banfield[1], **Jennifer A. Doudna[1*](doudna@berkeley.edu),** Adam M. Deutschbauer[3], 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.

Abstract

Knowledge of microbial gene function classically is derived from modifying the DNA of individual microbial strains in isolation from their natural environments and communities. Currently the ability to directly characterize individual microbial strains in the context of a natural microbial community, as well as assay their individual impacts on these communities using genetic experiments is limited. This is a highly limiting hurdle for the forward advancement of microbiology as most microorganisms naturally exist in complex communities, the vast majority of microbial species are difficult to or have not been cultivated, and the true behaviors of individual microbial cells as well as communities as a whole can likely only be assayed by interrogating cells in these natural contexts. Here we present a suite of integrated methods that allow for the *in situ* assay of genetic tractability of microbes to specific DNA delivery methods, targeted and locus specific delivery of DNA, and genetic modification, and application of targeted phages to specific microbial species existing in a community context. Additionally, we demonstrate a number of practical applications of these techniques by directly assaying and genetically perturbing specific microbial species directly within a synthetically assembled microbial consortia.

To approach the direct perturbation of individual microbes within communities we are developing three complementary isolation and selection independent approaches that address the following hurdles: (i) Isolation independent determination of genetic tractability; (ii) Targeted and isolation independent delivery of genetic cargo; (iii) High penetrance editing and species specific removal using engineered phages. To address the first hurdle we have developed a novel technique, Environmental Transformation Sequencing (ET-Seq), in which barcoded non-targeted transposons are delivered to a mixture of microbial cells, and discrete integration events are quantified without selection via sequencing, using three different DNA delivery approaches (conjugation, electroporation, and natural transformation without induced competence). Next, we present our development of a DNA-editing All-in-one RNA-guided CRISPR-Cas Transposase (DART) system that allows for the species and locus specific delivery of DNA into organisms identified as tractable by ET-Seq. Finally, given the low penetrance of DNA edits using classical transformation techniques we have developed engineered phage systems as repurposed nucleic acid delivery vectors for strain specific ablation in a dynamic community context.

To demonstrate the practical applications as well as synergistic combinations of these technologies we used ET-Seq to detect and quantify targeted site-specific integrations of DNA cargo delivered to a specific member of a microbial consortia with our DART vector without selection or isolation. Subsequently we demonstrate that this targeted quantification can be used to assay gene fitness within a microbial consortia by using DART to target and disrupt the *pyrF* locus in *Klebsiella michiganensis*, which is known to confer a fitness advantage in the presence of 5-fluoroorotic acid (5-FOA). Quantification of the *pyrF* inserts using ET-Seq show that after the community is treated with 5-FOA a positive selection can be observed *in situ*. We additionally demonstrate the utility of the DART vector for both targeted isolation of a single member of a microbial consortia through targeted delivery of an antibiotic resistance cassette, as well as for conferring novel metabolism to a targeted community member by delivery of a *lacZY* cassette allowing growth on lactose. Furthermore, we demonstrate optimal locus selection for high-efficiency phage engineering and report successful integration of varying-sized exogenous payloads for delivery to a target strain.

Overall, these techniques provide a novel avenue for the targeted and isolation independent analysis of specific microorganisms within complex microbial consortia thereby providing a route for genetic testing of microbiome-based hypotheses.

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

Towards integrated rhizosphere microbial community modeling through bottom-up COMETS genome-scale metabolic simulations and top-down data analysis

Ilija Dukovski[4*] (dukovski@bu.edu), Alan Pacheco[4], Dileep Kishore[4], Elysha Sameth[4], Jing Zhang[4], Nathaniel Borders[4], Peter Kim[3], Kateryna Zhalnina[1], Peter Andeer[1], **Daniel Segrè[4] (dsegre@bu.edu),** Adam Deutschbauer[1,2], and Trent Northen[1].
[1]Lawrence Berkeley National Laboratory, Berkeley; [2] University of California, Berkeley; [3]Sandia National Laboratory, Livermore; [4]Boston University, Boston.

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.

In order to disentangle and control the inter-microbial processes that shape rhizosphere communities, we have been developing multiple complementary computational approaches that can be iteratively refined through comparisons with experimental measurements. At the heart of our efforts to reach a detailed mechanistic understanding of how intracellular circuits shape microbe-microbe and microbe-plant interactions, is the use of genome-scale stoichiometric reconstructions of microbial metabolism, simulated using dynamic flux balance analysis (dFBA). In particular, we have been enhancing our comprehensive, open-source modular software for Computation Of Microbial Ecosystems in Time and Space (COMETS). COMETS extends dFBA to generate simulations of multiple microbial species in spatially structured environments. Notably, in COMETS, interactions between species arise as an emergent property of the metabolic dynamics of individual species. The newest version of COMETS (available at runcomets.org)[1] incorporates numerous innovative components that will greatly facilitate the construction of rhizosphere community models. These include a more accurate biophysical model of microbial biomass expansion, newly designed modules for evolutionary dynamics and extracellular enzyme activity, user-friendly Python and MATLAB interfaces, as well as comprehensive documentation and tutorials. Using COMETS, we have started constructing a

geometrically and molecularly accurate *in silico* version of a microfluidics device that creates rhizosphere relevant gradients (the µEcoFAB) This COMETS model will be used to predict the environmental dependence of community growth within the µEcoFAB. We have further used COMETS to implement an evolutionary algorithm that can search through vast spaces of environmental conditions to find those able to support growth and maintenance of communities with desired taxonomic composition or functional properties[2]. We envisage that this approach will identify community-level interventions (which can be achieved through CRISPR-Cas editing tools) likely to affect the community in a desired way.

In parallel to advancing the mechanistic approach summarized above, we have developed new data-driven methods for analyzing high-throughput genomics data. These allow us to generate co-occurrence networks that can be compared with mechanistic predictions. In particular, we have revisited the set of steps necessary for translating 16S ribosomal RNA (16S rRNA) amplicon datasets into networks of inter-microbial associations - a process that is prone to statistical artifacts. By performing a careful analysis of each step in this pipeline, we found that certain choices of algorithms and parameters affect the co-occurrence network more dramatically than others. We further generated an estimate of the most accurate and robust tools for inferring co-occurrence networks based on comparison with mock and synthetic datasets. Our new pipeline (available at https://github.com/segrelab/MiCoNE)[3], can therefore provide a default avenue for generating standardized co-occurrence networks, with the option of easily exploring the variations induced by alternative choices of tools and parameters. This pipeline can be used for integrating multiple datasets, as well as for generating networks that will help us identify community hubs on which to further focus predictive modeling of the outcome of rhizosphere perturbations.

References:

1. I. Dukovski, D. Bajić, J. M. Chacón, M. Quintin, J. C.C. Vila, S. Sulheim, A.. R Pacheco, D. B. Bernstein, W. J. Rieh, K. S. Korolev, A. Sanchez, W. R. Harcombe, D. Segrè: Computation Of Microbial Ecosystems in Time and Space (COMETS): An open source collaborative platform for modeling ecosystems metabolism: arXiv 2020, https://arxiv.org/abs/2009.01734

2. A. Pacheco, D. Segrè: An evolutionary algorithm for designing microbial communities via environmental modification, BioRxiv (2020), https://doi.org/10.1101/2020.11.25.398644

3. D. Kishore, G. Birzu, Z. Hu, C. DeLisi, K Korolev, D. Segrè: Inferring microbial co-occurrence networks from amplicon data: a systematic evaluation, BioRxiv (2020), https://doi.org/10.1101/2020.09.23.309781

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

Phenotypic Characterization of *Brachypodium distachyon* and a Synthetic Community for Dissecting Plant-Microbial Community Interactions in Fabricated Ecosystems (EcoFAB)

Hsiao-Han Lin[1] (hsiaohanlin@lbl.gov), Trenton K. Owens[1], Marta Torres[1], Mon O. Yee[1], Yifan Li[1], Kristine G. Cabugao[1], Kateryna Zhalnina[1], Lauren K. Jabusch[1], Peter F. Andeer[1], Romy Chakraborty[1], Jenny C. Mortimer[1,2](jcmortimer@lbl.gov), Adam M. Deutschbauer[1], and Trent R. Northen[1]

[1]Lawrence Berkeley National Laboratory, Berkeley CA; [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.

Plants grow in environments rich with microbes, where negative (pathogenic), neutral, or beneficial plant-microbial interactions may develop. These microbes are found as a complex community, and so interactions must be understood in the context of a community, rather than as an interaction between a single microbe and the plant. However, mechanistic understandings of plant-microbe interactions often derive from these reductionist single microbe systems. This over simplification creates several gaps in translating lab-produced knowledge to the field. In addition, the involvement of plant genetics in shaping these interactions remains largely unknown. The m-CAFEs project intends to overcome these gaps by adapting fabricated ecosystems (EcoFAB; www.eco-fab.org) to study plant-microbial community-interactions. The EcoFAB is a robust and reproducible system that enables plant microbial research with ecologically relevant experimental parameters [1,2]. To understand the interplay of plant-microbial community interactions, we use *Brachypodium distachyon* and a bacterial community composed of a defined synthetic community (SynCom) as our model system. The SynCom is composed of specific taxa enriched in grass soil, isolated from the switchgrass rhizosphere. For each SynCom bacterium, we sequenced its genome and developed metabolic models within KBase (Kbase.us). Next, we measured physiological traits for these microbes and constructed whole-genome, randomly barcoded transposon mutant libraries, when feasible. We also characterized several plant growth promoting traits of the individual strains, including siderophore and phytohormone production and mineral solubilization assays. Finally, we have characterized *B. distachyon*

growth and cell wall composition when grown in the EcoFAB, providing a robust baseline dataset for use in future experiments, including inoculation with the SynCom. Together this rich set of physiological data combined with predictive models will identify bacterial genes that mediate interactions within this *B. distachyon*-SynCom system, which we will validate using community editing technologies.

References

1. Gao, J. *et al.* Ecosystem Fabrication (EcoFAB) Protocols for The Construction of Laboratory Ecosystems Designed to Study Plant-microbe Interactions. *J. Vis. Exp.* (2018). doi:10.3791/57170

2. Sasse, J. *et al.* Multilab EcoFAB study shows highly reproducible physiology and depletion of soil metabolites by a model grass. *New Phytol.* **222,** 1149–1160 (2019). doi: 10.1111/nph.15662,

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

Bioinformatic Guided Enrichments to Develop Representative Rhizosphere Communities

Mon O. Yee[1*] (monyee@lbl.gov), Dawn Chiniquy[1], Spencer Diamond[2], Peter F. Andeer[1], Andrew R. Osborn[1], Amalia Soenens[3], Crystal Emery[3], Nigel J. Mouncey[1], Jillian Banfield[2], **Romy Chakraborty[1]**, Adam M. Deutschbauer[1], and Trent R. Northen[1]

[1]Lawrence Berkeley National Laboratory, Berkeley; [2] University of California, Berkeley, [3]General Automation Lab Technologies, Inc, San Carlos.

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

We aim to understand rhizosphere microbial communities at sufficient molecular resolution to predict responses to environmental and genetic perturbations. To this end, representative microbial enrichments and isolates are critical to facilitate downstream investigations with gene editing and computational models. In m-CAFEs, we are leveraging our expertise in metagenomics and fabricated ecosystems across labs to guide our enrichment strategies to obtain ecologically relevant yet challenging-to-cultivate taxa that are important to plant-microbe interactions in the rhizosphere.

We used soil from the rooting zone (0-10cm) at Angelo Coast Range Reserve, a well-studied field-site showing distinct microbial processes across depth[1], to grow the model grass *Brachypodium distachyon* in EcoFABs. EcoFABs are fabricated ecosystems designed to investigate the molecular basis of microbial interactions with plants (see Zhalnina *et al.* poster). In parallel, we also grew the same plant in conventional growth systems (i.e. pots, tubes) to assess the impacts of growth chambers on plant growth. After 2 weeks of incubation, 16S rRNA amplicon analysis of the microbial community revealed that the rhizosphere community was enriched in members of *Actinobacteriota, Bacteriodota,* and *Firmicutes* across all growth systems. This represents the community possibly enriched from root exudates and further guides our enrichment strategies.

We are using an enrichment strategy that focuses on reproducing geochemical properties of the Angelo soil and employing carbon sources representative of rhizosphere environments, to mimic environmentally relevant conditions. These enrichments are used to generate reduced complexity consortia that are stable and reproducible and strain isolation (particularly when novel microorganisms are enriched), and targets for genetic manipulation using m-CAFEs genetic tools. To increase the probability of novelty and diversity for editable targets, native soil was used as inoculum directly from the field-site to initiate enrichments. We leveraged metagenomic insights from Angelo soil to select a wide variety of enrichment substrates.. Our preliminary results revealed stark differences corresponding to different carbon sources and we aim to expand our enrichments to also include different incubation conditions beyond carbon sources such as pH and salinity.

Finally, in our third approach, we collaborated with scientists from the DOE Joint Genome Institute to use the GALT prospector platform[2] with the same Angelo soil. Using commercially available media, we were able to isolate field-abundant yet rarely cultivated taxa such as *Acidobacteria* and *Gammatimonadetes* using this massively parallel approach.

These enrichment approaches we have employed are distinct yet complementary to obtain multiple relevant rhizosphere microbial communities and isolates which are now in the pipeline to be tested for phenotypic impacts on plants (see Lin *et al.* poster) and are prime candidates for metabolic modeling (see Dukovski *et al.* poster) as well as for gene editing (see Diamond *et al.* poster).

References

1. Diamond, Spencer, et al. "Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms." *Nature microbiology* 4.8 (2019): 1356-1367.
2. https://www.galt-inc.com/

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 work with GALT was funded through LBNL being awarded their Microbiome Challenge project.

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.
• **References** can be **Publications** instead, if needed. Use any common style for these citations

Spanning laboratory ecosystem scales: insights into environmental complexity with EcoFABs andEcoPODs.

Kateryna Zhalnina[1*](kzhalnina@LBL.gov), Lauren K. Jabusch[1], Peter W. Kim[2], Marc Van Goethem[1], Joelle Sasse[1], Peter F. Andeer[1], Trenton K. Owens[1], Marta Torres[1], Dawn Chiniquy[1], Hsiao-Han Lin[1], Mon O. Yee[1], Catharine A. Adams[3], Albina Khasanova[1], Jonathan Diab[1], Shweta Priya[1], Benjamin Bowen[1], Romy Chackraborty[1], Jenny C. Mortimer[1,4], Aindrila Mukhopadhyay[1], Anup K. Singh[2], John Vogel[1], Adam M. Deutschbauer[1], and **Trent R. Northen[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]Sandia National Laboratory, Livermore; [3]University of California, Berkeley; [4]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.

Microbial rhizosphere communities are critical players in biogeochemical cycles, influencing plant nutrient availability, plant health, development, and stress tolerance. Plants fix CO_2 and release a large quantity of readily available carbon into the rhizosphere. This diverse pool of metabolites and root biopolymers affects microbial community assembly via a process known as the 'rhizosphere effect'. However, little is known about how these plant-microbe interactions are formed and regulated, and how they feedback to the environment. **Here we hypothesize that plants release specific metabolites that are selectively used by rhizosphere bacteria and this enables the plant to tailor its community assembly using exudate composition.** In this work, we use a suite of fabricated ecosystems, which span spatial and temporal scales to investigate complex interactions between plants, microbes, and the environment - from the small-scale **µEcoFAB**, the single-plant scale **EcoFAB,** and the larger **EcoPOD** (see below for descriptions)**.** These platforms enable us to explore complex interactions under fully controllable and highly replicated conditions and evaluate soil biogeochemical processes driven by plant-microbe interactions. These devices are compatible with conventional biosafety workflows to advance secure biosystems design by using contained and controlled environments of EcoFABs and EcoPODs to understand persistence and fate of engineered genes, microbes and synthetic communities through the environment.

µEcoFAB is a mm-scale fabricated ecosystem that mimics chemical gradients in the environment known to create micro-niches and contribute to the diversity of the microbial

communities in soil. Using a multi-layered design, the μEcoFAB generates 3 dimensional gradients (such as oxygen, synthetic exudates, and pH) in the microbial incubation chamber. This enables detailed examination of interactions and functions observed in the larger scale fabricated ecosystems We hypothesize that microbes known to colonize the rhizosphere will be enriched in regions with the highest synthetic exudate concentrations and low pH. **The EcoFAB** is a cm-scale chamber that captures key features of an ecosystem of interest (e.g., soil, plants). It includes main components of the ecosystem, including native, synthetic, or engineered microbial communities and different microbes (bacteria, archaea, fungi), plants, growth media (soil, hydroponics, sand), and is integrated with multiple measurement technologies, including high resolution imaging, and ports for collecting samples. The EcoFAB provides a sterile and highly controlled system that enables plant growth, phenotyping and plant-microbe imaging, sampling for metabolite, and microbial community composition analysis. We have been able to see unique establishment patterns in the rhizosphere by imaging multiple fluorescent strains of *Pseudomonas simiae* within EcoFABs. Since the presence of microorganisms in the rhizosphere does not necessarily indicate that they are metabolically active within the environment, we are identifying an active fraction of microbes in the rhizosphere using bio-orthogonal noncanonical amino acid tagging (BONCAT) combined with fluorescence-activated cell sorting (FACS). This was accomplished by using a modified EcoFAB that allowed cultivation of switchgrass in soil followed by BONCAT labeling and cell sorting, single-cell shotgun sequencing, and metabolomics. **The EcoPOD** is a meter-scale fabricated ecosystem that provides a contained and controlled environment equipped with sensors to regulate temperature, humidity, and other important climatic parameters both above and below ground thus complementing EcoFAB capabilities. Specifically, EcoPODs provide contained and controlled environments that reflect key aspects of field conditions to focus EcoFAB experiments on key microbes, metabolites, and genes. In turn, EcoFABs provide relatively low-cost and high-throughput capabilities for deconstructing molecular mechanisms to develop generalizable models, mechanisms, and principles for testing in EcoPODs. In the longer-term the EcoPODs will be linked to field studies to provide a complete translational path between laboratory and field experiments.

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

Crosstalk: Interkingdom interactions in the mycorrhizal hyphosphere and ramifications for soil C cycling

Erin E Nuccio[1*] (nuccio1@llnl.gov)

[1]Lawrence Livermore National Laboratory, Livermore, 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. My Early Career research will investigate 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, I will investigate 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. My 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, my Early Career research will dissect the interactions and signaling networks within the hyphosphere using a systems biology approach that leverages isotope tracers as well as 'omics technologies. 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, my 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. Work conducted at Lawrence Livermore National Laboratory was supported under the auspices of the U.S. DOE under Contract DE-AC52-07NA27344.

Microbial dynamics and syntrophic interactions at the pore scale: towards an integration of reactive transport and microbial cell models

Heewon Jung,[1,#] Christof Meile[1*] (cmeile@uga.edu), Nidhi Gupta[2 *] (ngupta@anl.gov), Jose Faria[2], Janaka Edirisinghe[2], Filipe Liu[2], Christopher Henry[2] and **Victoria Orphan**[3]

[1]University of Georgia, Athens, GA; [2]Argonne National Laboratory, Lemont, IL; [3]California Institute of Technology, Pasadena, CA; [#]now at the Department of Geological Sciences, Chungnam National University, South 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) a reactive transport model that establishes a connection between processes at the pore scale and the macroscopic environment and (2) novel flux balance community metabolic models of anaerobic oxidation of methane (AOM) consisting of methanotrophic archaea (ANME) and syntrophic sulfate-reducing bacteria.

The porescale modeling uses a Lattice-Boltzmann approach to solve the Navier-Stokes equation to establish the flow field, and advection - diffusion - reaction equations to model chemical distributions and account for microbial growth dynamics. Here, we apply the model to study the feedback between flow and microbial growth, in order to investigate the relationship between two key macroscopic characteristics of porous media, permeability and porosity, under conditions where bioclogging can occur. Simulation results for an idealized porous medium show biofilm growth at the pore scale which affects the evolution of porosity and permeability. The resulting spatially heterogeneous distribution of biomass leads to deviations from empirical macroscale models such as the Kozeny-Carman (KC) equation. For impermeable biofilms the deviations of porosity-permeability relations from the KC equation depend strongly on flow and reaction conditions, reflected in Péclet (*Pe*) and Damköhler numbers (*Da*). Simulations with porous and permeable biofilms reveal a substantially different evolution of porosity and permeability compared to nonporous and impermeable biofilms, highlighting the importance of microscale biofilm characteristics for macroscale hydrological properties of porous media. These results emphasize that accounting for microbial processes at the porescale (i.e. micrometer scale) can be important for understanding the evolution of chemical concentrations, microbial activities and flow at the macroscale.

Connecting micro- and macro-scales in the above application relied on a highly simplified, substrate-dependent process rate representing microbial activity. Thus, our ongoing efforts include the integration of omics-based in silico cell models to simulate sulfate-coupled AOM in a sedimentary environment.

Significant progress was made in our efforts to construct community flux models of AOM consortia. A critical first step is to improve our ability to construct accurate genome-scale metabolic models of the individual types of microbes that comprise these uncultured consortia: methanotrophic archaea (ANME) and sulfate-reducing bacteria. Toward this end, we curated pathways and model templates for these classes of microorganisms in our latest build of the ModelSEED genome-scale model reconstruction tool in the DOE Systems Biology Knowledgebase (KBase). The ModelSEED now has an archaea template with an archaea-specific biomass reaction. Additionally, it includes 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. Special attention was focused on the methanogenesis and related methanotrophy pathways in this curation, as these pathways are central to the ANME phenotype. ANME microbes are related to methanogenic archaea, but these organisms support “reverse methanogenesis” as the biochemical model for methane oxidation. This metabolic activity requires an efflux of electrons to a syntrophic partner, which is the role played by sulfate-reducing bacteria (SRB) in these systems. Thus, in addition to properly capturing the reactions and annotations for the methanogenesis pathway, we are also adjusting model flux constraints to ensure proper implementation of ANME behavior. We are performing similar curation to improve the representation of sulfur reduction metabolism in the ModelSEED, improving these pathways in all ModelSEED models. We applied 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. Ultimately, these predictions will be tested in the reactive transport simulations described above, as well as new experimental studies.

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.

Ecology, Diversity, and Biogeochemical Contribution of Viruses in Methane-Rich Sediments

Alon Philosof[1*], Aditi Narayanan[2], Maria Alvarez Sanchez[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 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.

In recent years, the fundamental role of viruses in marine ecological networks has become increasingly apparent. However, viral communities in sediment environments are largely unexplored, though initial evidence exists for their large contribution to the dissolved organic carbon pool. In anoxic sediments of cold methane seeps, methane is the primary energy source; anaerobic oxidation of methane coupled to sulfate reduction, performed by a microbial consortium of anaerobic methanotrophic archaea (ANME) and sulfate-reducing bacteria (SRB), underpins much of the biodiversity of the seep and beyond. To understand the role of viruses in this environment, 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, virus-like particles (VLP) were concentrated and then further purified using an Optiprep density gradient at two distinct densities (30 and 35%). Metagenomic assembly of these samples yielded more than 2200 complete and nearly complete viral genomes along with thousands more contigs longer than 15kb. Initial phylogenetic analyses show that the viral community from this environment is very diverse, mostly uncharacterized, and includes a number of novel distinct clades. We also profiled the diversity of VLP morphologies using transmission electron microscopy (TEM). Seep samples contain a variety of viral capsid morphologies, ranging from the classic head-tail phage to the spindle shapes of viruses known to infect archaea. In addition, incubations amended with anthraquinone-2, 6-disulfonate (AQDS), an external electron acceptor which decouples the process of AOM from sulfate reduction[1], showed an increase in the abundance of spindle-shaped viruses. Parallel incubations lacking

AQDS, in which the SRB partner was not decoupled from the ANME, showed a greater diversity of tailed viruses. These results are the first indication of the potential impact of the ANME-SRB syntrophy on the viral community in the cold seep environment. To further constrain the impact of viruses on AOM and sulfate reduction as well as their impact on the syntrophic relationship between ANME and SRB, we are developing a number of new methods to track viral activity and elemental composition. First, we adapted a protocol[2] for Biorthogonal Non-Canonical Amino Acid Tagging (BONCAT) with click-chemistry for fluorescently labeling newly-synthesized viruses. This will enable the quantification of viral production under different environmental conditions and host physiologies. We are also developing a method that combines viral-BONCAT fluorescence-activated sorting (FACS) with Single Virus Genomics[3] (SVG) in order to identify and sequence the genomes of newly produced viruses in sediments from active host cells. Optimization of protocols combining stable isotope probing with fluorescence identification and nanoscale Secondary Mass Spectroscopy (NanoSIMS) is also ongoing with sediment hosted viruses to assess the proportion of of carbon (methane) and nutrients (nitrogen) through the viral assemblage under different incubation conditions in the laboratory.

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.

Title: Uncovering Syntrophies within Methane-Oxidizing Microbial Consortia in Sediments: Synthesizing Insights from the Subcellular to the Population Scales

Authors: Yongzhao Guo[1], Haley Sapers[1,#], Ranjani Murali[1], Hang Yu[1], Grayson Chadwick[1], Rodney Tollerson[1], Kriti Sharma[1]* (ksharma@caltech.edu), John Magyar[1], Sam Webb[2], Tanja Woyke[3], Rex Malmstrom[3], Mark Ellisman[4], Tom Deerinck[4], Zhou Li[5], Robert Hettich[5], **Victoria Orphan* (vorphan@gps.caltech.edu)** [1]

Institutions: [1]California Institute of Technology, Pasadena, CA; [2]SLAC National Accelerator Laboratory, Menlo Park, CA; [3]SLAC National Accelerator Laboratory, Menlo Park, CA; [4]National Center for Microscopy and Imaging Research, La Jolla, CA; [5]Oak Ridge National Laboaratory, Oak Ridge, TN; [#]currently at York University, Toronto, ON, Canada

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 text:
Microbial-driven anaerobic oxidation of methane (AOM) accounts for up to 80% of methane sequestration in anoxic sediments, preventing this greenhouse gas from reaching the atmosphere. This sequestration is catalyzed primarily by consortia of ANaerobic MEthane-oxidizing 'ANME' archaea and syntrophic bacteria coupled through redox chemistry to sulfur, nitrogen, iron, and manganese. While the importance of AOM consortia for methane-oxidation in sedimentary environments is widely appreciated, the nutritional requirements and interdependencies of these diverse methanotrophic archaeal-bacterial syntrophies and their collective impact on nutrient cycling within the sedimentary ecosystem is not well understood.

Our group has developed several approaches to elucidate the syntrophies between sediment-hosted ANME archaea and their sulfate-reducing bacterial partners (SRB) at the intracellular, intercellular, and population scales, and over time. A key tool in our investigations is the disruption of extracellular electron transfer (EET) between ANME and SRB -- the main syntrophic association between the partners -- through the addition of an external electron acceptor (anthraquinone-2,6-disulfonate, or AQDS) to our incubations [1]. AQDS addition effectively enables ANME to carry out AOM autonomously without the SRB partner; by comparing cells under conditions of intact and disrupted syntrophy, we are able to reveal the nutrient flows dependent upon the syntrophic association.

Our use of 3D electron microscopy and synchrotron-based XRM and XANES spectroscopy enables investigation of ANME-SRB consortia *at the intracellular level,* and reveals storage granules important for nutritional syntrophies. Using serial block electron microscopy (SBEM) to image individual aggregates with nanometer resolution in three dimensions, and applying

supervised machine learning to image segmentation, we are reconstructing the 3D cell volume of several ANME-SRB aggregate morphologies to determine how macromolecular storage compounds (e.g. polyphosphate, carbon or iron storage granules) accumulate or dissimilate within individual archaeal or bacterial cells, and how these nutrients are stored within or shared between cells in the presence or absence of AQDS [2]. Complementing our SBEM imaging, we are using synchrotron-based X-ray fluorescence imaging and XANES spectroscopy at SSRL beamlines 2-3 and 14-3 to investigate elemental speciation in ANME-SRB aggregates. Using the recently upgraded beamline optics, which provide substantially improved signal and resolution, we have identified differences in sulfur-containing macromolecules in AOM consortia incubated under AQDS vs. SO_4^{2-} conditions.

We have developed pulse-chase multi-isotope labeling coupled with FISH-nanoSIMS to investigate consortia *at the aggregate level, and over time,* revealing the growth patterns of consortia as well as the phenotypic heterogeneity of growth patterns between aggregates. By adding $^{15}NH_4^+$, D_2O, $H^{13}CO_3^-$ or $H_2{}^{18}O$ at different incubation periods and tracking isotopic signals enriched in the ANME-SRB biomass using FISH-nanoSIMS, we have obtained some of the first *dynamic* single-cell activity measurements of these slow-growing aggregates. Additional FISH-nanoSIMS analyses reveal cell specific patterns in N_2 fixation among ANME and SRB. ANME archaea have been suggested to be dominant diazotrophs in AOM consortia; however it is not clear whether they are capable of this energetically demanding anabolic process in the absence of their syntrophic partner, nor is it clear how newly fixed nitrogen is partitioned between syntrophs. $^{15}N_2$ FISH-nanoSIMS in the presence and absence of AQDS confirmed active diazotrophic growth by ANME in the absence of an active SRB partner and active transfer of N-containing biomolecules into the extracellular matrix surrounding both partners. Other $^{15}N_2$ labeling experiments revealed a greater degree of variation in diazotrophic activity among diverse AOM consortia, with select SRB syntrophic partners (newly described Seep SRB1g group) shown to also be capable of N_2 fixation in association with ANME-2b [3].

Finally, *at the population level*, our use of proteomic and genomic approaches to compare metabolically-coupled and uncoupled aggregates reveal proteins and pathways important to the syntrophic lifestyle. We find distinct proteomic profiles in both the ANME and SRB in the presence and absence of AQDS. We observed a distinct shift from anabolic activity during metabolic coupling to catabolic activity when the syntrophy is disrupted. Proteome remodeling was observed in many protein families including nitrogen metabolism, amino acid biosynthesis, redox sensing, and cell surface composition. Taken together, our multi-scale approaches and observations will help inform and test model predictions of syntrophic associations and methane metabolism in the environment.

References/Publications

1. Scheller S et al 2016; Science 351: 703-707.
2. McGlynn S et al 2018; Appl Env Microbiol 84 (11): e00399-18
3. Metcalfe et al., 2021; ISME J 15: 377-396

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 **Award Number:** DE-SC0020373.

Systems analysis of a fast growing N_2-fixing cyanobacterium for production of advanced biofuels and nitrogen-containing petrochemical replacement compounds

Anindita Banerjee[1]* (anindita@wustl.edu), Deng Liu[1], Zi Ye,[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

https://sites.wustl.edu/photosynthbio/anabaena_33047/

Project Goals:

The overall objective of this project is to use an integrated systems biology approach to develop the filamentous cyanobacterium *Anabaena* sp. PCC 33047 as a model fast-growing, photosynthetic, diazotrophic production platform. The specific goals for this project are: 1) Construct a genome-scale metabolic model and predict genetic alterations that optimally direct fixed CO_2 and N_2 into target products. 2) Apply 13C and 15N assisted metabolomics and metabolic flux analysis to dissect the metabolism of the strain. 3) Develop an efficient genetic toolkit. 4) Demonstrate production of caprolactam and valerolactam in engineered *Anabaena* 33047. 5) Establish a stable consortium between *Anabaena* 33047 and a heterotroph for cost-effective bioproduction.

Abstract

Anabaena sp. ATCC 33047 is heterocystous cyanobacterium that thrives under very high light intensities and exhibits fast growth both in the presence and absence of fixed nitrogen, traits that make this strain an attractive platform for the cost effective production of nitrogen-rich compounds. The strain was known to be genetically intractable and hence not utilized in molecular and synthetic biology studies. During the course of this project we developed a genetic manipulation system that enabled us to make targeted modifications in its genome. We further characterized one of the modified strains, which shows greatly enhanced rates of nitrogen fixation. We also developed a genome scale metabolic model, *i*AnC892, for *Anabaena* 33047 to identify genetic interventions to overproduce valerolactam and caprolactam.

Phycobilisomes (PBS) are large antenna protein complexes in cyanobacteria that harvest light and funnel the energy to the photosynthetic reaction centers. When subjected to nitrogen deficient growth conditions, heterocystous cyanobacteria exhibit transient PBS degradation mediated by the NblA protein. As nitrogen fixation commences, PBS is resynthesized back to their normal levels in vegetative cells, but their abundance remains low in

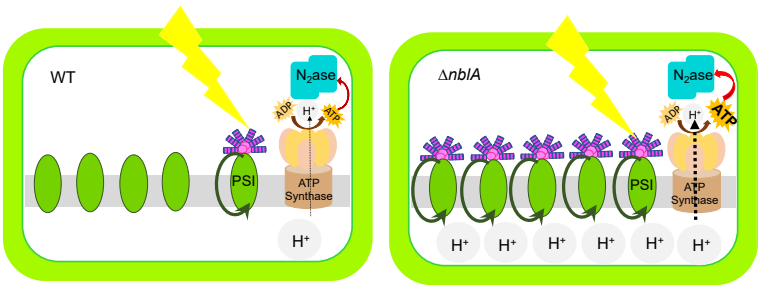


Figure 1. Schematics depicting the differences in the heterocysts of the WT and Δ*nblA* strains that contribute to enhanced nitrogenase activity. Higher abundance of phycobilisomes in the mutant heterocyst and their association with PSI centers leads to higher ATP generation and nitrogenase activity.

heterocysts. Earlier studies implicating a role for PBS in energy transfer to photosystem I (PSI) in heterocysts [1,2], instigated our efforts to investigate the effect of antenna modification on nitrogenase activity in the high light tolerant *Anabaena* 33047. To this end, we generated a Δ*nblA* mutant of *Anabaena* 33047 that retains large amounts of PBS in its heterocysts. Intriguingly, when subjected to high light the mutant exhibited 2.5 folds higher rates of nitrogen fixation compared to the WT. Analysis of the mutant indicated, increased cyclic electron flow, possibly resulting from higher PBS mediated energy transfer to PSI. This contributes to increased ATP synthesis and enhanced nitrogenase activity in the mutant (Figure 1) [3]. Thus the Δ*nblA* mutant of *Anabaena* 33047 offers an improved platform for the production of nitrogen rich compounds.

Genome-scale metabolic models (GSM) facilitate in-silico engineering of microbial metabolism for over production of target chemicals [4]. We developed a genome scale metabolic model, *i*AnC892, for *Anabaena* 33047 to identify genetic interventions to overproduce valerolactam and caprolactam [12]. The model was constructed by retrieving annotations from multiple databases: KEGG [5], MetaCyc [6] and ModelSEED [7] and a recently published model for the closely related *Anabaena* 33047 [8]. *i*AnC892 contains 953 unique reaction 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. The model was used alongside the strain design algorithm, OptForce [9], to identify genetic interventions that would lead to overproduction of caprolactam and valerolactam. The production of valerolactam and caprolactam were enabled by adding 5-aminovalerate pathway [10] and Adipyl-CoA pathway [11] respectively into *i*AnC892. The identified strategies recapitulated several of the previously reported genetic interventions that enhance the target compound production [12].

To explore the possibility of establishing a stable consortium between *Anabaena* 33047 and a heterotroph, several *E.coli* strains were co-cultured with this cyanobacterium under different growth conditions. Our study revealed that *Anabaena* 33047 can support the growth of *E.coli* Top10 cells when the strains are co- cultured in nitrogen free medium. Our preliminary results indicate that *Anabaena* 33047 can be a suitable partner in a synthetic photoautotrophic-heterotrophic consortium.

References

1. Cardona, T, Magnuson A: **Excitation energy transfer to Photosystem I in filaments and heterocysts of *Nostoc punctiforme***. *Biochim Biophys Acta* 2010, **1797**: 425-433.
2. Watanabe, M, Semchonok, DA, Webber-Birungi, MT, Ehira, S, Kondo, k, Narikawa, R, Ohmori, M, Boekema, E, Ikeuchi, M: **Attachment of phycobilisomes in an antenna-photosystem I supercomplex of cyanobacteria**. *Proc Natl Acad Sci U S A* 2014, **111**: 2512-2517.
3. Bandyopadhyay, A, Ye, Z, Benedicty, Z, Trtilek, M, Pakrasi, HB: **Antenna modification leads to enhanced nitrogenase activity in a high light tolerant cyanobacterium**. (Submitted) 2021.
4. Simeonidis E, Price ND: **Genome-scale modeling for metabolic engineering**. *J Ind Microbiol Biotechnol* 2015, **42**:327-338.
5. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome**. *Nucleic Acids Res* 2004, **32**:D277-280.
6. 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.
7. 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.
8. 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.

9. 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.
10. Zhang J, Barajas JF, Burdu M, Wang G, Baidoo EE, Keasling JD: **Application of an Acyl-CoA ligase from *Streptomyces aizunensis* for lactam biosynthesis**. *ACS Synth Biol* 2017, **6**:884-890.
11. Turk SCHJ, Kloosterman WP, Ninaber DK, Kolen KPAM, Knutova J, Suir E, Schürmann M, Raemakers-Franken PC, Müller M, de Wildeman SMA, et al.: **Metabolic engineering toward sustainable production of Nylon-6**. *ACS Synthetic Biology* 2016, **5**:65-73.
12. 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.** (Submitted) 2021.

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

Dynamic modeling of a synthetic *Clostridia* triple co-culture for producing medium-chain fatty acids from glucose

Charles Foster[1], Kamil Charubin[2], Eleftherios T. Papoutsakis[2], Costas D. Maranas[1]

[1]Pennsylvania State University, University Park

[2]University of Delaware, Newark

http://papoutsakis.org/?page_id=72

Project Goals: To construct and integrate a *Clostridium kluyveri* genome scale metabolic model into dynamic modeling framework with a novel cell fusion/growth model and hybrid *C. acetobutylicum* and *C. ljungdahlii* metabolic models, and deploy the resulting community model with existing strain design algorithms to identify engineering targets across the triple syntrophic culture to enhance diesel and aviation fuel precursor production.

Clostridial co-cultures have the potential to improve fermentation yields for value-added bioproducts through mutualistic cross-feeding which improves carbon conversion efficiency. In addition to 100% substrate conversion, *C. acetobutylicum* (*Cac*) and *C. ljungdahlii* (*Clj*) in syntrophic culture were shown to fuse membranes and exchange cytosolic contents, adding to their bioproduction potential. Hybrid cells with significant shifts in gene expression caused higher-than-expected fermentation yields, and *Clj* persisted at high abundance thanks to its acquisition of *Cac*-native metabolic enzymes during cell fusion. By introducing a dynamic genome-scale metabolic modeling framework with hybrid *Cac* and *Clj* metabolic models, we offer a mechanistic explanation for how cell fusion alters the growth phenotype and panel of metabolites produced by this binary community. Inclusion of hybrid metabolic models and a novel growth model accounting for cell fusion/proteome swap enables quantitative agreement with fermentation profiles that is beyond the scope of dynamic simulations in the absence of

such detail. We now seek to add value to this system by incorporating *C. kluyveri* (*Ckl*) to produce medium-chain fatty acids from ethanol and acetate excreted by *Cac* and *Clj*. To this end, we first assemble a genome-scale reconstruction of *Ckl* (iCKL841) containing 1989 reactions corresponding to 841 gene products inferred through gene homology analysis and gene-protein-reaction mapping from six existing *Clostridia* genome scale models. We then merge iCKL841 into a dynamic simulation of the triple synthetic co-culture which includes *Ckl* together with hybrid metabolic models of *Cac* and *Clj*. We are integrating this system with optKnock and optForce strain design algorithms to identify perturbation strategies for maximizing medium-chain fatty acid production from glucose.

Funding provided by the DOE Office of Science, Office of Biological and Environmental Research (Award Number DE-SC0019155).

Syntrophic Co-Cultures of *Clostridium* Organisms to Produce Higher Alcohols and Other C6-C8 Metabolites

Jonathan K. Otten[1,2,*] (jkotten@udel.edu), Kamil Charubin[1,2], Michael Dahle[1,2], Charles Foster[3], John Hill[1,2], Noah Willis[1,2], Dr. Costas Maranas[3] and **Dr. Eleftherios Terry Papoutsakis**[1,2]

[1] University of Delaware, Newark, DE; [2] Delaware Biotechnology Institute, Newark, DE;
[3] Pennsylvania State University, State College, PA

http://www.papoutsakis.org/ | http://www.maranasgroup.com/

Project Goals: The goal of this project is to develop syntrophic *Clostridium* co-culture systems for producing intermediate carbon-chain length metabolites (C4-C8) and their derivatives that can be used as chemicals or serve as biofuels and their precursors. Part of the effort is to develop O_2-independent fluorescent reporters which will allow us to determine the population dynamics of the dual and triple co-culture system in real time. Furthermore, new fluorescent reporters will also allow us to study the unique cell-to-cell interactions between organisms, which lead to the unique co-culture phenotype and performance. Genome scale models will also elucidate the interactions of co-culture organisms.

Multiple *Clostridium* organisms are of major importance for developing new technologies to produce biofuels and chemicals. Solventogenic clostridia are capable of utilizing a large variety of biomass-derived carbohydrates such as hexoses, pentoses, disaccharides, and hemicellulose, and can produce a good number of C2-C4 chemicals. Acetogenic clostridia can fix inorganic H_2, CO_2, and CO to generate C2 acids and alcohols. Other specialized clostridia possess diverse biosynthetic capabilities for production of a wide variety of metabolites including C4-C8 carboxylic acids and alcohols, which could serve as commodity chemicals, biofuels, or biofuel precursors.

While most previous work with clostridia focused on optimizing single organisms for production of biochemicals, microorganisms naturally live in complex communities where syntrophic interactions result in superior resource utilization. Here, we first examined a synthetic syntrophy consisting of the solventogen *C. acetobutylicum* (*Cac*), which converts simple and complex carbohydrates into a variety of chemicals, and the acetogen *C. ljungdahlii* (*Clj*), which fixes CO_2.[1] This synthetic co-culture achieved carbon recoveries into C2-C4 alcohols almost to the limit of substrate-electron availability, with minimal H_2 and CO_2 release. The co-culture exhibited unique direct cell-to-cell interactions and material exchange among the two microbes, which enabled unforeseen rearrangements in the metabolism of the individual species that resulted in the production of non-native metabolites, namely isopropanol and 2,3-butanediol.[1] The unique co-culture phenotype was possible only when both organisms were allowed to physically interact, which allowed them to form unique cell-to-cell fusions. To further investigate the extent of these interactions we have developed fluorescent *Cac* and *Clj* expressing fluorescent FAST[2] and HaloTag® proteins, respectively. When co-cultured together, both fluorescent strains exhibited cell-to-cell fusion, a never-before-seen phenomena. The cell fusion facilitated the whole-cell

exchange of protein and RNA material between both organisms and led to the formation of stable 'hybrid' bacterial cells which contained the cellular material of both organisms.

To expand our synthetic co-culture system, we formed dual and triple co-cultures including *C. kluyveri* (*Ckl*), which can metabolize ethanol and acetate to produce C6 and C8 carboxylic acids. Both *Cac* and *Clj* natively produce ethanol and acetate, and they have been engineered to produce higher amounts of ethanol while producing fewer unneeded chemicals. These organisms also convert the carboxylic acids produced by *Ckl* into their respective alcohols. In order to identify genes differentially expressed in co-culture conditions, an RNAseq study was performed on the binary *Cac*-*Clj* co-culture, using *Cac* and *Clj* monocultures as controls. Preliminary analysis of genes hypothesized to play key roles in the cross-species metabolic pathways showed upregulation in co-culture, supporting prior RT-qPCR results. Ongoing global transcriptional analysis will categorize differentially expressed genes based on the Clusters of Orthologous Groups of proteins (COGs) database to determine which gene classes play important roles in the co-culture phenotype. Future RNAseq studies will focus on the *Cac*-*Ckl* and *Cac*-*Clj*-*Ckl* co-cultures.

^{13}C-based Metabolic Flux Analysis (MFA) is used to gain insight into the regulation of cell growth and product formation pathways and to identify metabolic bottlenecks. Currently, use of stable-isotope tracers combined with measurements of isotopic labeling by mass spectrometry represents the state-of-the-art in flux determination. Metabolic fluxes are being studied using ^{13}C MFA in *Ckl*, *Cac*, and *Clj* under mono- and co-culture conditions to identify key changes in metabolism of each organism. We have introduced a dynamic genome-scale metabolic modeling framework and hybrid metabolism models to explore how cell fusion alters the growth phenotype and panel of metabolites produced by the binary *Cac*/*Clj* co-culture. Computational results agree quantitatively with experimental fermentation profiles and indicate *Clj* persists in the co-culture with the aid of *Cac* metabolic enzymes acquired during fusing events. We have assembled a genome-scale metabolic reconstruction (GSM) of *Ckl* containing 1989 reactions inferred through gene homology analysis and gene-protein-reaction mapping from six existing Clostridia GSMs. We are merging our *Ckl* GSM with *Cac* and *Clj* hybrid metabolism models in our dynamic modeling framework and using this system to explore perturbation strategies for maximizing medium-chain fatty acid production.

References

1 Charubin, K. & Papoutsakis, E. T. Direct cell-to-cell exchange of matter in a synthetic Clostridium syntrophy enables CO2 fixation, superior metabolite yields, and an expanded metabolic space. *Metabolic engineering* **52**, 9-19, doi:10.1016/j.ymben.2018.10.006 (2019).

2 Streett, H. E., Kalis, K. M. & Papoutsakis, E. T. A strongly fluorescing anaerobic reporter and protein-tagging system for Clostridium organisms based on the Fluorescence-Activating and Absorption-Shifting Tag (FAST) protein. *Applied and environmental microbiology*, AEM.00622-00619, doi:10.1128/AEM.00622-19 (2019).

SUPPORTED by the U.S. Department of Energy (Award No. DE-SC0019155).

Patterns of diversity in the North American Populus mycobiome

Michael Van Nuland[1*] (mvannula@stanford.edu), Caroline Daws[1], and **Kabir Peay**[1]

[1]Stanford University, Stanford, CA

Project Goals: Despite apparently strong geographic patterns in the dominant form of mycorrhizal symbiosis and the associated ecosystem consequences, ecologists have a limited understanding of why these patterns emerge. Why and how do ectomycorrhizal host trees outcompete arbuscular mycorrhizal host trees in certain ecosystems? Does climate play a direct or indirect role in determining the success of ectomycorrhizal symbiosis? Given that ectomycorrhizal fungi are themselves highly diverse, do changes in the ectomycorrhizal community expand the range of climates a host tree can grow in? For this project we are using *Populus* as a model to determine the mechanisms by which mycorrhizal symbiosis influences the distribution of tree species across North America.

Populus species are important bioenergy feedstocks and prevalent throughout North American forests[1]. Identifying biotic communities that may promote their persistence under different environmental conditions is vital to the sustainability of feedstock programs and forest conservation and management into the future. Microbial communities are especially important to consider as they form symbioses with host plants and influence their overall performance, stress tolerance, and disease susceptibility[2,3]. However, we know relatively little about how the distribution of microbial symbionts varies among and within *Populus* species, and how this relates to the broad environmental gradients underlying host tree species ranges. We surveyed five *Populus* species from across the United States to understand how host-associated fungal communities vary across major soil and climate gradients.

We collected leaf, root, and soil samples across five widespread Populus species ranges in the US, measured ectomycorrhizal fungi root colonization, and performed fungal (ITS1) amplicon sequencing (Illumina MiSeq) to characterize fungal community diversity and composition[4]. We tested how fungal community diversity and composition varies among tree species and broad environmental gradients. We used soil measurements and climate data to predict the amount of community turnover across each Populus compartment (Leaf, Root, and Soil) with generalized dissimilarity models[5]. We used distance decay analysis to examine how community structure varies over geographic space. We also identified distinct ecological clusters of fungal taxa using semi-partial correlations and visualized their natural clustering using co-occurrence networks. Our analyses showed that there are significant differences in the composition of fungi across the leaf, root and soil compartments of the *Populus* mycobiome, and that composition of these respective communities are affected by different combinations of climate and environmental variables, resulting in highly unique regional microbiomes. Finally, we show that different *Populus* tree species vary in their ectomycorrhizal affinity, and that this correlates with predicted litter decomposition rates (k) in a manner consistent with global biogeographic patterns of mycorrhizal root associations[6].

Populus associate with diverse fungal communities across their leaf, root, and soil compartments, which are structured by different soil and climate gradients across the US. Identifying these environmental drivers of host-mycobiome relationships will aid in designing sustainable feedstock programs and managing forest systems into the future.

References

1. Stettler, Bradshaw, Heilman, and Hinckley (1996) “Biology of Populus and its Implications for Management and Conservation”, NRC Research Press.
2. Busby, P. E., K. G. Peay, and G. Newcombe. 2016. Common foliar fungi of *Populus trichocarpa* modify *Melampsora* rust disease severity. *New Phytologist* **209**:1681-1692.
3. Peay, K. G. 2016. The mutualistic niche: mycorrhizal symbiosis and community dynamics. *Annual Review of Ecology and Systematics* **47**:143-164.
4. Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson, and S. P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**:581-583.
5. Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modeling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions* **13**:252-264.
6. Steidinger, B. S., T. W. Crowther, J. Liang, M. E. Van Nuland, G. D. A. Werner, P. B. Reich, G. Nabuurs, S. de-Miguel, M. Zhou, N. Picard, B. Herault, X. Zhao, C. Zhang, D. Routh, G. Consortium, and K. G. Peay. 2019. Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* **569**:404-408.

Funding statement. This work is supported by Early Career Research Grant DE-SC0016097 to KGP from the DOE Office of Science, Biological and Environmental Research Program.

Enhanced Resistance Pines for Improved Renewable Biofuel and Chemical Production

Gary F. Peter[1], gfpeter@ufl.edu, Daniel Ence[1], Mallory Morgan[1]

[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. The SE US also houses the U.S. pine chemicals industry, which is the oldest and one of the largest US renewable hydrocarbon chemical industries. Our focus is to increase constitutive terpene production to enhance loblolly and slash pine resistance to pests and pathogens. Enhanced genetic resistance in these commercial pine species is critical to protect against widespread losses as biotic pressures increase due to global warming, land-use change, and introduced exotic organisms.

Increasing pine terpenes is also well aligned with the needs of the developing bioeconomy. Today, commercial scale collection of pine terpenes occurs from live trees by tapping, from 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 convert efficiently 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, as pine terpene supply is currently limited by 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. In previous research we demonstrated that terpene defense traits are under genetic control and behave as quantitative traits, and have used genetic engineering to validate 12 genes that can significantly increase wood terpene content.

In objective one, we are integrating existing and new genome wide association (GWAS) 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. GWAS analyses of constitutive oleoresin flow, wood mono- and diterpene content and resin canal number with 82,000 biallelic SNPs were completed for our CCLONES population and constitutive oleoresin flow, mono- and diterpene

content are complete and resin canal number is in progress for our ADEPT 2 population. In the ADEPT2 population, we also measured induced oleoresin flow after treating clones with methyl jasmonate (MeJA). While our goal is to increase constitutive terpene defenses, we use MeJA to induce defense responses to identify the genes and genetic architecture of resinosis. Quantitative genetic and association analyses are in progress with ~2.2 million biallelic SNP markers. In our pseudo-backcross population between one F1 slash x loblolly hybrid genotype backcrossed to slash and loblolly genotypes, we collected constitutive oleoresin flow for future QTL mapping. We measured constitutive and induced resin flow in mature USFS known high-yielding slash pine selections in two seed orchards. We tapped 124 ramets from 31 clones in FL and 55 ramets from 14 clones in GA. Analysis to further characterize these high-yielding genotypes is underway.

We conducted a time course experiment to identify early, mid and late genes expressed in differentiating resin ducts. In addition to available cell specific RNAseq data, we also created 88 new RNAseq libraries from cambial zone tissue collected from days 0, 1-14, 17, and 21 after treatment with MeJA, an inducer of new axial resin canals. We also constructed 43 libraries from transgenic pines expressing pine genes with significantly greater wood terpene content. Anticipating integrating these datasets, we created a database of all genomic resources including porting pine ref v1.01 to which most SNPs were designed to v2.01. We created a more complete transcriptome with 64,671 genes, including existing contigs, PacBio reads and pine ref v2.01 predicted transcripts. This improved transcriptome has 80% full-length transcripts for mapping RNAseq reads for differential expression and functional annotation of significant SNP markers.

In objective two, we are using 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

Historic Precipitation Regimes Influence Microbial Population Dynamics in Response to Seasonal Rewetting in Mediterranean-Grassland Soil

Steven J. Blazewicz*[1] (blazewicz1@llnl.gov), Ben Koch[2], Ella Sieradzki[3], Katerina Estera-Molina[3], Mengting Yuan[3], Marissa Lafler[1], **Mary Firestone[3]**, **Bruce Hungate[2]**, **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 Science, 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: Microbial activity is rapidly stimulated by the rewetting of dry soils, resulting in a pulse of carbon mineralization and nutrient availability. This phenomenon is important because significant portions of annual net ecosystem production can be lost through microbially driven C mineralization in a single rewetting event, and predicted climate changes in semi-arid life zones could profoundly change soil C dynamics and nutrient availability. While there has been much interest in the immediate response of indigenous communities to soil wet-up, little is known about the effects of reduced spring rainfall on subsequent microbial population responses in semi-arid soils undergoing a fall wet-up.

To determine how reduced spring rainfall affects losses of recently fixed C and taxon-specific population dynamics, we performed a wet-up experiment using soils that had been field-labeled with $^{13}CO_2$ by the Firestone group as part of their 'Phage and Fauna' Genomic Sciences project. Sixteen rainfall-manipulation plots were exposed to $^{13}CO_2$ (labeled) or $^{12}CO_2$ (unlabeled) for 5 days to label photosynthetic soil inputs under two precipitation regimes: the historical average water (100%) and 50% water reduction (treatment). Following the annual summer dry period, our SFA team collected and incubated soils at LLNL with four isotopic treatments ($^{12}C^{18}O$, $^{13}C^{18}O$, $^{12}C^{16}O$, $^{13}C^{16}O$) in 184 microcosms. '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 h) for DNA-quantitative stable isotope probing (qSIP), total CO_2, and $^{13}CO_2$ production.

While the differences between historical precipitation treatments were minimal for both total soil C respired and newly fixed CO_2 respired, we observed a large treatment effect on microbial response. Quantitative SIP analysis of 16S rRNA genes showed the prior spring's precipitation significantly affected cumulative total community growth and death rates following the fall wet-up. Overall, growth began more rapidly and sustained higher rates for 100% precipitation soils with 4.5 times more cumulative growth during the incubation. Mortality was also more rapid for 100% precipitation soils, with cumulative mortality 5 orders of magnitude larger in the 100% treatment. Population specific response patterns also differed across treatments. Growth response was dominated by Firmicutes, Actinobacteria, and Proteobacteria in 100% precipitation soils, whereas there was very little Actinobacteria response and delayed and reduced Firmicutes response in the 50% precipitation soils. Mortality was also markedly different, with 119 ASVs (amplicon sequence variant) found to have measurable death rates in 100% precipitation treatment, while only 4 ASVs were observed as dying in 50% treatment. To investigate mechanisms of mortality we are analyzing SIP-metagenomes to characterize microbial and viral activity and functionality following wet-up. Preliminary results show active viruses responded differently to our two treatments. Our results suggest historic precipitation patterns can have a significant impact on future microbial activities, turnover, and persistence. Future work will investigate the underlying mechanisms of different responses and their impact on soil C.

This research is based upon work supported by the LLNL 'Microbes Persist' Soil Microbiome SFA, funded 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 subcontracts to the University of California, Berkeley and Northern Arizona University. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

Cell Size Constraints on Microbial Ecophysiology and Implications for Soil Carbon Cycling

Authors: Gianna Marschmann*[1] (glmarschmann@lbl.gov), **Jinyun Tang[1], Ulas Karaoz[1], Jennifer Pett-Ridge[3], Eoin Brodie[1,2]**

Institutions: [1]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Department of Environmental Science, Policy, and Management, University of California, 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 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: Size is one of the most important biophysical traits influencing organismal ecology and evolution. Cell volume varies significantly across species and is related to patterns in genome, protein, cellular envelope, and ribosomal content that together can be used to predict key physiological traits of microorganisms that play important roles in soil organic matter turnover and stabilization. Here we coupled an existing allometric scaling model of cellular composition with a genome-informed dynamic energy budget (DEB) representation of microbial metabolism. Allometric predictions for protein synthesis rates that form the core of DEB metabolism were compared to pure culture growth efficiency experiments from the literature. In order to predict the scaling of traits and tradeoffs in rhizosphere and bulk soil communities, the kBASE microTrait software was used to extract fitness traits from isolate genome sequences for model parameterization.

Resource acquisition strategies of soil microorganisms may either be uptake-optimized when precursor compounds (e.g. from root exudation) are available, or depolymerization-optimized when resources are scarce or complex (e.g. from detritus). Due to the size contrast between microbial cells and low molecular weight compounds, uptake is limited mostly by accessible transporter proteins on the cell surface. We find that cellular carbon allocation to the production of transporter proteins, necessary to sustain a cell's maximum specific growth rate, increases with cell size. At the same time, the return on investment into constitutive extracellular enzyme production decreases with increasing cell size.

Microbial growth efficiency is an emergent property of numerous traits in addition to environmental conditions and resource properties. It has also been explicitly linked to the rate and yield of protein synthesis. Recent studies emphasize, that growth rate/yield trade-offs are not universal, but depend strongly on metabolic kinetics and environmental conditions. We find that our allometric scaling approach mirrors protein synthesis phenotypes from the literature. The emergent tradeoff curve of rate and yield as predicted from DEB theory, however, is nonlinear. A metabolic tradeoff between rate and yield is only apparent at high growth rates, while at low growth rates the shape of the tradeoff curve is dictated by storage compound accumulation. The size of the tradeoff region in which a combination of rate and yield is maximized increases with increasing cell size, suggesting higher flexibility in larger organisms consistent with the diffusion-constraint hypothesis in cell size evolution.

This work investigates the consequences of biophysical and physiological constraints in the DEB model by forcing it with different root exudation profiles in batch mode. We are investigating interactions between substrate preference and growth yield and compare emergent guild dynamics to existing metagenomic classifications of rhizosphere and bulk soil adapted communities.

On-going work is focused on coupling our DEB model to a plant-microbial model of soil organic matter turnover. Using a substrate-explicit modeling approach, we will examine biophysical and biochemical properties of low molecular weight compounds that are stabilized on mineral surfaces in distinct soil regimes. We anticipate the fully-coupled model to provide a rank-ordering of biophysical, life-history, and metabolic traits that distinguish microbial strategies. Together, this modeling concept will provide a platform to connect genome-level properties of organisms and phenotypic traits relevant to ecological fitness in the rhizosphere, and ultimately connect microbial ecological processes to soil biogeochemical function.

This research is based upon work supported by the LLNL 'Microbes Persist' Soil Microbiome SFA, funded 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. Part of this work was performed at Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

Prevalence and Automated Curation of Local Errors from Metagenomic Assembled Genomes

Rohan Sachdeva[1*] (rohansach@berkeley.edu), Livia Moura[2], Dylan Chivian[3], **Jennifer Pett-Ridge**[4], and **Jillian F. Banfield**[1,3]

[1]Innovative Genomics Institute, University of California, Berkeley CA; [2]University of São Paulo, São Paulo, Brazil; [3]Lawrence Berkeley National Laboratory, Berkeley CA; [4]Lawrence National 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 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: Metagenome assembled genomes (MAGs) recovered from complex communities are now the primary basis for understanding microbes in their natural environments. Consequently, the recovery of genomes that accurately reflect true biological organisms is essential. Contemporary metagenomic projects can produce thousands of genomes and the computational assemblies that comprise these mass-produced MAGs contain characteristic errors. Assembly errors perturb or even preclude functional predictions and accurate phylogenetic analyses and confound biochemical studies so limit the full potential utility of MAGs. These errors can be repaired, but because the process is time-consuming and requires human-guided manual curation, it is rarely performed. To address this, we are developing FixAME, a KBase software toolkit, for automatically curating and improving assemblies that does not require human-guided intervention. Additionally, to understand the prevalence of assembly errors, we quantified the number of assembly errors in assemblies from commonly used assembly programs across different environments. Following the complete development of FixAME we will be able to scale-up in order to curate and improve assemblies in the thousands of MAGs in public databases.

This research is based upon work supported by the LLNL 'Microbes Persist' Soil Microbiome SFA, funded 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 University of California, Berkeley. Additional funds were provided by a grant to encourage new tool development for the

DOE Systems Knowledgebase, KBase, SCW1656. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

The Influence of Drought on Carbon-Use Efficiency and Soil Carbon Formation in Rhizosphere and Detritusphere Microbial Communities

Noah Sokol*[1] (sokol1@llnl.gov), Megan Foley[2], Katarina Estera-Molina[3], Alex Greenlon[3], Eric Slessarev[1], Jose Liquet[1], **Bruce Hungate[2], Mary Firestone[3], Steve 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 Science, Policy, and Management, University of California, Berkeley, Berkeley CA

Website: https://sc-programs.llnl.gov/soil-microbiome

Project Goals: Soil organic matter (SOM) is critical to soil health and Earth's climate. Microorganisms play key roles in SOM turnover via their metabolic activities, cellular biochemistry, and extracellular products. We hypothesize that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape SOM 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: Microbial residues are dominant ingredients of persistent soil organic matter (SOM). Via the 'microbial carbon pump,' plant carbon from litter and rhizodeposition is processed by the microbial community *en route* to forming SOM. As the largest actively cycling reservoir of carbon, SOM is of critical importance to the global carbon balance and soil health. Yet, major uncertainty surrounds the microbial ecophysiological traits that regulate the microbial carbon pump, and how the relative importance of these traits varies in different soil microhabitats (e.g. the rhizosphere and detritusphere) and under different moisture conditions. Most attention to date has focused on the role microbial carbon-use efficiency (CUE), because this trait captures the partitioning of carbon to new microbial growth versus microbial respiration. For this reason, CUE has been widely posited to be positively related to MAOM formation. Yet there exists virtually no mechanistic evidence in support of this relationship to date.

We conducted a 12-week ^{13}C tracer study to track the movement of two dominant sources of plant carbon – rhizodeposition and root detritus – into soil microbial communities and SOM pools under normal moisture (15 ± 4.2 %) or droughted (8 ± 2%) conditions. Using a continuous $^{13}CO_2$-labeling system, we grew the annual grass *Avena barbata* in controlled growth chambers and measured formation of SOM from ^{13}C-enriched rhizodeposition. As the plants grew, we harvested rhizosphere and bulk soil at three time points (4, 8, and 12 weeks) to capture changes in SOM pools and microbial community dynamics. In a second set of microcosms, we tracked the formation of SOM derived from ^{13}C-enriched *A. barbata* root detritus during 12 weeks of decomposition, harvesting detritusphere and bulk soil at 4, 8, and 12 weeks. In a third set of

microcosms, we studied the combined influence of rhizodeposition and root detritus, separately tracking the contributions from each root C source using a reciprocal ^{13}C-labeling design.

For all harvest points, we density fractionated the soil to isolate the slower-cycling, mineral-associated organic matter (MAOM) fraction, as well as the faster-cycling particulate organic matter (POM) fraction. We also measured a suite of microbial ecophysiological traits that we predict are important in soil carbon formation and persistence. Here, we present data on ^{13}C-MAOM formation, as well as microbial CUE, measured via the ^{18}O-H_2O method.

We found CUE was on average 22% higher in the rhizosphere than in the detritusphere, and 34% higher under normal moisture versus drought conditions. The magnitude of difference in CUE between normal moisture vs. drought treatments also increased through time in both the rhizosphere (coeff. = 0.62; p=0.04) and detritusphere (coeff.=0.69, p=0.02). Similarly, we observed greater ^{13}C-MAOM formation under normal moisture vs. drought conditions in both the rhizosphere (25% increase) and detritusphere (13% increase). Notably, we did not find support for a positive relationship between CUE and ^{13}C-MAOM formation. Across treatments, there was a weak negative relationship between CUE and ^{13}C-MAOM formation ($r^2 = 0.11$, $p<0.01$). There were, however, key differences in both the direction and magnitude of this relationship based on microbial habitat, moisture status, and time (three-way interaction; F=3.75, p=0.05).

Overall, we found the CUE-MAOM relationship was context-dependent, and that CUE was often a poor predictor of MAOM formation. By extension, there may be other microbial ecophysiological traits (e.g. production of extracellular polymeric substances, growth rate) that better predict MAOM formation than CUE alone. This implies that new research must focus on a broader suite of microbial traits – including but not limited to CUE – to model the role of microbes in MAOM formation and SOM persistence.

This research is based upon work supported by LLNL 'Microbes Persist' Soil Microbiome SFA, funded 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 subcontracts to the University of California, Berkeley and Northern Arizona University. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

Virus Activity in Soil Revealed Through SIP-Metagenomics

Olivier Zablocki[1*] (zablocki.4@osu.edu), Alex Greenlon[2], Christine Sun[1], Ahmed Zayed[1], **Steven Blazewicz[2], Jennifer Pett-Ridge[2], Matthew Sullivan[1]**

[1]The Ohio State University, Columbus, OH; [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 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: Viruses are known to have critical ecosystem functions, via mechanisms such as nutrient cycling through host lysis, host metabolism reprogramming and shuffling of host genes via horizontal gene transfer [1,2]. In soils these mechanisms are under-explored, however recent work suggests that viruses may impact their surroundings by expressing a variety of auxiliary metabolic genes (AMGs) involved in carbon cycling [3] and even sporulation [4]. Here, we investigated the ecology of the viral component of bulk metagenomes sampled from three grassland Californian soils that occur along a precipitation gradient, and evaluated the subset of viruses that were 'active' via a stable isotope probing (SIP) experiment with ^{18}O-water.

Across the three sites, 8,617 viral operational taxonomic units ('vOTUs') were identified (a 5-fold gain over previous similar surveys), 37% of which were distributed into 495 novel viral clusters (approximate genus-level taxonomy). Virus diversity varied significantly (Shannon's H, p-value < 0.005) between the three soil sites, and was most diverse in the driest soil. Virus communities were strongly structured by soil site origin (PERMANOVA, p-value=0.001), and vOTU composition was strongly influenced by soil composition, pH and inorganic elements present. Host-linkage trends revealed that Actinobacteriota – specifically Mycobacteria - were the dominant host taxa (range: 68 – 83%) across the three soils. These were followed by Proteobacteria or Acidobacteria hosts, depending on the soil origin. The level of virus activity (determined via SIP incubation) varied significantly (range: 29 – 70% active viruses from total viruses) between soil sites, and followed an inverse relationship with soil moisture. Actinobacteriophages (mainly mycobacteriophages) were the most active viruses in two soils, except for Angelo. At the wettest site, viruses of Bacteriovoracia (order: Bdellovibrio; common soil-dwelling obligatory bacterial parasites) were among the most active. Uniquely in the driest soil site, we found active Archaeal viruses, predicted to infect members of the

Nitrososphaeraceae. This family plays important roles in nitrogen cycling, implying that viral infection could impact soil nutrient cycling. Lastly, 3% of the active pool of viruses encoded ecologically-relevant AMGs, including those for carbon utilization, nitrogen and energy metabolism. In summary, our data suggest that soils harbor substantial viral diversity, most of which is unknown; these viruses are actively replicating and can manipulate keystone microbial taxa responsible for nutrient transformation in soils.

References

1. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537:689–93.
2. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA Viral Macro- and Microdiversity from Pole to Pole. Cell. 2019;177:1109-1123.e14.
3. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, et al. Host-linked soil viral ecology along a permafrost thaw gradient. Nat Microbiol [Internet]. 2018;3:870–80. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30013236
4. Trubl G, Jang H Bin, Roux S, Emerson JB, Solonenko N, Vik DR, et al. Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. mSystems. 2018;3:1–21.

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 under Award Number SCW1632 to the Lawrence Livermore National Laboratory, and a subcontract to the Ohio State University. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

Transcriptomic and Metabolomic Analysis of Nitrogen and Carbon Metabolism in *Saccharomyces cerevisiae* and *Rhodotorula toruloides*

William Woodruff[1,2] (ww12@illinois.edu), Sujit S. Jagtap,[1,2] Jing-Jing Liu,[1,2] Anshu Deewan,[1,2] Hanna E. Walukiewicz,[1,2] Eun Ju Yun,[1,3] Yong-Su Jin,[2,4] and **Christopher V. Rao**[1,2]

[1]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana Champaign, Urbana; [2] DOE Center for Advanced Bioenergy and Bioproducts Innovation (CABBI); [3]Department of Biotechnology, Graduate School, Korea University, Seoul, Republic of Korea; [4]Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana

https://cabbi.bio/

Project Goals: The goal of this research is to understand the metabolic response of yeast to changes in nitrogen and carbon availability. Accumulation of valuable lipids and lipid-derived compounds in yeasts occurs during nitrogen starvation, and an improved understanding of nitrogen metabolism may enable development of engineered strains that can accumulate lipids in nitrogen-rich media. At the same time, the effects of varying carbon sources on the metabolism of oleaginous yeasts such as *R. toruloides* remain poorly characterized. In this study, the role of Ure2, a key regulator of nitrogen catabolite repression, was studied in a *ΔURE2* strain of *S. cerevisiae,* and the effects of changing carbon source on the metabolism of *R. toruloides* were investigated using a multi-omics approach.

The production of lipids and valuable oleochemicals in yeast is affected by changes in both nitrogen and carbon availability.[1,2] Lipid accumulation occurs under nitrogen limiting conditions, the response to which is mediated by the nitrogen catabolite repression (NCR) pathway.[2,3] Ure2 is a key transcriptional regulator of this pathway.[4] To investigate the role of this regulator, gene expression and intracellular metabolite concentrations were measured in a *ΔURE2* strain of *S. cerevisiae*. Transcriptomic changes consistent with nitrogen starvation and a selective autophagic response were observed. This mutant strain also accumulated less trehalose and glycogen, and it produced more lipid and ethanol. *URE2* is therefore a potential target for engineering yeast strains capable of lipid accumulation on nitrogen-rich substrates.

While *R. toruloides* can grow on a wide variety of substrates, the choice of medium substantially affects growth rates and product formation, while the metabolic basis for these differences remains poorly understood. To investigate these responses, gene expression and intracellular metabolite concentrations were measured on *R. toruloides* grown on glucose, xylose, acetate, and soybean oil. These substrates were chosen because they can all be obtained from plant biomass. Most observed changes were consistent with upregulation of known substrate utilization pathways; however, poor expression of xylulokinase was observed on xylose. This poor expression results in arabitol accumulation through an arabitol dehydrogenase bypass and opens the possibility of targeted metabolic engineering to improve xylose utilization.[5]

References

1. Wiebe, M.G. et al. Lipid production in batch and fed-batch cultures of *Rhodosporidium toruloides* from 5 and 6 carbon carbohydrates. *BMC Biotechnology* 12, 26, doi:10.1186/1472-6750-12-26 (2012).
2. Zhu, Z. et al. A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. *Nature Communications*, 1112, doi:10.1038/ncomms2112 (2012).
3. Yaegashi, J. et al. *Rhodosporidium toruloides*: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotechnology for Biofuels* 10, 241, doi:10.1186/s13068-017-0927-5 (2017).
4. Georis I. et al. The yeast GATA factor Gat1 occupies a central position in nitrogen catabolite repression-sensitive gene activation. *Molecular and Cellular Biology* 29(13):3803-15, doi:10.1128/MCB.00399-09 (2009).
5. Jagtap, S.S. and Rao, C.V., Production of D-arabitol from D-xylose by the oleaginous yeast *Rhodosporidium toruloides* IFO0880. *Applied Microbiology and Biotechnology* 102 (1), 143-151 (2018).

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.*

Nanoparticle-Mediated Transformation of Sorghum towards the Determination of a Subcellular Metabolic Network Map

Christopher T. Jackson[1,*], 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
[*]ctjackson@berkeley.edu

Project Goals: The goal of this project is to build an integrated pipeline to characterize metabolic interactions and pathways at a cellular level, using a combination of computational prediction, metabolic network modeling, and high-throughput experimental testing. This pipeline will be divided into three stages in order to develop a high-resolution subcellular map of small molecule metabolism in Sorghum and Brachypodium: a) generating localization predictions using bioinformatic algorithms, b) testing those predictions using nanotechnology mediated transformation of fluorescently tagged target proteins and high-sensitivity confocal imaging, and c) using the experimental data to generate new compartmentalized metabolic network models as well as refining existing pathway models. This project will initiate the creation of a repository for subcellular locations of metabolic enzymes, yielding important insight into the structure and function of metabolic networks in model systems as well as economically important crop species.

Understanding plant metabolic networks is essential to enable the efficient engineering of resilient and sustainable bioenergy crops. Although model species such as Arabidopsis have extensive resources to draw from, there is still lack of information in less well studied species such as sorghum. Sorghum is a challenging species to work with, as it has very poor transformation efficiencies. Here, we have been implementing new transformation methods which will allow us to rapidly test the bioinformatic predictions of enzyme subcellular locations. Initial tests using vectors with green fluorescent protein (GFP) under the control of *CaMV35S* and maize *Ubiquitin* promoters have shown transient expression of GFP in sorghum leaves, indicating successful carbon nanotube-mediated transformation. However, optimization is necessary due to inconsistencies between carbon nanotube (CNT) batches.

We identified that the polymers used to load vectors onto CNTs could be toxic to plants and play a role in batch-to-batch variability. We sought to quantitatively gauge and better understand these effects using associated toxicity marker genes. We identified that *PR1*, a marker for biotic stress, is upregulated in response to infiltration. Using RT-qPCR, we are exploring a library of amine-containing polymers to identify trends in chemical structure and functionality that minimize toxicity while providing an ideal platform for biomolecule conjugation to CNTs. At the same time, we are exploring covalent attachment methods, including EDC-NHS (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide, *N*-Hydroxysuccinimide) and triazine-based chemistries, to identify new avenues for the attachment of cationic polymer and subsequent electrostatic loading of biomolecular cargo.

We are exploring several alternatives to CNT-mediated leaf transformation. In order to increase throughput of transient transformation screenings, protocols for sorghum protoplast isolation and GFP delivery into protoplasts using CNTs have been tested. CNTs at the tested conditions are likely to be toxic as controls containing only CNTs also showed fluorescence. As an alternative to CNTs, we are optimizing particle bombardment parameters to deliver DNA into sorghum callus cells.

To broaden the knowledge base of plant transformation methods and to pivot our efforts from limited lab access as a result of COVID19, we have also written two articles: (1) an in-depth review on current sorghum biotechnology and challenges that need to be addressed to efficiently improve sorghum transformation and (2) a perspectives article on the use of nanotechnology to advance plant genetic engineering.

References:

1. Silva, T.N., Thomas, J.B., Dahlberg, J., Rhee, S.Y., Mortimer, J.C. Progress and challenges in sorghum biotechnology, a multi-purpose feedstock for the bioeconomy. In prep.
2. Demirer, G.S., Silva, T.N., Jackson, C.T., Thomas, J.B., Ehrhardt, D., Rhee, S.Y., Mortimer, J.C., Landry, M.P. Nanotechnology to advance CRISPR/Cas genetic engineering of plants. Nature Nanotechnology (2021). In press.

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

High-Throughput Determination of a Subcellular Metabolic Network Map of Plants

Jason Thomas[1*](jthomas@carnegiescience.edu), Suryatapa Ghosh Jha[1], Kevin Radja[1,3],Tallyta Silva[2], William Dwyer[1], Justin Krupp[1], Charles Hawkins[1], Angela Xu[1], Bo Xue[1], Christine Aquino[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; University of California, Berkeley, CA, [3]; Virginia Tech, Blacksburg, VA

Project Goals: The goal of this project is to build an integrated pipeline to characterize metabolic interactions and pathways at a cellular level, using a combination of computational prediction, metabolic network modeling, and high-throughput experimental testing. This pipeline will be divided into three stages in order to develop a high-resolution subcellular map of small molecule metabolism in Sorghum and Brachypodium: a) generating localization predictions using bioinformatic algorithms, b) testing those predictions using nanotechnology mediated transformation of fluorescently tagged target proteins and high-sensitivity confocal imaging, and c) using the experimental data to generate new compartmentalized metabolic network models as well as refining existing pathway models. This project will initiate the creation of a repository for subcellular locations of metabolic enzymes, yielding important insight into the structure and function of metabolic networks in model systems as well as economically important crop species.

Advances in our understanding of plant metabolism have underpinned many traits that contribute towards improving plant productivity. To identify, by predictive modeling and experimentation, and engineer desirable metabolic traits, such as maximizing biomass production under suboptimal conditions or reallocation of biomass from carbohydrates to lipids, we must decode the complex metabolic networks. Subcellular compartmentalization of metabolic reactions through the locations of enzymes is critical to understanding, modeling, and engineering plant metabolism. Yet, localization of the majority of the predicted enzymes in *S. bicolor* and *B. distachyon* is not yet known. The paucity of experimentally validated information in most plants, especially in the DOE flagship bioenergy plants, severely limits scientists and engineers to assess the performance and translatability of computational tools and resources.

In this project we have developed an integrated pipeline that combines computational prediction, metabolic network modeling, and high-throughput experimental testing using state of the art technologies in live confocal imaging, nanomaterial-mediated plant transformation with target metabolic enzymes, and metabolic network modeling. The enzyme localization prediction pipeline, named Compartmentalization Of METabolic networks (COMET), collects annotation data from existing metabolic pathway databases. For sorghum, the database SorghumbicolorCyc was created by using the E2P2 software to predict enzymatic function of the proteins in the sorghum genome sequence, then using the Pathologic and SAVI software to call the presence in sorghum of metabolic pathways from the Metacyc reference database. COMET implements a novel network-based classifier, MetaboLoc, to infer compartmentalization of metabolic pathways based on information available from GFP experiments and high-quality predictions from DeepLoc, an existing sequence-based classifier. When compared against DeepLoc, our classifier MetaboLoc achieved near equal performance in several of the subcellular compartments surveyed and

even outperformed the sequence-based classifier in the vacuole and Golgi apparatus. By circumventing reliance on sequence information, MetaboLoc enables the COMET pipeline to predict localization for over 95% of the total reactome, regardless of the species. The final compartmentalized network can serve as input for downstream analyses like metabolic domain enrichment or comparison between species

To validate COMET predictions, 48 candidate metabolic genes in *Arabidopsis thaliana* were chosen for subcellular localization. To achieve this, Gateway cloning technology was used to fuse these candidates with mCherry fluorescent tags, and subsequently expressed in *Nicotiana benthamiana* leaf cells using transient, Agrobacterium-mediated plant transformation techniques. Current efforts are focused on utilizing high resolution confocal live imaging (EMCCD spinning disk and Leica HyD point scanning) to determine the locations of these fusion proteins *in planta*, and to confirm these subcellular localization results using known organellar markers. The dataset collected from these validations will be used for developing the network maps and refine current models and to apply the methods to *S. bicolor* and *B. distachyon.*.

Overall, this project aims to holistically decipher the complexity of plant metabolic networks in order to engineer pathways to tackle impending challenges regarding climate change, food security, and the availability of sustainable energy sources.

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

Title: An Integrated Machine-Learning Framework for Reliable Host Prediction of Uncultivated Phages

Authors: Simon Roux[1] (sroux@lbl.gov), Andrew Tritt[2]

Institutions: [1]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Project Goals: Environmental viral diversity is quickly being mapped through large-scale metagenomics (meta)analyses. A major challenge of such approach, compared to traditional viral isolation, is the lack of host information for uncultivated viruses. Over the last several years, multiple tools have been released to predict host taxonomy from (partial) genome sequences of uncultivated viruses. Here, we review the major types of host predictions proposed so far, and describe the prototype of a new neural network framework able to integrate results from multiple tools in a single reliable host genus prediction.

Abstract text: Viruses are critical components of soil microbial ecosystems. By shaping microbial communities' structure and altering host cell metabolism during infection, viruses exert strong constraints on microbiomes, with downstream effects on nutrient cycling and metabolic outputs. Viral genomic diversity in the environment, especially in soil, is progressively being mapped primarily through the assembly of novel viral genomes from metagenomes. In the last five years alone, the number of such "uncultivated virus genomes" available in public databases has increased by more than 3 orders of magnitudes (from a few hundreds to several millions[1]). These genomes have enabled multiple discoveries on the diversity and distribution of viruses across different ecosystems, yet one limitation inherent to these datasets is the lack of host information for these viruses.

Linking novel uncultivated viruses to their host(s) is a critical step towards understanding the influence and potential impacts of these viruses on microbiomes and ecosystems. Accordingly, more than 15 different tools have been released over the last few years aiming at predicting virus:host pairs and/or host taxonomy for uncultivated viruses. Overall, these rely on four major types of genomic signal: (i) sequence similarity to known viruses, (ii) sequence similarity to putative host genomes, including due to host-encoded CRISPR-Cas systems, horizontal gene transfer, and provirus integration, and (iii) similarity in terms of nucleotide composition (i.e., k-mer frequency) between virus and host genome[2-5]. These different approaches differ greatly in their recall (i.e., ability to predict host taxonomy for as many viruses as possible) and false-discovery rate (i.e., frequency at which the predicted taxonomy correspond to the correct host), so that aggregating the results obtained from different tools on a single virus is challenging[6-7].

We describe here the prototype of a machine-learning framework taking as input multiple results of host prediction tools and using a deep neural network structure to provide as output a single host taxonomy prediction, at the genus rank, along with a confidence score. We illustrate how

this integration step maximizes both recall and precision, enabling robust host prediction for more input sequences than any individual tool. In addition, rather than directly predicting a host taxonomy, the neural network is designed to learn to distinguish patterns of "reliable" and "unreliable" host prediction based on a combination of all signals considered. Hence, it is not limited by the virus-host pairs currently described, and could be applied to entirely novel virus and host communities.

Overall, integrating independent signals for host prediction appears to be a promising approach to progressively populate uncultivated virus genome databases with reliable host information at a satisfactory taxonomic resolution (i.e., genus rank). Coupled with a continuing expansion of the global collection of isolated phages and innovative experimental methods to link uncultivated viruses to hosts, these pave the way towards a more comprehensive reconstruction of virus:host network from complex microbial communities.

References/Publications

1. Roux, Simon, David Páez-Espino, I-Min A Chen, Krishna Palaniappan, Anna Ratner, Ken Chu, T B K Reddy, et al. 2020. "IMG/VR v3: An Integrated Ecological and Evolutionary Framework for Interrogating Genomes of Uncultivated Viruses." Nucleic Acids Research, 1–12. https://doi.org/10.1093/nar/gkaa946.
2. Edwards, Robert A., Katelyn McNair, Karoline Faust, Jeroen Raes, and Bas E. Dutilh. 2016. "Computational Approaches to Predict Bacteriophage-Host Relationships." FEMS Microbiology Reviews 40 (2): 258–72. https://doi.org/10.1093/femsre/fuv048.
3. Galiez, Clovis, Matthias Siebert, François Enault, Jonathan Vincent, and Johannes Söding. 2017. "WIsH: Who Is the Host? Predicting Prokaryotic Hosts from Metagenomic Phage Contigs." Bioinformatics 33 (19): 3113–14. https://doi.org/10.1093/bioinformatics/btx383.
4. Ahlgren, NA, Jie Ren, Yang Young Lu, Jed A. Fuhrman, and Fengzhu Sun. 2016. "Alignment-Free d2 * Oligonucleotide Frequency Dissimilarity Measure Improves Prediction of Hosts from Metagenomically-Derived Viral Sequences." Nucleic Acids Research 45 (1): 39–53. https://doi.org/10.1093/nar/gkw1002.
5. Zhang, Ruoshi, Milot Mirdita, Eli Levy Karin, Clovis Norroy, Clovis Galiez, and Johannes Soeding. 2020. "SpacePHARER: Sensitive Identification of Phages from CRISPR Spacers in Prokaryotic Hosts." BioRxiv, 2020.05.15.090266. https://doi.org/10.1101/2020.05.15.090266.
6. Wang, Weili, Jie Ren, Kujin Tang, Emily Dart, Julio Cesar Ignacio-Espinoza, Jed A Fuhrman, Jonathan Braun, Fengzhu Sun, and Nathan A Ahlgren. 2020. "A Network-Based Integrated Framework for Predicting Virus–Prokaryote Interactions." NAR Genomics and Bioinformatics 2 (2): 1–19. https://doi.org/10.1093/nargab/lqaa044.
7. Zhang, Fan, Fengxia Zhou, Rui Gan, Chunyan Ren, Yuqiang Jia, Ling Yu, and Zhiwei Huang. 2019. "PHISDetector: A Web Tool to Detect Diverse in Silico Phage-Host Interaction Signals." BioRxiv. https://doi.org/10.1101/661074.

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. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

Non-Repetitive Promoters and Ribosome Binding Sites to Control *Clostridium autoethanogenum* Gene Expression Levels during Syngas Fermentation

Audrey Harris[1†*] (audrey.harris@lanzatech.com), Nick Fackler[1†], Rebecca O'Toole[2], Daniel Cetnar[2†], Ayaan Hossain[2], Steven D. Brown[1], Michael Köpke[1], and **Howard M. Salis**[2]
† These authors contributed equally to this work

[1]LanzaTech, Skokie, IL; [2]The Pennsylvania State University, University Park, PA

Project Goals: *Clostridium autoethanogenum* is an emerging platform organism that can convert C1 gas feedstock (e.g. waste gas) into valuable products, including acrylates and polymers. New genetic tools are needed to tune gene expression levels in *C. auto* for metabolic pathway and network engineering. For tunable transcriptional control, we constructed over 1000 highly non-repetitive promoters with *C. auto* specifications, characterized their transcription rates during syngas fermentation, and validated these measurements as compared to natural *C. auto* promoters. For tunable translational control, we constructed a toolbox of designed, synthetic ribosome binding sites and characterized their expression levels. Altogether, we demonstrate the ability to tune the expression levels of many proteins across over a 100,000-fold range in *C. auto* without introducing repetitive DNA sequences, facilitating metabolic pathway and network optimization.

C. autoethanogenum is a non-model organism that has become an emerging platform for industrial-scale bioconversion of C1 gas feedstock into valuable chemicals. Successes using *C. auto* include operation of a waste gas to ethanol commercial plant at a scale of 16 million gallon/year and successful demonstration of a pilot plant that utilizes gasified waste streams from lignocellulosic biomass and municipal solid waste (MSW) (1-2). As new pathways are introduced into *C. auto* to produce valuable products, its gene expression levels must be tuned to maximize production titers and growth rates. New genetic parts are needed that have highly non-repetitive sequences (to avoid homologous recombination) as well as characterized transcription and translation rates to tune gene expression levels.

Expanding on our previous work, we designed, constructed and characterized over 1000 highly non-repetitive promoters that varied *C. auto* transcription rates by over 100,000-fold during syngas fermentation conditions. We applied our new algorithm, the Non-Repetitive Parts Calculator, to carry out this rational design (3). We selected a representative subset of these non-repetitive promoters as well as a set of commonly used *C. auto* promoters and additionally measured their transcription rates. Separately, we designed, constructed, and characterized a set of synthetic ribosome binding sites, designed by the RBS Calculator v2.1 model, and characterized their translation rates using GusA and NanoLuc reporter assays in both *E. coli* and *C. auto*.

Altogether, these non-repetitive genetic parts will be utilized to tune enzymes expression levels in multi-enzyme pathways in *C. auto* as well to construct Extra-long sgRNA Arrays (ELSAs) to carry out many-gene CRISPRi for rewiring of metabolic fluxes (4, 5). We are also utilizing these datasets to test and improve biophysical models of gene expression in AT-rich genomes, such as *C. auto*.

References/Publications:

1. Köpke, M., Simpson, S.D. Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Curr Opin Biotechnol* **65**,180-189. https://doi.org/10.1016/j.copbio.2020.02.017
2. Fackler, N., Heijstra, B., Rasor, B., *et al.* Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. *Annual Review of Chemical and Biomolecular Engineering*, (in review).
3. Hossain, A., Lopez, E., Halper, S.M. *et al.* Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. *Nat Biotechnol* **38,** 1466–1475 (2020). https://doi.org/10.1038/s41587-020-0584-2.
4. Fackler, N., Heffernan, J., Juminaga, A., *et al.* Transcriptional Control of Clostridium autoethanogenum using CRISPRi. *Synthetic Biology*, (in review).
5. Reis, A. C., Halper, S. M., Vezeau, G. E., Cetnar, D. P., Hossain, A., Clauer, P. R., & Salis, H. M. (2019). Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nature Biotechnology*, *37*(11), 1294–1301. https://doi.org/10.1038/s41587-019-0286-9.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0019090 for funding of this project.

Multiplex Genome Engineering for Bioproduction of 3-Hydroxypropionic Acid and 1,3-Propanediol from Waste Gases

Fungmin (Eric) Liew (Eric.liew@lanzatech.com),[1*] Hunter Zeleznik,[1], Megan Garber,[1] Rasmus Jensen,[1] Loan Tran,[1] Archer Smith,[1] Steven D. Brown,[1] **Howard M. Salis[2]**, and Michael Köpke[1]

[1]LanzaTech, Skokie, IL; [2]The Pennsylvania State University, University Park, PA.

Project Goals: Gas fermentation is a commercially scalable platform for the sustainable biomanufacturing of valuable chemicals from abundant, low cost C1 feedstocks (1,2). In this project, we aim to engineer gas-utilizing Clostridia strains to produce 3-hydroxypropionic acid (3-HP), which is an ideal bio-renewable precursor to acrylates and polymers (acrylonitrile, acrylamide, acrylic acid and acrylate esters) with a global market estimated as 3.6 million tons per year. Through combinatorial analysis, we compared the levels of 3-HP production from different biosynthesis routes. In order to systematically redirect metabolic flux towards 3-HP biosynthesis, we developed and employed Extra-Long sgRNA Arrays (ELSAs) (3) to knockdown gene expression levels & reduce undesired reaction fluxes, guided by genome-scale metabolic modeling (GSM). The expected outcome of this project is the development of sophisticated multiplex genome editing tools for a non-model microorganism and a bioprocess that efficiently converts industrial waste gas into 3-HP at commercially relevant productivity and selectivity.

Gas fermentation has emerged as a promising biorenewable platform for manufacturing valuable chemicals from gaseous, non-food feedstocks that would normally be considered pollutants or waste. These gases include carbon dioxide (CO_2; a greenhouse gas) and carbon monoxide (CO; a harmful pollutant that will be oxidized to CO_2 when released in the atmosphere). LanzaTech is a worldwide leader in gas fermentation having commercialized and scaled up the production of ethanol from CO/CO_2 gas mixtures using *Clostridium autoethanogenum* as the whole-cell biocatalyst. Gas feedstocks are sourced from lignin-derived syngas, steel mill waste gas, and biorefinery waste gas, providing ample commercial opportunities for upgrading negative value pollutants into valuable co-products (1,2).

In this project, we are engineering industrial *C. autoethanogenum* strains to manufacture 3-hydroxypropionic acid (3-HP) with commercially relevant metrics (selectivity, titer, yield, and productivity) from a syngas feedstock. Following evaluation of 25 different 3-HP biosynthesis pathways using our customized GSM, we focused our effort on two metabolic routes towards 3-HP. A highly efficient, modular cell-free expression plasmid assembly system (4) was employed to build two combinatorial libraries (up to 3780 permutations per library), which were subjected to plasmid sequencing to investigate promoter-gene diversity. These libraries were then transformed into *C. autoethanogenum* generating >1000 strains.

Screening of these combinatorial strains in our high-throughput, automated anaerobic biofoundry (1) showed strains that produced 3-HP and its downstream product, 1,3-propanediol (1,3-PDO) at various levels. 1,3-PDO is an important chemical with market size of $490 million and we previously showed that conversion of 3-HP to 1,3-PDO in *C. autoethanogenum* occurs via

aldehyde::ferredoxin oxidoreductase (AOR) enzymes (5). A LC-MS method was developed to confirm the identity of 3-HP and measure pathway intermediates to aid metabolic engineering efforts. In continuously stirred tank reactor (CSTR) using synthetic gas blend as feedstock, these recombinant strains produced 3-HP and 1,3-PDO at high titer and productivity. To further enhance the production of 3-HP, we are employing Extra-Long sgRNA Arrays (ELSA) (3) to simultaneously knockdown multiple gene expression levels to reduce competing metabolic fluxes, guided by our GSM (6).

References/Publications

1. Köpke, M., and Simpson, S.D. (2020). Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Current Opinion in Biotechnology* **65**: 180-189.
2. Fackler, N., Heijstra, B., Rasor, B., *et al.* Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. *Annual Review of Chemical and Biomolecular Engineering*, (in review).
3. Reis, A.C., Halper, S.M., Vezeaus, G.E. et al. (2019). Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. *Nature Biotechnology* **37**(11): 1294-1301.
4. Karim, A.S., Liew, F., Garg, S. *et al*. (2020). Modular cell-free expression plasmids to accelerate biological design in cells. *Synthetic Biology* **5**(1): ysaa019.
5. Liew, F., Henstra, A.M., Köpke, M. *et al.* (2017). Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metabolic Engineering* **40**: 104-114.
6. Marcellin, E., Behrendorff, J.B., Nagaraju, S. (2016). Low carbon fuels and commodity chemicals from waste gases – systematic approach to understand energy metabolism in a model acetogen. Green Chemistry **18**: 3020-28.

We acknowledge the U.S. Department of Energy Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program for funding of this project under Contract DE-SC0019090 for funding of this project.

Elucidation of aromatic catabolic pathways in white-rot fungi

Carlos del Cerro[1], Erika Erickson[1], Tao Dong[1], Kelsey J. Ramirez[1], Allison R. Wong[2], Elizabeth K. Eder[2], Samuel O. Purvine[2], Hugh D. Mitchell[2], Karl K. Weitz[2], Lye Meng Markillie[2], Meagan C. Burnet[2], David W. Hoyt[2], Rosalie K. Chu[2], Jan-Fang Cheng[3], and **Davinia Salvachúa**[1*] (davinia.salvachua@nrel.gov)

[1]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO; [2]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA; [3]Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA

This project aims to investigate the hypothesis that white-rot fungi can simultaneously depolymerize lignin extracellularly and catabolize depolymerization products intracellularly as carbon and energy sources. Evaluating this hypothesis will provide deeper understanding of the role of white-rot fungi in facilitating carbon sequestration in Nature. Additionally, identifying the most promising fungal strains for lignin turnover and catabolism will catalyze future efforts in genetic tool development to enable metabolic engineering in white-rot fungi for lignin bioconversion to bioproducts.

Lignin is the second most abundant plant-based biopolymer on Earth and represents up to 40% of the energy density of lignocellulosic biomass. Even though lignin is a massive natural carbon and energy reservoir, only a small group of basidiomycete fungi, namely white-rot fungi (WRF), have evolved the ability to efficiently depolymerize and mineralize lignin to CO_2 and H_2O. Considerable research efforts have been undertaken to understand how WRF depolymerize lignin but the biochemical reactions that convert lignin into CO_2 have been largely neglected. In fact, it is unclear if WRF intracellularly catabolize lignin-derived aromatic compounds to utilize them as a carbon and energy source, or rather if lignin is depolymerized and mineralized extracellularly merely to facilitate access to cellulose and hemicellulose for use as a primary carbon source.

To date, we have employed ^{13}C-isotope labeling, systems biology approaches, and *in vitro* enzyme assays to definitively demonstrate that two WRF, *Trametes versicolor* and *Gelatoporia (Ceriporiopsis) subvermispora*, funnel carbon from lignin-derived aromatic compounds into central carbon metabolism via intracellular catabolic pathways [1]. Specifically, ^{13}C-isotopic labeling approaches showed that these WRF utilize poplar-derived aromatic compounds (e.g. 4-hydroxybenzoic acid (4-HBA)) as a carbon source. *In silico* genome analysis led us to hypothesize a complete catabolic pathway for 4-HBA and identify multiple homologous sequences for enzymes with putative oxidative decarboxylase, hydroxylase, and ring-opening dioxygenase activities, which are among the main biochemical reactions acting on aromatic compounds. Spatial and differential proteomic and metabolomic analyses supported the proposed catabolic pathways and showed alternative catabolic steps in *T. versicolor* that were not present in *G. subvermispora*.

Based on the *in silico*, proteomics, and transcriptomics results, we down-selected enzymes for further *in vitro* characterization, and we have assigned a function to six fungal enzymes (including oxidative decarboxylases, hydroxylases, and ring-opening dioxygenases). Interestingly, even though we selected homologous enzyme pairs from both WRF with similar -omics trends, in a few cases only one of the studied fungi showed activity for the proposed substrate. Based on all the observations from this study [1], we hypothesized that 4-HBA preferentially undergoes oxidative decarboxylation to hydroquinone and subsequent hydroxylation to 1,2,4-benzenetriol in *G. subvermispora* before ring cleavage, whereas 4-HBA would preferentially undergo hydroxylation to protocatechuate and further oxidative decarboxylation to 1,2,4-benzenetriol in *T. versicolor*. Examining additional decarboxylases and hydroxylases as well as enzymes from other protein families that can perform the same or similar oxidative reactions, such as cytochromes P450 with aromatic hydroxylation activity, will also be key for elucidating enzyme preferences for specific substrates. Overall, this work forms the foundation of a new research area based on lignin catabolism by WRF, which could be further exploited to convert the undervalued biopolymer lignin into value-added compounds.

Publication

[1] Carlos del Cerro, Erika Erickson, Tao Dong, Allison R. Wong, Elizabeth K. Eder, Samuel O. Purvine, Hugh D. Mitchell, Karl K. Weitz, Lye Meng Markillie, Meagan C. Burnet, David W. Hoyt, Rosalie K. Chu, Jan-Fang Cheng, Kelsey J. Ramirez, Rui Katahira, Wei Xiong, Michael E. Himmel, Venkataramanan Subramanian, Jeffrey G. Linger, and Davinia Salvachúa. Intracellular pathways for lignin catabolism in white-rot fungi. ***PNAS***, in press. **2021**.

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. A portion of the research has been performed using EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. A portion of the work has been also conducted by the U.S. DOE Joint Genome Institute, a DOE Office of Science User Facility, supported by the DOE Office of Science under Contract No. DE-AC02-05CH11231. Part of this work was also supported by a Laboratory Directed Research and Development project at the National Renewable Energy Laboratory, operated by the Alliance for Sustainable Energy, LLC for the U.S. DOE under contract DE-AC36-08GO28308.

Using aggregated field collections data and the novel R package "fungarium" to investigate fungal traits

Hunter J. Simpson, **Jonathan S. Schilling**[1] (schillin@umn.edu), Claudia Schmidt-Dannert[1], Jiwei Zhang[1], David Hibbett[2], Igor Grigoriev[3], Young-Mo Kim[4]

[1]University of Minnesota, Saint Paul, [2]Clark University, [3]Joint Genome Institute (JGI) [4]Pacific Northwest National Laboratory (PNNL)

URL: http://schillinglab.cfans.umn.edu

Project Goals: Fungi dominate the biological decomposition of wood and other lignocellulosic plant tissues in nature. These saprotrophs offer us a proven model for making energy, sustainably, from biomass. They also offer those with commercial interests a range of pathways for unlocking sugars embedded in lignin. Their strategies range from 'white rot' mechanisms that remove lignin to gain access to polysaccharides to 'brown rot' mechanisms that selectively extract sugars, leaving most lignin behind. This metabolic diversity could be harnessed, industrially, but research has generally been focused more toward white rot delignification pathways. White rot fungi can unsheath polysaccharides by selectively removing lignin, a capacity that historically attracted interest for the potential to extract intact fibers for papermaking. Modern bioenergy schemes, however, do not aspire for intact fibers - instead, the goal is to depolymerize polysaccharides to release fermentable sugars (saccharification), saving lignin as a co-product, if possible. This is a better fit for the carbohydrate-selective pathways of brown rot fungi, but our grasp of fungal brown rot metabolism lags behind what we know about white rot.

Our collaborative project is aligned to address these gaps, with the **goal** of producing an integrated regulatory model for brown rot. Our proposed objectives insure stand-alone advances, but will also synergize to push ideas forward in a systems context.

Objective 1 is to identify fungal gene regulation patterns that distinguish brown rot fungi from fungi with other decay modes (e.g., white rot). We are comparing fungi among relevant lineages but with varied carbohydrate-selectivities. We are culturing these strains on solid wood wafers, spatially mapping gene expression and then overlaying fungal/wood metabolite patterns to enable temporally-resolved functional genomics. These maps can isolate patterns unique to brown rot and can target characterization.

Objective 2 focuses on characterization, starting with a short list generated in an earlier transcriptomics study, and progressively adding objective 1 gene targets. We are using routine single-/multi-cellular *in vitro* transformation pipelines, but complementing this with efforts to develop a brown rot transformation system, enabling *in vivo* manipulations (e.g., Crispr-Cas9).

Objective 3 is to use metabolomics to map metabolite-expression feedback over time, providing networks of gene regulation. This approach promises to advance our understanding of this unique brown rot strategy, beyond current ROS-centric models toward a systems view.

Abstract:

Fungal traits offer predictive characters useful in ecology and for prospecting new strains for bioconversion. Archived sporocarp data, like the collection and observation records accessible through The Mycology Collections Portal (MyCoPortal), are well-suited for trait investigations considering that these records circumvent the need for field work, are geographically and temporally diverse, and often have detailed and trait-relevant environmental metadata.

There are, however, inefficiencies and inadequacies in the MyCoPortal online interface that affect dataset generation and trait searching, and many of the available records have outdated or misspelled taxon names as well as misspelled location names. Thus, we created the R package *fungarium*, which enables the efficient download of complete MyCoPortal datasets from within the R environment, enhances the identification of trait-relevant records, confirms or updates taxon names based on current scientific consensus while also accounting for spelling errors, and fixes misspelled location names. Utilizing this package and MyCoPortal data, we demonstrated methods for analyzing taxonomic, geographic, and temporal patterns in ecological traits, using fire-association as an example.

We found that fire-association, which was quantified via fire-associated enrichment factors (fire-associated records/total records), differed substantially between taxa and these differences were qualitatively supported by existing literature, as hypothesized. Fire-association varied between counties and years as well, but these patterns did not correlate with burned acreage as expected. This lack of correlation was linked to sampling bias within the MyCoPortal data and limitations of the burned acreage dataset used (i.e. Monitoring Trends in Burn Severity). However, both confounding factors are likely depend on the trait analyzed and external dataset used. Overall, the *fungarium* package and associated methods presented here effectively enable the use of archived sporocarp data for future ecological trait studies.

References:

Simpson, H.J. and **Schilling, J.S.** (in press) Using aggregated field collections data and the novel R package "fungarium" to investigate fungal fire association. Mycologia

Castaño, J.D., Zhang, J., Anderson, C.E., and **Schilling, J.S.*** (2018) Oxidative damage control as a brown rot fungus attacks wood using oxygen radicals. Applied and Environmental Microbiology 84(22):e01937-18

Zhang, J., Figueroa, M., **Castaño, J.D.,** Silverstein, K., **Schilling, J.S.*** (2019) Gene regulation shifts shed light on fungal adaptation in plant biomass decomposers. mBio 10:e02176-19.

Funding statement

This work was supported by the U.S. Department of Energy Office of Science Grant DE-SC0019357 from the Office of Biological and Ecological Research (BER) to J.S.S., by user Facility grant #48607 at the Environmental Molecular Sciences Laboratory (to J.S.S.), by NSF as a Graduate Research Fellowship (to C.E.A.), and by Fulbright Colombia (to J.C.).

Meta-Analysis Identifies Pleiotropic Loci Controlling Phenotypic Trade-offs in Sorghum

Ravi V. Mural[1*] (rmural2@unl.edu) Marcin Grzybowski[1], Chenyong Miao[1], Alyssa Damke[2], Sirjan Sapkota[3,4], Richard E. Boyles[4,5], Maria G. Salas Fernandez[6], Patrick S. Schnable[6], Brandi Sigmon[2], Stephen Kresovich[4,7], and **James C. Schnable**[1]

[1]Center for Plant Science Innovation and Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE USA; [2]Department of Plant Pathology, University of Nebraska-Lincoln, Lincoln, NE USA; [3]Advanced Plant Technology Program, Clemson University, Clemson, SC USA; [4]Department of Plant and Environment Sciences, Clemson University, Clemson, SC USA; [5]Pee Dee Research and Education Center, Clemson University, Florence, SC, USA; [6]Department of Agronomy, Iowa State University, Ames, IA USA; and [7]Feed the Future Innovation Lab for Crop Improvement Cornell University, Ithaca, NY USA

Project Goals:
Here we sought to assemble publicly accessible and internally generated trait datasets from a common association population of sorghum to identify genes associated with phenotypic variation in one or more phenotypes to serve as a training dataset for machine learning models which will predict which additional sorghum genes most likely to exhibit loss of function phenotypes.

Abstract text.

Most of the agronomic traits are complex in nature and are influenced by many genetic loci and environmental conditions as well as their interactions. However, there are some instances where some traits show correlated variations as they are influenced/controlled by same genetic loci/gene which is pleiotropic in nature. Not only the investigation and understanding of the genetic association of complex traits but also pleiotropy and genotype by environment interactions are of paramount importance in strategizing the breeding program for crop improvement. The current approaches to link phenotypic variation to natural genetic variation using quantitative genetic tools are primarily conducted one trait at a time, because scoring multiple traits across large population is labor intensive and economically challenging. Also, most of the individual research groups focus on few of the specific traits of interests. Thus to bring together all the traits scored on a community population we conducted a meta-analysis of Genome Wide Association Studies (GWAS), employing a set of 234 separate trait datasets, including both published and unpublished trait datasets for the Sorghum Association Panel (SAP) with 406 sorghum genotypes, and multiple genetic marker datasets to empirically evaluate both the degree of saturation achieved by current genetic marker sets and the degree to which detectable loci controlling phenotypic variation in the SAP tend to be pleiotropic or non-pleiotropic using a multi-trait approach based on meta-analysis and adaptive shrinkage. Our results based on comparison of GWAS conducted with two independently generated marker sets to screen this population, suggest that the existing genetic marker sets are not enough to saturate the genome and capture only 35 to 45% of potentially detectable loci controlling variation for traits scored in this population. Cross-GWAS comparisons showed a limited evidence of pleiotropy, thus a multivariate adaptive shrinkage approach was adopted. The multivariate approach aided in recovering both known pleiotropic effects of existing loci and various new

pleiotropic effects, which might go unaccounted in direct comparison of GWAS results. The known pleiotropic effects detected in this study included the significant impacts of known dwarfing genes on root architecture, and the new pleiotropic loci detected were consistent with known trade-offs in sorghum development. These results demonstrate the potential of detecting new associations as new, denser genetic marker datasets are generated for the Sorghum Association Population.

Funding statement.
This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant no. DE-SC0020355 to JCS and BS

Overexpression of the Phosphatidylcholine: diacylglycerol cholinephosphotransferase (*PDCT*) Gene Has Increased Carbon Flux Toward Triacylglycerol (TAG) Synthesis in Camelina (*Camelina sativa*) Seeds

*Hesham M. Abdullah[1] (abdull76@msu.edu), Na Pang[1], Benjamin Chilcoat[2], Yair Shachar-Hill[1], Om Parkash Dhankher[2], and **Danny J. Schnell**[1]

[1]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA
[2]Stockbridge School of Agriculture, University of Massachusetts Amherst, MA 01003, 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 aims to increase fixed carbon allocation to triacylglycerol (TAG) by identifying metabolic bottlenecks that control seed oil production and engineering these limiting steps to increase TAG production in developing Camelina seeds. We describe here the positive impact of engineering the *PDCT* gene, encoding phosphatidylcholine:diacylglycerol cholinephosphotransferase 1, on both the levels and composition of Camelina seed oil.

Abstract

In our previous studies, we utilized metabolomic and transcriptomic profiling approaches in developing Camelina seeds to further understand the routes, rates, and regulation of pathways that determine seed and oil yields to increase Camelina's productivity[3]. We revealed the potential limiting factor(s) in oil synthesis pathways, and accordingly, we selected several candidate genes/enzymes for metabolic engineering of Camelina. Among those genes, in the described work, we targeted the overexpression of the *PDCT* gene, a homolog to the Arabidopsis Reduced Oleate Desaturation 1, *ROD1* gene (TAIR ID: AT3G15820), which encodes Phosphatidylcholine:diacylglycerol cholinephosphotransferase 1 enzyme[4].

PDCT is proposed to act as a gate keeper responsible for the interconversions of diacylglycerol (DAG) and phosphatidylcholine (PC) pools[4,5]. On this basis, we hypothesized that increased PDCT activity in developing Camelina seeds would enhance metabolic carbon flux toward increased levels of TAG and alter oil composition to make it more compatible with the proposed industrial uses of Camelina oils. To test this hypothesis, we engineered Camelina by expressing its *PDCT* gene under the control of the seed-specific phaseolin promoter. Transgenic Camelina plants exhibited significant increases in seed mass and seed oil content, overall higher seed and oil yields and altered polyunsaturated fatty acid (PUFA) content compared to their parental wild-type (WT) plants. Further, the preliminary results from embryo culturing coupled with [^{14}C]acetate labeling of developing Camelina embryos indicated increased rates of fatty acid incorporation into glycerolipids. This resulted in higher total radiolabeled lipid content in the PDCT transgenic lines, particularly in the TAG and DAG lipid classes, relative to that in WT embryos.

We conclude that overexpression of the TAG biosynthetic gene, PDCT, appears to be a positive strategy to achieve a synergistic effect on the flux through the TAG synthesis pathway, thereby further increasing oil yields in *Camelina sativa*.

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. Abdullah, H.M., Chhikara, S., Akbari, P. et al. Comparative transcriptome and metabolome analysis suggests bottlenecks that limit seed and oil yields in transgenic Camelina sativa expressing diacylglycerol acyltransferase 1 and glycerol-3-phosphate dehydrogenase. Biotechnol Biofuels 11, 335 (2018). https://doi.org/10.1186/s13068-018-1326-2

4. Lu C, Xin Z, Ren Z, Miquel M, Browse J. An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of Arabidopsis. Proc Natl Acad Sci U S A. 2009 Nov 3;106(44):18837-42. doi: 10.1073/pnas.0908848106. Epub 2009 Oct 15. PMID: 19833868; PMCID: PMC2774007.

5. Yang W, Wang G, Li J, Bates PD, Wang X, Allen DK. Phospholipase Dζ Enhances Diacylglycerol Flux into Triacylglycerol. *Plant Physiol*. 2017;174(1):110-123. doi:10.1104/pp.17.00026

Acknowledgment: This research is funded by the Department of Energy (ARPAe DE-0000200 & BER DE-SC0018269)

The metabolic origins of non-photorespiratory CO_2 release during photosynthesis: A metabolic flux analysis

* Yuan Xu[1] (xuyuan5@msu.edu), Xinyu Fu[2], Thomas D. Sharkey[2, 3], **Yair Shachar-Hill[1]**, Berkley J. Walker[1,2]

[1]Department of Plant Biology, [2]Plant Research Laboratory, [3]Department of Molecular Biology and Biochemistry
Michigan State University, East Lansing, MI 48824, USA

URL: https://groecamelina.natsci.msu.edu/

Project Goals

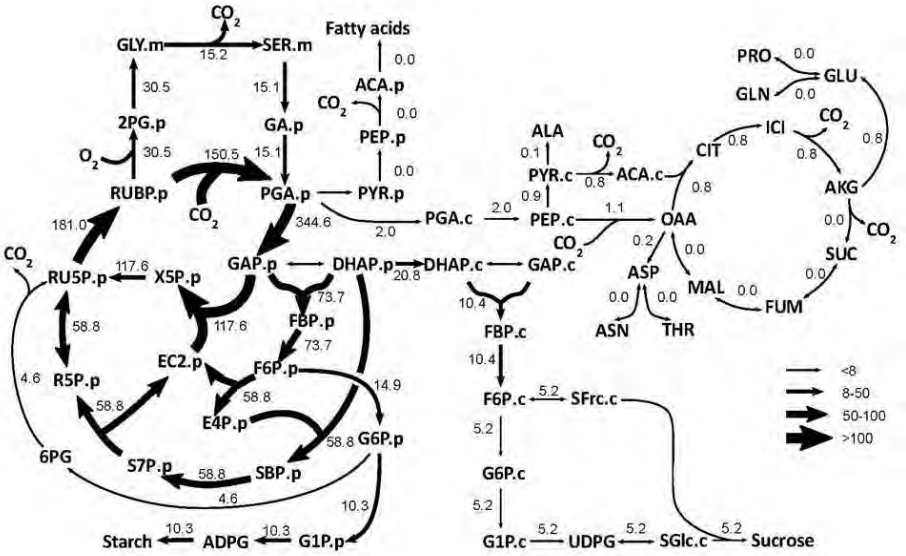
The overall goal of this project is to increase the triacylglycerol yield of the model oilseed crop plant, *Camelina sativa*, to increase its usefulness for producing fuels and chemical feedstocks. Camelina shows promise as a biofuel crop and is widely used as a model oilseed plant. A near relative of *Brassica napus* and *Arabidopsis thaliana* it is easily transformed, requires low agronomic inputs, and is naturally resistant to both biotic and abiotic stress; however its yields are lower than major oilseed crops. The aims of this sub-project were to establish and improve metabolic flux analysis tools to quantify fluxes through central metabolism in photosynthesizing Camelina leaves and to apply this approach to determine the source(s) of non-photorespiratory CO_2 release in the light, which lowers photosynthetic efficiency.

Abstract

Respiration in the light (R_L) releases CO_2 in photosynthesizing leaves and occurs independently from photorespiration. Since R_L lowers net carbon fixation, understanding it could help improve plant carbon-use efficiency and modeling of crop photosynthesis. Although R_L was identified more than 75 years ago, its biochemical mechanisms remain unclear. To identify reactions contributing to R_L, we mapped metabolic fluxes in photosynthesizing source leaves of the oilseed crop and model plant camelina (*Camelina sativa*). We performed a flux analysis using $^{13}CO_2$ isotopic labeling patterns of central metabolites during time course, gas exchange and carbohydrate production rate experiments. To quantify the contributions of multiple potential CO_2 sources with statistical and biological confidence, we increased the number of metabolites measured and reduced biological and technical heterogeneity by using single mature source leaves and quickly quenching metabolism by directly injecting liquid N_2; we then compared the goodness-of-fit between these data and data from models with alternative metabolic network structures and constraints. Our analysis predicted that R_L releases 5.2 μmol g^{-1} FW hr^{-1} of CO_2, which is consistent with a value of 9.3 μmol g^{-1} FW hr^{-1} estimated by CO_2 gas exchange. The flux analysis indicated that ≤10% of R_L results from TCA cycle reactions, which are widely considered to dominate R_L. Further analysis of the results indicated that oxidation of glucose-6-phosphate to pentose phosphate via 6-phosphogluconate (the G6P/OPP shunt) can account for >93% of CO_2 released by R_L.

The methods established in this study are being applied in the broader research on improving Camelina productivity to: (a) measuring changes in leaf central metabolism in transgenic Camelina plants with increased rates of CO_2 assimilation; (b) mapping leaf carbohydrate turnover during photosynthesis; and (c) to provide experimentally derived flux maps for improvement of predictive stoichiometric flux analysis by Flux Balance Analysis.

Central carbon assimilatory metabolic fluxes in photosynthetic *Camelina sativa* leaves.



Fluxes are shown in numbers and also depicted by the variable width of arrows. Fluxes were estimated by ^{13}C INST-MFA using the INCA software suite constrained by the metabolic network model and experimental inputs including mass isotopomer distributions of measured metabolites, net CO_2 assimilation, starch synthesis rate, sucrose synthesis rate and amino acid export rate. Fluxes were not constrained by measured R_L. Flux units are expressed as μmol metabolite g FW^{-1} hr^{-1}. The model network is compartmentalized into cytosol (".c"), which includes mitochondrial and peroxisomal reactions, plastid (".p"), and mitochondria (".m"). Metabolite pools (principally vacuolar) that do not become labeled on the time scale of the experiments are modeled but not shown here.

References

1. **Laisk AK (Agu K** (1977) Kinetika fotosinteza i fotodykhaniia C3- rastenii.

2. **Ma F, Jazmin LJ, Young JD, Allen DK** (2014) Isotopically nonstationary ^{13} C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. Proc Natl Acad Sci **111**: 16967–16972

3. **Preiser AL, Fisher N, Banerjee A, Sharkey TD** (2019) Plastidic glucose-6-phosphate dehydrogenases are regulated to maintain activity in the light. Biochem J. doi: 10.1042/BCJ20190234

4. **Ratcliffe RG, Shachar-Hill Y** (2006) Measuring multiple fluxes through plant metabolic networks. Plant J. doi: 10.1111/j.1365-313X.2005.02649.x

5. **Young JD** (2014) INCA: A computational platform for isotopically non-stationary metabolic flux analysis. Bioinformatics. doi: 10.1093/bioinformatics/btu015

Acknowledgment: This research is funded by the Department of Energy (ARPAe DE-0000200 & BER DE-SC0018269)

Integration of a Synthetic CO_2 Fixation Cycle into *Camelina sativa*

Nathan Wilson,[1*] (njwilso6@ncsu.edu) Heike Sederoff [1*](hwsedero@ncsu.edu) Amy Grunden,[1] Brianne Edwards,[1] Swathi Barampuram,[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

https://groecamelina.natsci.msu.edu/

Project Goals: To overcome the limitations of photosynthetic CO_2 fixation via Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO) in plants, we created a RuBisCO-independent synthetic CO_2 fixation cycle based on enzymes from bacterial autotrophs. This condensed, reversed CO_2 fixation cycle consists of 5 enzymes expressed in the nucleus and imported into chloroplasts to generate glyoxylate from succinate (1). We show here the integration of partial and complete crTCA cycle enzyme in Camelina sativa chloroplast and its effect on physiology and gene expression.

Abstract:

Photosynthetic CO_2 fixation is catalyzed by Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO), the most abundant enzyme on Earth. It's high abundance in plant chloroplasts is necessary due to its very low activity and specificity for CO_2. Attempts to improve the activity or specificity of RuBisCO have yielded little progress so far. We have focused on engineering a RuBisCO independent CO_2 fixation cycle into the chloroplast of *Camelina sativa* to increase overall CO_2 assimilation. This synthetic CO_2 fixation cycle is based on enzymes from autotropic bacteria utilizing a reverse TriCarboxylic Acid (TCA) cycle to fix CO_2. The minimal condensed reverse TCA cycle (crTCA) consists of 5 bacterial enzymes that generate glyoxylate from succinate and CO_2/bicarbonate.

We have shown that this engineered crTCA cycle can assimilate CO_2 *in vitro*. After codon-optimization, we were able to show that these enzymes can be expressed in plants. The genes were transformed into the nucleus as fusions containing chloroplast targeting sequences. Chloroplast-localized crTCA enzymes showed activity after purification.

In this study, stable, chloroplast-localized expression of the crTCA cycle in *Camelina sativa* is used to assess changes in photosynthetic parameters. Transgenic crTCA lines have increases in CO_2 assimilation rates under elevated CO_2 levels, greater efficiency in electron usage, and

differences in morphology compared to WT plants. To identify mechanisms beyond the changes CO_2 fixation, we carried out comparative transcriptome analysis from leaf material of transgenic camelina plants expressing the full or partial crTCA cycles with null segregant and empty vector lines. Using differential gene expression analysis, we were able distinguish distinct patterns between the different genotypes. Network analysis identified correlations between the expression of individual crTCA enzymes with changes in specific *Camelina* gene clusters.

While at least parts of the crTCA cycle are apparently functioning in assimilating CO_2, one of the major hurdles is the high abundance of RuBisCO, that competes with the comparatively lower abundance of the crTCA cycle enzymes. We are currently evaluating the full potential of the crTCA cycle in vivo by reducing the endogenous RuBisCO protein using an antisense approach.

References

1. **Bar-Even A., et al.** 2010. Design and analysis of synthetic carbon fixation pathways. *PNAS.* DOI: 10.1073/pnas.0907176107

Acknowledgement: This research is funded by the Department of Energy (ARPAe AR-0000207 & BER DE-SC0018269).

Computational Front End for Creating Universal Phage-based Vectors for Bacteria

Catherine M. Mageeney[1] (cmmagee@sandia.gov), Kelly P. Williams[1], Joseph Schoeniger[1]

[1]Systems Biology, Sandia National Laboratories, Livermore, CA

Project Goals:

The goal for the Intrinsic Control for Genome and Transcriptome Editing in Communities project is to create and deliver technologies to detect the genetic and biochemical state of bacteria. The ability to engineer genomes of isolatable and non-isolatable species is necessary for the growth of the bioeconomy. However, engineering organisms requires addressing key biosafety, biocontainment, and biosecurity concerns. We will create modular technologies that can sense and control altered states of microorganisms in a community setting. We also have developed software to identify numerous bacteriophages - viruses infecting bacteria - which will be used to deliver the modular technologies to any bacterial species within a community, even non-isolable members. In the long-term, we expect our work will create a new foundation for characterization and control of microbial communities providing the scientific foundation for secure bioeconomy.

Abstract

Bacteriophages (phages) are extremely diverse and are hypothesized to infect every group of bacteria. Temperate phages - viruses with the capability to integrate into the bacteria genome - are natural engineers of bacterial populations. The standard approach to collecting phages is to painstakingly isolate them from environmental samples and they are typically not host adapted. We instead exploit the large number of prophages integrated within bacterial genomes and will use these phages as vectors for engineering any microbe, whether isolatable or not.

We have developed two software Islander (1) and TIGER (2) software, that detects genomic islands (GIs) in any bacterial or archaeal sequence. Islander identifies GIs integrated within tRNA genes and containing a tyrosine integrase. This based on the tRNA signature after integration of a GI, one full tRNA and one fragment of the same tRNA. Islander output revealed that integrase genes are generally located at one end or the other of the island. TIGER uses this knowledge to find GIs in any genomic site, containing serine or tyrosine integrases. TIGER uses a ping-pong BLAST search from the mid-point of the integrase to identify the precise GI end coordinates. Our unique software is extensive and precisely maps the mobilome, which provides information for improving biocontainment of heterologous sequences by enabling avoidance of capture of these sequences by intrinsic mobility mechanism.

TIGER and Islander coupled together have the capability to find large numbers of GIs within bacterial or archaeal genomes with the precise end points known. In a pilot study of 2168 genomes, we identified 6415 GIs(2). We are currently in the process of scaling this up to 288,000 genomes available from NCBI. To create a computational front end for phage engineering, TIGER contains a module to type GIs as prophages. This module is based on gene content within the GI, and we found 2418 prophages in our study of 2168 genomes (2), revealing 37.8% of GIs are prophages, about 1 prophage per genome. When we scale this number to the

number of sequenced bacterial genomes, we expect to build a database of more than 288,000 new prophages.

We have previously created an engineering platform to convert our prophages into lytic phages with the capability of killing their target bacteria for therapeutic purposes (3), and proved the principle by developing a phage cocktail targeting the pathogen *Pseudomonas aeruginosa* PAO1. From our prophage database created by TIGER and Islander, we examine those from the target genus, especially those that are close phylogenetic relatives of our target strains. We verify these phages are active by inducing prophages excision from the bacterial genome using mitomycin C, followed by deep sequencing, and non-canonical junction analysis using Juxtaposer (4). This reveals which phages have excised from their bacterial genome, as well as phages that are actively replicating. We identified five active prophages from two *P. aeruginosa* strains. We created synthetic phage genomes using long PCR with overlapping joints, which allowed for the removal of the integrase gene, followed by Gibson Assembly to reassemble the circular phage genome. We rebooted the phages through electro-transformation and chemical transformation. These phages were then tested in liquid cultures and the *Galleria mellonela* -waxworm- model system for therapeutic killing ability.

We are continuing to build the computation front end phage engineering platform to identify phages of interest for a variety of species, including *Streptomyces*, *Cyanobacteria*, and *Escherichia coli*. We plan to identify active prophages in near relatives of our targets of interest, engineer genomes with cargo for detection of environmental states of the bacteria, and reboot the phages using a variety of methods including electroporation and novel-conjugation based methods.

References

1. Hudson CM, Lau BY, & Williams KP (2015) Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res* 43(Database issue):D48-53.
2. Mageeney CM*, et al.* (2020) New candidates for regulated gene integrity revealed through precise mapping of integrative genetic elements. *Nucleic Acids Res*.
3. Mageeney CM*, et al.* (2020) Computational Basis for On-Demand Production of Diversified Therapeutic Phage Cocktails. *mSystems* 5(4).
4. Schoeniger JS, Hudson CM, Bent ZW, Sinha A, & Williams KP (2016) Experimental single-strain mobilomics reveals events that shape pathogen emergence. *Nucleic Acids Res* 44(14):6830-6839.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Field Work Proposal Number 20023046." 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: SAND2021-1393 A

Title: Developing mutant resources for pennycress to domesticate and improve crop resilience

Ratan Chopra[1,2*] (rchopra@umn.edu), M. David Marks[2], **John Sedbrook[3]**

[1]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108
[2]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108
[3]School of Biological Sciences, Illinois State University, Normal, IL 61790 USA

Project Goals: First-generation domesticated pennycress varieties have limited genetic variation, which hampers their adaptability and resilience against abiotic and biotic challenges. Therefore, crucial work remains to identify genetic variants conferring stress tolerance and resilience for incorporation into next generation elite pennycress varieties. Future pennycress varieties will also require optimized lifespans for a range of latitudes and cropping systems, and improved root architectures and physiologies to maximize water and nutrient scavenging as well as carbon sequestration. To attain these goals, we are developing mutant resources to screen for resilience traits or variants in candidate genes identified using eco-evolutionary computational genomics.

Abstract: The oilseed species *Thlaspi arvense* (pennycress)—a weed that was only recently removed from the wild—has the potential to provide new sources of food and bioproducts when grown as a winter cover crop without requiring new land. By taking advantage of extensive gene and phenotype knowledge in the related plant *Arabidopsis*[1], progress has been made to rapidly identify and stack crucial traits needed to domesticate the plant, allowing it to fit within current crop cycles and to have improved seed harvestability and nutritional content[2]. The first-generation domestication varieties have limited genetic variation, which hampers their adaptability and resilience against abiotic and biotic challenges. Much work has been performed using *Arabidopsis* as a model to identify gene mutations that may improve resilience. We have used classical mutagenesis and whole genome sequencing to recreate and characterize such mutations in pennycress. To date, we have whole-genome re-sequenced ~500 EMS mutagenized pennycress lines. A total of ~2.7 million mutations were identified and the mutation rate per line was estimated to be an average of 14.09 variants/Mb in the EMS population. Functional annotation of SNPs generated by EMS treatments suggested 82% of the pennycress genes had one or more alleles. Of these, 5,000 genes contained a mutation-introduced stop codon. We are on track to sequence an additional 500 mutants to obtain more than two alleles in each of the predicted genes of pennycress. We anticipate that this dataset will contain a spectrum of weak to strong alleles for genes previously identified in Arabidopsis that are predicted to improve resilience. We will be presenting the progress on the development of this mutant resource.

References

1. Chopra, R. *et al.* Translational genomics using Arabidopsis as a model enables the characterization of pennycress genes through forward and reverse genetics. *Plant J.* (2018). doi:10.1111/tpj.14147
2. Chopra, R. *et al.* Identification and stacking of crucial traits required for the domestication of pennycress. *Nat. Food* **1**, 84–91 (2020).

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 DOE BER DE-SC0021286 and by the USDA-NIFA under award number 2019-69012-29851

Title: Advancing field pennycress as a new oilseed biofuels feedstock that does not require new land commitments – Improving pennycress stand establishment

Ratan Chopra[1], **John Sedbrook[2]**, and M. David Marks[1] (marks004@umn.edu)*

[1]University of Minnesota, St. Paul, MN; Department of Plant and Microbial Biology
[2] Illinois State University, Normal, IL; School of Biological Sciences

Project Goals:

Pennycress (*Thlaspi arvense* L.) is being domesticated to serve as new cash/cover crop for the Midwestern United States. Currently land is left fallow approximately 8 months of the year between the fall harvest and spring sowing of traditional summer crops. During the fallow period the land is susceptible to soil erosion, nutrient leaching, and runoff pollution. Winter pennycress can be grown during the fallow period as a cover crop to provide ecosystem services to protect the land. Importantly, pennycress produces a harvestable oil seed that can be used for the production of biofuels and other products. As a weed, pennycress seeds exhibits a high degree of dormancy that results in a long-lived seed bank. In addition, the latent dormancy reduces stand establishment. The main goal of the work presented in this poster is to identify pennycress mutants that exhibit reduced dormancy and better stand establishment.

Abstract text.

As a member of the Brassicaceae, pennycress is closely related to the model plant Arabidopsis thaliana. Pennycress shares important attributes with Arabidopsis including self-compatibility and a small genome exhibiting a reduced degree of gene duplication[1]. Prior work has shown that there is a mostly one to one correspondence between genes in pennycress and Arabidopsis. Further, it has been shown that the same spectrum of mutants found in Arabidopsis can be created in pennycress[2]. In this report mutations that reduce dormancy are sought. Pennycress seeds need both light and a relatively high-water potential to germinate. Currently, the most common methods for sowing pennycress are through either surface broadcasting or via very shallow drilling. This makes pennycress establishment dependent on fall rains that can be rare in the Midwestern United States. The seeds of most crop plants can be drilled deep enough to make contact with soil containing a sufficiently high-water potential to promote uniform germination. Therefore, pennycress mutants with reduced requirements for light and high water potential will potentially allow pennycress seeds to be more deeply drilled in order to improve stand establishment. In addition, such traits should reduce the formation of long-lived seed banks.

Dormancy is a well-studied trait in Arabidopsis[3]. Numerous Arabidopsis mutants have been identified that reduce dormancy. The goal of this work was to identify similar mutants in pennycress. In theory these mutants can be identified via classical mutagenesis or produced through gene editing. The work described in this report combines classical mutagenesis with modern genomics to enable rapid reverse genetic approaches to identify recessive mutations that promote pennycress domestication. This approach relies on conducting whole genome sequencing of a population of mutagenized individuals. The goal is to produce a database called a mutant gene index (MGI) that contains a list of multiple mutations in every gene in pennycress. The MGI can be used to identify individuals in the mutagenized population that contains

mutations that confer useful traits. Thus, to identify pennycress mutants with reduced seed dormancy, prior work on dormancy in Arabidopsis has been used to identify gene targets. Data presented in this report is on the characterization of two pennycress mutants selected as likely targets to show reduce seed dormancy.

Three key traits are needed to improve field establishment. These include reduced dormancy to speed germination under optimal conditions, the ability to germinate in the dark to allow seeds to be more deeply drilled, and the ability to germinate at a lower water potential. Given these needs, three types of in lab experiments were designed to characterize two chosen candidate mutants named mutant A and mutant B. In summary, it was found that both mutants germinated faster than the parental controls under optimal conditions. In addition, mutant B germinated much better than either mutant A or the control in the dark. Finally, under reduced water potential mutant A germinated better than either mutant B or the control. Limited field testing has shown that mutant B, the best dark germinator, does established better than either mutant A or the control when drilled up to an inch into the soil. In the future additional field tests will be conducted and these studies will include the double mutant.

References

1. Dorn, K. M., Fankhauser, J. D., Wyse, D. L. & Marks, M. D. A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. *DNA Res.* **22**, 121–131 (2015).
2. Chopra, R. *et al.* Translational genomics using Arabidopsis as a model enables the characterization of pennycress genes through forward and reverse genetics. *Plant J.* (2018). doi:10.1111/tpj.14147
3. Yan, A. & Chen, Z. The Control of Seed Dormancy and Germination by Temperature, Light and Nitrate. *Bot. Rev.* **86**, 39–75 (2020).

This research was supported by the Plant Feedstock Genomics for Bioenergy: A Joint Research Funding Opportunity. Funds provided by National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2018-67009-27374.

Interrogating pennycress natural and induced variation to improve abiotic stress tolerance and oilseed bioenergy crop resilience

Moises Exposito-Alonso,[1] Sue Rhee,[1] M. David Marks,[2] Dmitri Nusinow,[3] Chris Topp,[3] Daniel Jacobson,[4] Pubudu Handakumbura,[5] Winthrop Phippen,[6] Karen Sanguinet,[7] Andrea Gschwend,[8] and **John Sedbrook**[9*] (jcsedbr@ilstu.edu)

[1]Carnegie Institution for Science, Stanford, CA; [2]University of Minnesota, St. Paul, MN; [3]Donald Danforth Plant Science Center, St. Louis, MO; [4]DOE Oak Ridge National Laboratory, Oak Ridge, TN; [5]DOE Pacific Northwest National Laboratory EMSL, Richland, WA; [6]Western Illinois University, Macomb, IL; [7]Washington State University, Pullman, WA; [8]Ohio State University, Columbus, OH; and [9]Illinois State University, Normal, IL

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: 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 that meets the U.S. Renewable Fuels Standard. Academic, governmental, and industrial stakeholders are working closely to commercialize domesticated pennycress varieties by 2022 that can yield over 1680 kg ha^{-1} (1500 lb ac^{-1}) of seeds producing 600 liters ha^{-1} (65 gal ac^{-1}) of oil annually. However, these first-generation varieties have limited genetic variation, which hampers their adaptability to and resilience against abiotic and biotic challenges. Therefore, crucial work remains to identify genetic variants conferring stress tolerance and resilience for incorporation into next generation elite pennycress varieties. Future pennycress varieties will also require optimized lifespans for a range of latitudes and cropping systems, and improved root architectures and physiologies to maximize water and nutrient scavenging as well as carbon sequestration. To attain these goals, interdisciplinary teams employing eco-evolutionary computational genomics are identifying key genetic variants that have enabled pennycress to locally adapt and colonize all temperate regions of the world. Knowledge of metabolic and cellular networks derived from first principles are being generated to guide precise laboratory efforts aimed at identifying superior abiotic stress resilience gene variants. Candidate gene variants are being identified by screening large sequence-indexed mutant populations and by employing cutting-edge CRISPR genome editing strategies. Advanced high-throughput

phenotyping and in-field analytical methods will be employed to validate pennycress lines exhibiting superior abiotic stress tolerance resulting in consistent and higher seed yields. This project will deliver speed-breeding methods to facilitate the introduction of superior allelic variants into a wide range of commercial pennycress varieties to precisely adapt them to the desired local environments. Many of the 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.

Oxygen Generation via Water Splitting by a Novel Biogenic Metal Binding Compound

Alan A. DiSpirito[1] (aland@iastate.edu), Philip Dershwitz[1], Nathan L. Bandow[1], Marcus T. McEllistrem[2], Rafael A. Heinze[2], Matheus Fonseca[2], Joshua C. Ledesma[1], Jacob R. Jennett[1], Ana M. DiSpirito[1], Navjot S. Athwal[1], Mark S. Hargrove[1], Thomas A. Bobik[1], Hans Zischka[3] and **Jeremy D. Semrau[4]**

[1]Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology. Iowa State University, Ames, IA; [2]Department of Chemistry, University of Wisconsin-Eau Claire, WI; [3]Institute of Molecular Toxicology and Pharmacology, Helmholtz Center Munich, German Research Center and Environmental Health, Ingolsteadter Landstrasse, Germany; [4]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI

https://emmb.engin.umich.edu

Project Goals: The overall goal of this project is to determine how significant microbial competition for copper is *in situ*, particularly how such competition affects net methane and nitrous oxide emissions. By better understanding how microbes compete for trace nutrients (i.e., copper) at a molecular level, we can scale such competition to ecosystem functioning, i.e., how microbial competition can be modeled to predict emerging microbial community composition and activity. Doing so will enable us to better understand how interactions within microbial communities control net greenhouse gas emissions, and allow us to determine to what extent copper uptake systems may be public vs. private goods.

Abstract
Methanobactins (MBs) are small (<1,300 Da) post-translationally modified copper-binding peptides and represent the extracellular component of a copper acquisition system in some methanotrophs. Interestingly, MBs can bind a range of metals, with some reduced after binding, e.g., Cu^{2+} reduced to Cu^{+} and Au^{3+} to Au^{0}. Other metals, however, are bound but not reduced, e.g., K^{+}. The source of electrons for selective metal reduction has been speculated to be water but never empirically shown. Here, using $H_2{}^{18}O$, we show that when MB from *Methylocystis sp*. strain SB2 (MB-SB2) and *Methylosinus trichosporium* OB3b (MB-OB3) were incubated in the presence of either Au^{3+}, Cu^{2} , and Ag^{+}, $^{18,18}O_2$ and free protons were released. No $^{18,18}O_2$ production was observed either in presence of MB-SB2 or MB-OB3b alone, gold alone, copper alone, silver alone or when K^{+} or Mo^{2+} was incubated with MB-SB2. The discovery that MB will couple the oxidation of H_2O to metal reduction and the release of O_2 suggests that methanotrophs expressing MB may be able to maintain their activity in hypoxic/anoxic conditions through "self-generation" of dioxygen required for the initial oxidation of methane to methanol. Such an ability may be an important factor in enabling methanotrophs to not only colonize the oxic-anoxic interface where methane concentrations are highest, but also tolerate significant temporal fluctuations of this interface. Given that genomic surveys often show evidence of aerobic methanotrophs within anoxic zones, the ability to express MB (and thereby generate dioxygen) may be an important parameter in facilitating their ability to remove methane, a potent greenhouse gas, before it enters the atmosphere.

This work was supported by the Department of Energy, Award # DE-SC0020174.

Methanotrophs Produce Diverse Chalkophores to Compete for Copper

Christina S. Kang-Yun[1*] (csrkang@umich.edu), Alan A. DiSpirito[2], and **Jeremy D. Semrau**[1]

[1]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan
[2]Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa

https://emmb.engin.umich.edu

Project Goals: The overall goal of this project is to determine how significant microbial competition for copper is *in situ*, particularly how such competition affects net methane and nitrous oxide emissions. By better understanding how microbes compete for trace nutrients (i.e., copper) at a molecular level, we can scale such competition to ecosystem functioning, i.e., how microbial competition can be modeled to predict emerging microbial community composition and activity.

Abstract

Methanotrophs, microorganisms that use methane as their sole source of carbon and electron, play a critical role in the biogeochemical cycling of carbon. These intriguing microbes consume substantial amounts of methane and thus serve as a significant sink for methane in the environment. Expression and activity of alternative forms of methane monooxygenase (MMO), which is responsible for the initial conversion of methane to methanol, is controlled by copper availability or the canonical "copper-switch".

Methanotrophs have multiple mechanisms of copper uptake. Some methanotrophs of the *Methylocystaceae* family within the Alphaproteobacteria express a copper-binding compound or chalkophore – methanobactin (MB). MB is a small ribosomally synthesized, post-translationally modified polypeptide, that is characterized by two heterocyclic rings and associated enethiol groups that bind copper with extremely high affinity. Methanotrophs belonging to the Gammaproteobacteria, such as those of the *Methylococcaceae* family, rely on an outer membrane protein (MopE or CorA) for copper uptake, as well as some *Methylococcaceae* secreting a chalkophore similar to MB, but with much weaker affinity for copper.

Given the importance of copper in methanotrophy, it is highly likely that methanotrophs actively compete for copper, which raises an intriguing question. Since MB is secreted into the environment and copper-MB complexes are then taken up, can these complexes be "stolen" by other microbes, similar to siderophore theft observed between certain microbes?

Herein we show that both "cheating" and "competing" exist between methanotrophs for copper. Specifically, *Methylomicrobium album* BG8, lacking genes for MB biosynthesis, carries a gene for TonB-dependent transporter required for MB uptake (*mbnT*). However, deletion of the putative *mbnT* does not affect the copper acquisition ability of *M. album* BG8, suggesting this gene is either not involved in MB uptake, or this microbe has an alternative pathway for copper

uptake. In addition, *M. album* BG8 is not starved for copper in the presence of MB, but is in the presence of triethylenetetramine (TRIEN), an abiotic copper chelator. Specifically, *M. album* BG8 produces a MB-like substance when grown in the presence of both copper and TRIEN. Preliminary analyses of this potentially novel chalkophore indicated a size of approximately 1125 Da and characteristic UV-Vis spectra that are responsive to varying concentrations of copper. These results indicate that MB may be considered to be a "public good," and that there may be yet to be characterized chalkophores that can affect environmental copper availability as well as microbial community composition and activity.

This work was supported by the Department of Energy, Award # DE-SC0020174.

Integrated deep-learning computational approach to proteome annotation

Authors: Mu Gao[1]* (mu.gao@gatech.edu), Jianlin Cheng[2], Jerry Parks[3], Dwayne Elias[3], Ada Sedova[3], and **Jeffrey Skolnick**[1]

Institutions: [1]Georgia Institute of Technology, Atlanta, GA; [2]University of Missouri, Columbia, MO; [3]Oak Ridge National Laboratory, Oak Ridge, TN.

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 infer protein structure information 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: The ability to predict the structure and function of a protein-coding gene from its sequence is a grand challenge in biology. Advances in next generation sequencing technologies have led to an exponential increase in the size of genomic datasets. Genome functional annotation, the assignment of validated molecular functions to the majority of the protein coding genes in a genome, has been challenged by these massive datasets. Experimental methods offer the gold standard for proof-of-function, but even high-throughput experiments are orders of magnitude too slow compared with the speed at which gene-sequencing big-data is generated, causing a new technology bottleneck. With growing computing power and the success of advanced machine-learning analyses, computational methods could help eliminate this bottleneck by accurately inferring protein function and thereby providing experimentally testable hypotheses.

Current computational technologies to infer function are often based on evolutionary inference through sequence comparison with known, annotated protein sequences. However, these methods fail when the sequence similarity is low, e.g., at 30% sequence identity. For many organisms, this may represent a significant portion of the genome. To tackle this important issue, our team is developing a suite of novel high-performance computing (HPC) and deep learning-based computational methods that predict structural information, and then apply it to help predict the function of proteins with low sequence similarity to any known annotated protein. Using deep learning for the prediction of the structure of these low-similarity proteins has recently achieved some dramatic breakthroughs. The inclusion of structure-related information from these predictions can help to fill in the knowledge gaps for this proteomic “dark matter.”

There are a number of ways that deep-learning inspired structural inference can help infer function. One such method is SAdLSA, which is trained to conduct sequence alignment from deep-learning protein structural alignments [1]. The implicit structural information used by deep learning reveals structural similarities not apparent from standard sequence comparison. SAdLSA has been deployed on the Summit supercomputer and can use hundreds of GPUs to search for hidden matches to experimentally deter crystal structures. We applied SAdLSA to 559 uncharacterized protein sequences in *Desulfovibrio vulgaris*, a model organism for sulfur-reducing bacteria. Scanning a large sequence library of 83,000 sequences, the pilot runs of SAdLSA on *Desulfovibrio vulgaris* found some significant hits for over 25% of these uncharacterized sequences. Preliminary analysis on just a few sequences has already revealed interesting predictions. For example, one bacterial protein's top structural match points to human PHPT1, a eukaryotic phosphohistidine phosphatase with no known prokaryotic counterpart. Moreover, the sequence alignment prediction is corroborated by the MULTICOM2 method, discussed below.

Predicting the full three-dimensional structure of a protein can provide a wealth of essential information derived from analyzing this structure using it to model binding interactions. The multi-task deep learning method DeepDist [2] is based on residual convolutional neural networks and predicts inter-residue distances from protein sequences via both regression and multi-classification. DeepDist was used to predict inter-residue distances for template-free (*ab initio*) protein structure prediction in the latest version of the comprehensive protein structure prediction system, MULTICOM2, which was ranked 7th out of 146 predictors in the tertiary structure prediction and 3rd out of 136 predictors in the inter-domain structure prediction in the CASP14 experiment. We are currently deploying several of MULTICOM2's tasks on Summit in order to train on larger datasets for even more accurate models. After a structure is predicted, fold analysis, and binding pocket analysis is performed, and molecular interactions can be predicted and modeled with simulations. This provides more information about the protein's function, and its placement in metabolic and interaction networks. Together the methods and applications we are developing on leadership computing resources will help provide a new generation of solutions to help break the genome functional annotation bottleneck.

References/Publications

1. Gao, M. and J. Skolnick, *A novel sequence alignment algorithm based on deep learning of the protein folding code.* Bioinformatics, 2020.
2. Wu, T., Z. Guo, J. Hou, and J. Cheng, *DeepDist: real-value inter-residue distance prediction with deep residual convolutional network.* bioRxiv, 2020: p. 2020.03.17.995910.

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021303.

Title: Bayesian Conditional Auto-Regressive LASSO Models to Learn Sparse Networks with Predictors

Authors: Yunyi Shen[1]* (yshen99@wisc.edu) and **Claudia Solis-Lemus**[1]

Institutions: [1]University of Wisconsin-Madison, Madison, WI

Website: https://github.com/YunyiShen/CAR-LASSO

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

1. Develop novel statistical theory on a Bayesian regression framework that can predict changes in microbial compositions due to experimental or environmental factors. Our novel model will simultaneously account for the microbial interactions and predictor effects
2. Produce an efficient sampling scheme to estimate the posterior distribution of model parameters that is scalable enough to meet the ever-growing needs of high-dimensional soil microbiome data
3. Implement a publicly available open-source R package that would allow scientists to utilize our novel model in their own soil microbiome data to make inferences about the specific predictors (experimental or environmental) that are playing a role in shaping the microbial compositions in soil

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. In particular, many critical soil processes such as mineral weathering, and soil cycling of mineral-sorbed organic matter are governed by mineral-associated microbes. Understanding the composition of microbial communities and what environmental factors play a role in shaping this composition is crucial to comprehend soil biological processes and to predict microbial responses to environmental changes.

However, the inter-connectivity of microbes-environment is still not fully understood. One of the reasons for this gap in knowledge is the lack of statistical tools to infer connections among microbial communities while simultaneously accounting for predictors in a unified framework. Indeed, there is a need for statistical models that can decode microbes' reaction to the environment and interactions among microbes simultaneously. The model should have the ability to correctly incorporate prior knowledge from controlled experiments and its implementation should be scalable enough to meet the ever-growing needs of the high-dimensional soil microbiome big data.

We introduce a novel Bayesian conditional auto-regressive (CAR) LASSO model to infer a sparse network structure with nodes for responses and for predictors. Directed edges between a predictor and a response represent conditional links, and undirected edges among responses represent correlations. Specifically, our model estimates a microbial network that represents the dependence structure of a multivariate response (e.g. abundances of microbes) while simultaneously estimating the effect of a set of predictors that influence the network (e.g. diet, weather, experimental treatments). In addition, our method produces a sparse interpretable graph via LASSO penalization.

We also propose an adaptive extension of the CAR LASSO model so that different shrinkage can be applied to different edges which allows the incorporation of edge-specific prior knowledge. Indeed, the conditional representation of our model coefficients and adaptivity allow us to adequately encode prior knowledge obtained by specific experimental interventions and agrees with the experimenter's intuition on average behavior of nodes under experiments. In addition, our model is able to equally handle small and big data and is computationally inexpensive through an efficient Gibbs sampling algorithm. With hierarchical structure, we extend the model to binary, counting and compositional responses by adding an appropriate sampling distribution to the core Normal model. Finally, we apply our model to a public soil microbial composition dataset. Researchers isolated soil aggregates from three land management systems in central Iowa to test if the aggregate-level microbial responses are related to plant community and management practices. Our work estimates the soil microbiota network and identifies the most important connections between microbes and the most important microbes-environment links.

References/Publications

1. Shen, Yunyi, and Claudia Solis-Lemus. "Bayesian Conditional Auto-Regressive LASSO Models to Learn Sparse Networks with Predictors." *preprint arXiv:2012.08397* (2020).

Funding statement: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021016.

High quality anaerobic fungal genome assemblies and annotation for study and optimization of lignocellulose conversion

Casey Hooker[1,2*] (hookerc@purdue.edu), Ethan Hillman[1,3], Javier Muñoz Briones[1,3], and **Kevin Solomon[1,2,3†]**

[1] Agricultural & Biological Engineering, Purdue University, West Lafayette, IN; [2]Laboratory of Renewable Resources Engineering (LORRE), Purdue University, West Lafayette, IN; [3] Purdue University Interdisciplinary Life Sciences (PULSe) Program, Purdue University, West Lafayette, IN

[†]Present Address: Chemical & Biomolecular Engineering, University of Delaware, Newark, DE

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]. However, the specific role of individual enzymes in maintaining hydrolytic efficiency remains unknown due to a lack of genetic tools that facilitate testing of gene function in its natural context. Thus, there is a critical need to create methods that manipulate CAZyme expression and rapidly interrogate gene function in anaerobic fungi to identify targets that will advance biofuel and bioproducts production.

In this project, we study three novel specimens of anaerobic fungi. Species characterization confirm that they are unique and exhibit high enzymatic activity against a range of untreated lignocellulolytic substrates regardless of lignin composition (e.g. poplar, sorghum, alfalfa, corn stover)[6]. Anaerobic fungi tailor the secretion of CAZymes to adapt to differences in substrate composition and achieve consistently high-levels of synergistic activity. To better understand this response, we need to develop tools for genetic engineering of anaerobic fungi.

As a first step towards tool development, we assembled new fungal genomes to identify key regulatory sequences for future parts. Previous genomes were highly fragmented in to as many as 30,000 scaffolds. That is, many potential parts identified from these datasets could be non-functional due to truncations in the sequence due to poor assembly. However, 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 family of organisms with chromosomal resolution. For example, our assembly of *Piromyces* sp. UH3-1 incorporated 99% of the genome into 12 chromosomes, a 94% reduction in fragmentation relative to the historically best anaerobic fungal genome assemblies.

Annotation of these rich datasets for parts discovery is underway. We have developed a bioinformatic pipeline to identify conserved AT-rich promoters that may drive CAZyme expression and are currently synthesizing more than a dozen at the JGI for evaluation along with reporter proteins and other putative parts for gene expression. In parallel, we have begun to identify and test other regulatory sequences such as centromere binding sequences (CEN) and autonomous replicating sequences (ARS) needed to create self-replicating plasmids that may be used to deliver new genes and control cell phenotypes. We are also acquiring transcriptomes and proteomes of our fungal isolates across various substrates in partnership with the DOE-JGI and EMSL. Using these resources, we are identifying parts and environmental conditions that regulate them to develop novel tools for stable gene expression.

In addition to developing specific parts, we have also studied the genetics of anaerobic fungi to inform strategies for manipulation of CAZymes and other genes. Genome annotation and nuclei labelling suggest that all our isolates are haploid (only one chromosomal copy per species), which streamlines future genetic engineering testing and development. Genome architecture analysis has also identified several features such as long-terminal repeat sequences and horizontal gene transfer events, which may be mutational hotspots on the genome to avoid in future strain engineering for stable constructs.

In summary, the ongoing work has 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)

This project is supported by the Office of Biological and Environmental Research through the DOE Office of Science (Early Career Program) Contract No DE-SC0020117. 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).

Characterizing Algal Exo-Metabolites and Their Impacts on Algae and Associated Bacteria

Vanessa Brisson[1]* (brisson2@llnl.gov), Xavier Mayali,[1] Ty Samo,[1] **Trent Northen**,[2,3] and **Rhona Stuart**[1]

[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; and [3]The DOE Joint Genome Institute, Berkeley, CA.

https://bio-sfa.llnl.gov

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy crop production through a community systems biology understanding of microbial consortia that are closely associated with bioenergy-relevant crops. We focus on host-microbial interactions in algal ponds and perennial grasses, with the goal of understanding and predicting the system-scale consequences of these interactions for biomass productivity and robustness, the balance of resources, and the functionality of surrounding microbial communities. Our approach integrates 'omics measurements with quantitative isotope tracing, characterization of metabolites and biophysical factors, genome-enabled metabolic modeling, and trait-based representations of complex multi-trophic biological communities, to characterize the microscale impacts of single cells on system scale processes.

Complex metabolic interactions between microalgae and their associated microbial communities influence algal physiology and growth, with implications for both bioenergy production and global biogeochemical cycling. The suite of extracellular metabolites produced and exchanged between algae and bacteria are potentially important mediators of these interactions. Here, we present research taking the first steps toward a more global profile of algal exo-metabolites and their role in algal-bacterial interactions. We set out to determine: (a) whether phylogenetically distinct algal taxa had distinct exo-metabolomes, (b) how specific algal exo-metabolites affect algal growth, (c) whether algal exo-metabolomes have predictable dynamics and (d) the effect of specific exo-metabolites on phycosphere-associated bacteria.

We profiled the extracellular and cell associated metabolites of four phylogenetically diverse algal strains using liquid chromatography tandem mass spectrometry (LC-MS/MS). We characterized metabolites from axenic cultures of the freshwater green alga *Chlamydomonas reinhardtii*, *Desmodesmus* sp. strain C406, which can grow in either fresh or salt water, the saltwater algal *Microchloropsis salina*, and the diatom *Phaeodactylum tricornutum*. The metabolite profiles of these algae differed significantly, reflecting their phylogenetic diversity. For each alga, the extracellular and cell associated metabolite profiles were also distinct, indicating that the extracellular metabolites were not primarily the product of cell lysis, but instead were more likely excreted or produced extracellularly. We used the global natural products social networking (GNPS) pipeline to generate putative identifications for some

metabolites. Putatively identified metabolites included known phytohormones and plant growth promoters, as well as signaling molecules. We hypothesized that some of these metabolites might be involved in intra-algal signaling, and thus might affect algal physiology and growth. We investigated the impacts of 10 metabolites enriched in the extracellular pools of one or more algal strains on each of the axenic algal cultures. We demonstrated that some metabolites significantly increased total chlorophyl in some algal cultures, while other metabolites decreased chlorophyll.

We sought to further characterize the temporal dynamics of the exo-metabolome of the model diatom *P. tricornutum* through its growth cycle. Using a targeted LC-MS/MS analysis, we identified 60 metabolites, including phytohormones, several B-vitamins and B-vitamin derivatives, and a suite of aromatic organic acids. We found that different sets of metabolites accumulated in the extracellular metabolite pool during the lag, early log, late log, and stationary growth phases, suggesting that there is growth phase dependent regulation of the exo-metabolome. Based on our analysis, we selected a set of 15 *P. tricornutum* exo-metabolites for further study to better understand how these metabolites may be involved in algal-bacterial interactions. The selected metabolites had different accumulation patterns across the range of *P. tricornutum* growth phases. We investigated the ability of 12 phycosphere-associated bacterial isolates to grow in a defined saltwater medium with each of these metabolites provided individually as the carbon source. We found that several of the metabolites could support growth of some bacterial isolates. The patterns of substrate use differed between bacterial isolates, suggesting that algal exo-metabolites could modulate microbial communities by selectively supporting the growth of different bacterial strains in the surrounding communities. Together, these results reveal the broad diversity of algal exo-metabolites, which are likely to have distinct functions in governing algal microbiomes and algal ecophysiology.

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. LLNL-ABS-818816.

Integrated Experimentation and Trait-based Model Exploration of Algal-Bacterial Interaction Mechanisms.

Authors: Yiwei Cheng,[1] Ty J. Samo,[2] Peter K. Weber,[2] Nicholas J. Bouskill,[1] Ulas Karaoz[1], Noele Norris,[1] Rhona K. Stuart,[2] Xavier Mayali,[2] **Eoin Brodie (elbrodie@lbl.gov)[1,3*]**

Institutions: [1]Lawrence Berkeley National Laboratory, Climate and Ecosystem Sciences Division, Berkeley, CA, USA; [2] Lawrence Livermore National Laboratory, Nuclear and Chemical Sciences Division, Livermore, CA, USA**;** [3] Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA.

Website: https://bio-sfa.llnl.gov/

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy crop production through a community systems biology understanding of microbial consortia that are closely associated with bioenergy-relevant crops. We focus on host-microbial interactions in algal ponds and perennial grasses, with the goal of understanding and predicting the system-scale consequences of these interactions for biomass productivity and robustness, the balance of resources, and the functionality of surrounding microbial communities. Our approach integrates 'omics measurements with quantitative isotope tracing, characterization of metabolites and biophysical factors, genome-enabled metabolic modeling, and trait-based representations of complex multi-trophic biological communities, to characterize the microscale impacts of single cells on system scale processes.

Microalgal interactions with heterotrophic bacteria have substantial implications for carbon cycling in natural and engineered aquatic ecosystems, including beneficial effects on net primary productivity, but the relative importance of different mutualistic mechanisms is rarely explored from a functional trait perspective. In this work, *Phaeodactylum tricornutum*, a biofuel relevant microalgal species, was cultured with and without a heterotrophic bacterium of the *Marinobacter* genus, under cyclical light conditions and two nutrient regimes. *P. tricornutum* grown with *Marinobacter* exhibited a 10% increase in growth rate over the axenic culture. Carbon exchange and productivity increases were supported by a 19% enhancement in carbon (C) fixation by individual *P. tricornutum* cells as shown by nanoscale stable isotope probing (nanoSIP). The proposed beneficial roles of the heterotrophic bacteria include: (1) relief of algal CO_2 limitation through localized bacterial respiration of algal exudates (a function of biomass and growth yield); (2) oxidative stress alleviation via usage of O_2 as a terminal electron acceptor in bacterial respiration. These processes are coupled in reality, and to explore the relative contributions of these two mechanisms to enhanced algal growth, we have developed a dynamic energy budget (DEB) model of interactions between these phototrophic and heterotrophic microorganisms. Explicit representation of the competitive dynamics between DIC and O_2 in RuBisCO permitted direct comparison of the relative impact of the two mechanisms. Benchmark simulation results

show that the two mechanisms can account for most but not all of the enhanced carbon fixation, suggesting the possible importance of other mechanisms. The simulations specifically evaluated algal C exudation rates, bacterial growth yield (via maintenance costs), in addition to the affinities of algal RuBisCo for CO_2 versus O_2. In doing so, the model revealed that algae with higher carbon exudation rates and bacteria with higher maintenance costs (and therefore lower carbon use efficiency) combine to result in higher algal growth yield. Therefore, factors that regulate algal retention and excretion of carbon, and microbial metabolic and biophysical traits that regulate carbon-use efficiency, are likely to influence the dynamics of algal-bacterial associations and the fitness of both parties. This provides a series of hypotheses that are being explored through integrated computational trait inference, experimentation and dynamic energy budget modeling.

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 Biofuels SFA, FWP SCW1039, LLNL-ABS-818820. Part of this work was performed at Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231.

Context Dependent Mycorrhizal Resource Exchange in Bioenergy Cropping Systems

Authors: Rachel Hestrin[1*] (hestrin1@llnl.gov), Megan Kan[1], Jeff Kimbrel[1], Jessica Wollard[1], Christina Ramon[1], Marissa Lafler[1], Prasun Ray[2], Rina Estera-Molina[3], Christina Fossum[3], Aaron Chew[3], Tasnim Ahmed[3], Ilexis Chu-Jacoby[3], David Sanchez[3], Melanie Rodriguez Fuentes[3], Cynthia-Jeanette Mancilla[3], Emily Kline[3], Laura Adame[3], Courtney Swink[1], Jose Liquet y Gonzalez[1], Jamie Brown[4], Bruce Hungate[4], Amrita Bhattacharyya[5], Steven Blazewicz[1], Erik Oerter[1], **Kelly Craven[2], Mary Firestone[3], Peter Weber[1], Rhona Stuart[1], Erin Nuccio[1], Jennifer Pett-Ridge[1]**

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Noble Research Institute, Ardmore, OK; [3]University of California, Berkeley, CA; [4]Northern Arizona University, Flagstaff, AZ; [5]Lawrence Berkeley National Laboratory, Berkeley, CA

https://bio-sfa.llnl.gov/

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy crop production through a community systems biology understanding of microbial consortia that are closely associated with bioenergy-relevant crops. We focus on host-microbial interactions in algal ponds and perennial grasses, with the goal of understanding and predicting the system-scale consequences of these interactions for biomass productivity and robustness, the balance of resources, and the functionality of surrounding microbial communities. Our approach integrates 'omics measurements with quantitative isotope tracing, characterization of metabolites and biophysical factors, genome-enabled metabolic modeling, and trait-based representations of complex multi-trophic biological communities, to characterize the microscale impacts of single cells on system scale processes.

Mutualistic associations between plants and mycorrhizal fungi can enhance plant productivity, resilience to stress, and carbon (C) allocation belowground. A better understanding of mycorrhizal 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 resource exchange between *Panicum hallii*—a model species closely related to switchgrass—and two mycorrhizal fungi: the arbuscular mycorrhizal (AM) fungus *Rhizophagus irregularis* and the sebacinoid mycorrhizal fungus *Serendipita bescii*. Both fungi have been found growing in association with a wide range of plant species, including switchgrass and several other bioenergy crops. Due to differences in their genomic repertoires, we hypothesize that each fungus confers plant benefits and ecosystem services through unique mechanisms.

We grew *P. hallii* with and without *R. irregularis* and *S. bescii* in microcosms containing 'live' soil harvested from a marginal Oklahoma pasture. We restricted the soil moisture in half of the microcosms in order to assess plant and mycorrhizal response to water limitation. Additionally, half of the microcosms were grown in a $^{12}CO_2$ atmosphere and half in a $^{13}CO_2$ atmosphere. The microcosms were harvested destructively at 5, 8, and 12 weeks after the onset of $^{13}CO_2$ labeling.

This approach allows us to track plant- and mycorrhizal-derived ^{13}C into other microbial taxa, soil C pools, and C fluxes (CO_2, volatiles, dissolved organic C). We used root and hyphal exclusion chambers to examine biogeochemical fluxes and multipartite microbial interactions in spatially distinct ecological niches.

Although AM and sebacinoid fungi were identified in the native soil microbial community, qPCR analyses show that the *R. irregularis* and *S. bescii* strains used as inoculants were more abundant in roots and soils harvested from inoculated microcosms. Overall, total plant biomass was slightly higher in microcosms inoculated with *R. irregularis* than in uninoculated microcosms and those inoculated with *S. bescii*. The relationship between mycorrhizal inoculation and total plant biomass diminished over time under water-limited conditions, but not under water-replete conditions. Additionally, the ratio of shoot:root biomass was higher in plants grown under water-replete conditions—particularly those inoculated with *S. bescii*. Isotopic analyses and NanoSIMS imaging show that plants allocated a substantial quantity of C to their mycorrhizal partners and other C pools. After 12 weeks of growth in a $^{13}CO_2$ atmosphere, recently photosynthesized ^{13}C accounted for up to 10% of rhizosphere soil C (0.4 mg ^{13}C excess g^{-1} soil) and 0.7% of hyphosphere soil C. Although much of this plant- and mycorrhizal-derived C was retained below ground, some returned to the atmosphere as CO_2. We are using quantitative stable isotope probing (qSIP) to assess taxon-specific microbial response to mycorrhizal presence and soil moisture availability. Our results shed light upon the complex and dynamic nature of plant-microbe interactions and their potential role in sustainable bioenergy crop production.

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-818770

Impacts of physical proximity and metabolite diffusion on algal-bacteria interactions

Hyungseok Kim[1]* (hskimm@mit.edu), Jeffrey A. Kimbrel,[2] Ty Samo,[2] Jessica Wollard,[2] Christina Ramon,[2] **Rhona Stuart**,[2] Peter K. Weber,[2] Cullen R. Buie,[1] and Xavier Mayali[2]

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA; [2]Physical and Life Science Directorate, Lawrence Livermore National Laboratory, Livermore, CA.

https://bio-sfa.llnl.gov/

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy crop production through a community systems biology understanding of microbial consortia that are closely associated with bioenergy-relevant crops. We focus on host-microbial interactions in algal ponds and perennial grasses, with the goal of understanding and predicting the system-scale consequences of these interactions for biomass productivity and robustness, the balance of resources, and the functionality of surrounding microbial communities. Our approach integrates 'omics measurements with quantitative isotope tracing, characterization of metabolites and biophysical factors, genome-enabled metabolic modeling, and trait-based representations of complex multi-trophic biological communities, to characterize the microscale impacts of single cells on system scale processes.

Metabolic interactions between biofuel-producing microalgae and heterotrophic bacteria (the algal microbiome) dominate the elemental cycling of these highly productive phototrophic ecosystems, where bacteria can incorporate up to half of the photosynthetically fixed carbon. These metabolic interactions occur at the scale of single microbial cells, but most methods to investigate these phenomena are not optimized for measurements at this small scale, making it challenging to capture interactions mediated by diffusion and/or cell attachment. To address this, we have taken two approaches. First, in collaboration with MIT, we have built a co-culture hydrogel microplate for algae and bacteria where the microalgal culture, *Phaeodactylum tricornutum*, can grow to a concentration ~20 times greater than batch conditions, and different bacterial populations can be incubated at different distances from the algal source of organic matter (Fig. 1). Second, using fifteen model bacterial co-cultures growing with *P. tricornutum*, we quantified bacterial impact on algal growth under different nutrient and light conditions, and organic algal-derived C and N incorporation at the single cell level using isotope tracing and NanoSIMS.

Using the co-culture microplate, our experiments quantified how two algal-associated bacteria responded at the population level to the presence of the algae grown at different distances. We also examined the impact of inorganic nutrients and algal organic matter diffusion on bacterial community structure development. These data, combined with a diffusion-based numerical derivation, help us propose a new model of bacterial responses towards algal exudate as a

function of algal growth state. Concurrently, using laboratory batch cultures and NanoSIMS, we found surprising variability in the net C and N fluxes between algae and bacteria, which led us to identify a continuum of metabolic interactions from bacterial strains with high levels of C and N exchange through the metabolism of complex organic matter, to strains with low to undetectable exchange. These data challenge the generally accepted view that algal-associated bacteria require algal-derived organic matter for growth, particularly organic nitrogen. We also provide direct evidence of bacterial incorporation of nitrate, suggesting that algal-associated bacteria could switch from commensal or mutualists to competitive under N limitation, and confirming hypotheses generated by our microplate hydrogel experiments.

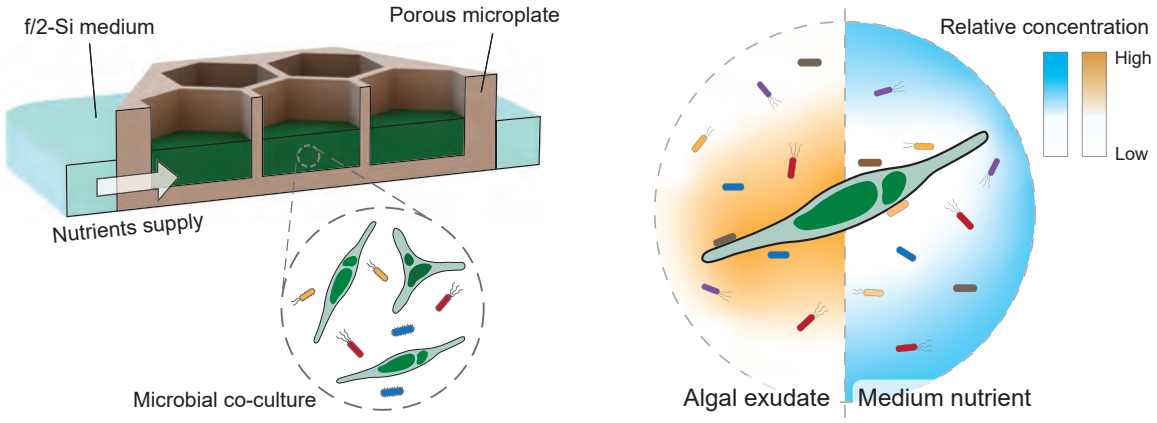


Figure 1. (Left) Schematic of a porous microplate system to co-culture algae and bacteria. (Right) Schematic of a proposed model of nutrient gradients in a phycosphere with bacterial responses.

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. LLNL-ABS-818796.

Characterizing algicidal bacteria and their mechanisms in antagonistic algal-bacterial interactions

Megan M. Morris (morris81@llnl.gov)[1], Jeffrey Kimbrel[1], **Rhona Stuart**[1], **Xavier Mayali**[1]

[1]Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, CA, USA

https://bio-sfa.llnl.gov/

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy crop production through a community systems biology understanding of microbial consortia that are closely associated with bioenergy-relevant crops. We focus on host-microbial interactions in algal ponds and perennial grasses, with the goal of understanding and predicting the system-scale consequences of these interactions for biomass productivity and robustness, the balance of resources, and the functionality of surrounding microbial communities. Our approach integrates 'omics measurements with quantitative isotope tracing, characterization of metabolites and biophysical factors, genome-enabled metabolic modeling, and trait-based representations of complex multitrophic biological communities, to characterize the microscale impacts of single cells on system scale processes.

Heterotrophic bacteria play critical biogeochemical roles in surface aquatic ecosystems where they interact with photosynthetic microalgae via remineralization of algal-derived organic matter. The mutualistic nutrient exchange between algae and associated bacteria can shift towards commensalism, for example when opportunistic bacteria directly exploit algal resources, attacking and lysing algae and benefitting from lysed cellular nutrients. There are a multitude of factors that contribute to the strategy, mechanism, and activity of antagonistic bacteria; however, these are not fully understood and may vary widely depending on taxonomic identity of each partner and the ecological context in which the two partners interact. It is critical that we better understand algal-bacterial antagonism, as bacterial attack on algal populations may have drastic implications for nutrient flux in both natural ecosystems and engineered biofuel aquacultures. Using genome-resolved metagenomics, amplicon sequencing, advanced microscopy, and growth assays, we aimed to characterize the identity, genomic capacity, and mechanism of a novel bacterium with algicidal effects on the model photosynthetic diatom *Phaeodactylum tricornutum*.

The putative algicidal bacterium of this investigation is an unculturable bacterial species growing within a laboratory enrichment of a reduced-diversity bacterial community co-cultured with *P. tricornutum*. We sought to first quantify the effect of the algicidal community on the growth of *P. tricornutum*, and second to identify the algicidal bacterium within this community. In repeated, independent laboratory experiments we showed that the algicidal bacterium attacks *P. tricornutum* during the lag and early exponential growth phases, crashing the diatom population and preventing it from reaching high cell abundances in late log and stationary growth phases. The rate and magnitude of the bacterially-induced diatom population decrease were exacerbated by abiotic stress. We showed that increased temperatures were correlated with a more rapid and irreversible decline in *P. tricornutum*. To identify the bacterial species within the community

enrichment, we used 16S rRNA gene amplicon sequencing. Specifically, we sub-sampled the original bacterial community, and sub-enrichments which varied in their algicidal capacity – high, moderate, to no algicidal effect on *P. tricornutum*. We conducted a repeated time-series sampling of the bacterial community in conjunction with growth of *P. tricornutum* from inoculation (day 0) to mid-stationary phase (day 7). Bacterial sub-enrichments which varied in their algicidal phenotype were distinct in community structure and diversity, and we identified a taxon unique to the sub-enrichments with high algicidal capacity against *P. tricornutum*. Specifically, this taxon, an unclassified Alphaproteobacterium, reached high relative abundances in conjunction with the decline of *P. tricornutum*, comprising of up to 60 % relative abundance within the bacterial community. This suggests that the algicidal Alphaproteobacterium benefits in some capacity from the lysis of *P. tricornutum*. However, the exact strategy and mechanism of the algicidal activity by the unclassified Alphaproteobacterium are still unclear and will be the focus of follow-up experiments. Currently, we are describing the genomic capabilities and metabolic potential of the algicidal Alphaproteobacterium through a metagenome-assembled-genome (MAG) of this organism, as the recovered genome may elucidate the genetic underpinnings of the algicidal phenotype. In addition to characterizing a novel algicidal bacterium, this investigation more broadly contributes to our understanding of the spectrum of algal-bacterial interactions, and this has implications for both environmental and applied sciences.

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. LLNL-ABS-818889

System-level analysis of metabolism of *Kordia algicida* and its interaction with *Phaeodactylum tricornutum*

Ali Navid[1*] (navid1@llnl.gov), Samantha Shen[2], Jeffrey Kimbrel[1], Patrik D'haeseleer[1], Xavier Mayali[1], **Rhona Stuart[1]**

[1]Physics & Life sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Georgetown University, Washington, DC

https://bio-sfa.llnl.gov/

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy crop production through a community systems biology understanding of microbial consortia that are closely associated with bioenergy-relevant crops. We focus on host-microbial interactions in algal ponds and perennial grasses, with the goal of understanding and predicting the system-scale consequences of these interactions for biomass productivity and robustness, the balance of resources, and the functionality of surrounding microbial communities. Our approach integrates 'omics measurements with quantitative isotope tracing, characterization of metabolites and biophysical factors, genome-enabled metabolic modeling, and trait-based representations of complex multi-trophic biological communities, to characterize the microscale impacts of single cells on system scale processes.

Planktonic microalgae play an outsized role in regulating the dynamics of earth's ecosystem. Their activity affects the global oxygen supply, the food chain, global biogeochemical cycling, and climate. Due to their ecological importance, as well as the industrial interest in using algae for production of renewable biofuels, understanding the factors that control algal productivity are crucial for devising strategies for mitigating damaging effects of climate change. Interactions between algae and bacteria are a major factor in the fate of algal populations. This includes parasitic interactions, wherein "algicidal" bacteria can lyse algal cells in order to acquire nutrients. The controls on algicidal bacteria are not well defined, so while they are sometimes detectable at low levels in algal microbiomes, we currently cannot predict what conditions will lead to an algicidal-induced algal population crash. In order to gain a system-level understanding we conducted a number of computational analyses of an algicidal bacterium's interaction with a model algae.

Kordia algicida is a marine bacterium that has exhibited algicidal behavior when co-cultured with a variety of algae[1], including *Phaeodactylum tricornutum*, a long-studied model organism for analysis of phytoplankton ecology and physiology. *P. tricornutum* is also recognized for its potential for biofuel production[2]. In order to gain a better understanding of the predatory algal-bacterial interactions we have examined the metabolism of *K. algicida* and its interaction with *P. tricornutum in silico*.

We studied the metabolic needs of *K. algicida* by developing a human curated genome-scale model of *K. algicida* strain OT-1. We previously showed that combining annotations from multiple sources will provide us with a more complete annotation and subsequently metabolic network reconstruction[3]. We initiated the model development process by using a new app developed by members of our team for the DOE KBase platform that combines annotations of *K. algicida* OT-1 genome from a number of sources. We then used the KBase's "Build Metabolic Model" app to generate a draft genome-scale model. The model was then curated using data from the literature as well as new bioinformatic analyses. The curated model was used to examine the metabolic capabilities and needs of *K. algicida* for a variety of different conditions. The results point to a metabolism adapted to consuming microbial biomass components such as amino acids and nucleotides as well as a variety of sugars and small organic acids. The analyses also showed that *K. algicida* is auxotrophic for a number of amino acids.

Using the results from the system-level analyses of *K. algicida* and published experimental measurements[1,4], we developed a computational model of the dynamics of interaction between *K. algicida* and *P. tricornutum*. It has been shown that the *K. algicida* uses a regulated protease excretion mechanism for algal lysis[1]. The model is composed of a series of coupled ordinary differential equations that solve for the transient changes in concentration of microbes, critical metabolites, and the protease. We used the model to examine various interaction scenarios and discovered that because of *K. algicida*'s dependence on *P. tricornutum* as its primary source of some amino acids, unregulated lysis of the algae will eventually result in cessation of growth by the bacterium. Additionally, significant accumulation of the protease in the co-culture media will result in decreased growth for *K. algicida* and elimination of *P. tricornutum*. These findings have provided novel avenues of research to experimentally test with *in vitro* cultures and with increasing community complexity.

References

1 Paul, C. & Pohnert, G. Interactions of the algicidal bacterium Kordia algicida with diatoms: regulated protease excretion for specific algal lysis. *PloS one* **6**, e21032 (2011).

2 Chisti, Y. Biodiesel from microalgae. *Biotechnology Advances* **25**, 294-306, doi:https://doi.org/10.1016/j.biotechadv.2007.02.001 (2007).

3 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).

4 Sohn, J. H. *et al.* Kordia algicida gen. nov., sp. nov., an algicidal bacterium isolated from red tide. *International journal of systematic and evolutionary microbiology* **54**, 675-680 (2004).

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. LLNL-ABS-818967

Title: Coupled Metabolomics and Transcriptomics Analyses Reveal Active Dynamics of Infection in Virocells

Authors: Roya AminiTabrizi,[1*] (royaaminit@email.arizona.edu), Cristina Howard-Varona,[2] Jane Fudyma,[1] Melissa Duhaime,[3] Malak Tfaily,[1,4], and **Mathew Sullivan**

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. Within this overall project, we established bacterial virus (phage)-host model systems to investigate metabolic reprogramming in virus-infected cells (a.k.a. virocells). Such work is critical for establishing baseline approaches for understanding microcells across ecosystems, including soils, and to translate these findings into biogeochemical data needed to develop predictive computational models of soil microbial functioning and impacts.

Abstract:

Viruses control the microbes that provide essential planetary services through infection. Yet, studying them in complex ecosystems such as soils presents many challenges, ranging from establishing laboratory model systems to analyzing infections and measuring viral infection impacts on the ecosystem. To develop foundational approaches to study viruses in soils, here we used a known, ecologically-relevant bacterium (Pseudoalateromonas) and two contrastingly different infecting phages with (podovirus HP1 and siphovirus HS2) as model system for analytical tool development for analyzing infections and measuring their ecosystem impacts. Specifically, we chose a nutrient-challenged environment (phosphorus, P, limitation) to mimic natural conditions. Here we sought to push multi-omics approaches for two *Pseudoalteromonas* phage-host model systems (podovirus HP1 and siphovirus HS2) to improve mechanistic understanding of how phage and host respond to each other during infection with and without phosphorus (P) limitation. Because metabolite profiling responses are particularly under-studied in these systems, we also measured extra-cellular metabolites via high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS) to follow the dynamics of database-captured exometabolites. This revealed a dynamic and complex response to phage infection in both virus-infected cells (a.k.a. virocells), with a notable increase in polyphenols in both virocells under P-limitation, though with contrasting temporal responses across the two virocells. While these molecules' production as a response to stresses is typical in plants, their production in bacterial cells due to phage infection has not yet been reported. To assess metabolomic response further, we leveraged MAGI (metabolite, annotation, and gene integration), which revealed fatty acid metabolites were elevated in virocells under P-limitation. To improve viral gene annotations (e.g., polyphenolic biosynthesis genes are not annotated), we applied gene-metabolite correlation networks and found that the majority of genes correlating with polyphenols were of unknown functions as opposed to other classes of compounds that were linked to genes with known potential functions. Finally, metabolite-metabolite correlation networks, described as "fingerprints" of metabolic systems, revealed a dynamic viral infection pattern in both virocells at each infection time point. These networks revealed the potential underlying enzymatic system and biomarkers of infection as

indicated by clusters of metabolites separated by time. Together, these findings bring new biological understanding of phage-host interactions and the impacts of nutrient limitation on their dynamics, but also simply provide a roadmap for such analyses to be conducted as virocells are increasingly explored across diverse ecosystems. Our study highlights novel insights in phage-host interactions and provided tools that can extend to new soil model phage-host systems.

References/Publications

1. 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.
2. 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.

Expanding standards in viromics: *in silico* evaluation of viral identification, taxonomy, and auxiliary metabolic genes (AMGs) curation

Akbar Adjie Pratama [1, 2], Ben Bolduc [1, 2], Ahmed A. Zayed [1, 2], Zhi-Ping Zhong [1, 2,] Jiarong Guo [1, 2], Dean Vik [1, 2], Maria Consuelo Gazitua [3], James Wainaina [1, 2], Simon Roux [*4], Matthew B. Sullivan [* 1, 2, 5]

[1] Department of Microbiology, Ohio State University, Columbus, OH, United States
[2] Center of Microbiome Science, Ohio State University, Columbus, OH, United States
[3] Viromica Consulting, Santiago, Chile
[4] Lawrence Berkeley National Lab, Berkeley, CA, United States
[5] Department of Civil, Environmental and Geodetic Engineering, Ohio State University, Columbus, OH, United States
*Corresponding Authors: Simon Roux, Matthew B. Sullivan
Email address: sroux@lbl.gov, sullivan.948@osu.edu

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. Specifically, we aim to provide recommendations for how best to analyze (dsDNA) viruses in viromes and bulk metagenomic samples. We use *in-silico* datasets that mimic viromes and bulk metagenomes with varied inference from non-virus 'distractor' sequences to evaluate (*i*) options for viral identification, (*ii*) genomic fragment sizes for viral classification via gene-sharing networks, as well as (*iii*) provide guidelines for best practices for the evaluation of candidate auxiliary metabolic genes (AMGs). These analyses and results contribute to the growing set of community-driven benchmarks and guidelines in the field of environmental virology.

Abstract:
Metagenomic approaches have been critical for revealing viral roles across diverse ecosystems, driving force in microbial diversity and nutrient cycling. However, the emergent field of viral ecogenomics would benefit from comparative benchmarking to better standardize and enable comparison across datasets. Here we constructed *in silico*-generated datasets that mimicked features of viromes and bulk metagenomes, and used these to provide guidelines and highlight potential pitfalls of viral metagenomic analyses. We compared the performance of the most commonly-used viral identification tools, evaluate viral taxonomic assignations, and propose guidelines to systematize the evaluation of candidate AMGs.

The *in silico* benchmarking of five commonly-used viral identification tools show that gene-content-based tools consistently performed well for long (≥3 kbp) contigs, while *k*-mer- and blast-based tools were uniquely able to detect viral sequences from short (<3 kbp) contigs. Notably, however, the performance increase of *k*-mer- and blast-based tools for short contigs was obtained at the cost of increased false positives (sometimes up to ~40%), particularly when eukaryotic or mobile genetic element sequences were included in the test datasets. For viral classification, variously sized genome fragments were assessed using gene-sharing network analytics to quantify drop-offs in taxonomic assignments, which revealed correct assignations ranging from ~90% (whole genomes) down to ~50% (3 kbp sized genome fragments). Finally, we highlight how fragmented assemblies can lead to erroneous identification of AMGs, and outline a comprehensive workflow that can be used to curate candidate AMGs in viral genomes assembled from metagenomes. Together these benchmarking experiments provide guidance for researchers seeking to best detect and characterize the myriad viruses 'hidden' in diverse sequence datasets.

Funding statement: This research was supported by the U.S. Department of Energy (#DE-SC0020173 and #248445), and the Gordon and Betty Moore Foundation (#3790) to MBS. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. An award from the Ohio Supercomputer Center (OSC) to MBS supported computing resources used here.

Viral ecology through time and along a permafrost thaw gradient

Christine L. Sun*[1,2,3†] (sun.2508@osu.edu), Akbar Adjie Pratama[1,2,3†], Consuelo Gazitua[4], Ahmed Zayed[1,2,3], Dylan Cronin[1,2,3], Lindsey Solden[1], Benjamin Bolduc[1,2,3], the IsoGenie Project Field Teams 2010-2017, the IsoGenie Project Coordinators, Virginia 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. [†]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.

Permafrost is thawing due to elevated temperatures resulting from climate change. It is important to understand how thawing permafrost will impact the release of greenhouse gasses since permafrost accounts for 30%-50% of global soil carbon (C). Microorganisms play a critical role in the terrestrial C cycle, but viruses are less well-studied in soils. However, in marine systems, viruses are known to impact C cycling by controlling host microbial communities via predation, transferring genes between hosts, and metabolically reprogramming host cells via encoding auxiliary metabolic genes (AMGs). Here, we examined viruses identified in 379 bulk soil metagenomes derived along a permafrost thaw gradient (palsa, bog and fen) in Stordalen Mire, Sweden, over eight years (2010-2017), and use these data to establish baseline ecological understanding. In total, we identified 8,597 unique viral operational taxonomic units (vOTUs; ≥ 5 kb contigs dereplicated at 95% average nucleotide identity and 80% coverage), which gene-sharing network analyses suggested represent 1,609 genera, of which 617 are novel (compared to NCBI Viral RefSeq). Ecologically, many vOTUs were shared across sites (e.g., 50% of bog vOTUs were shared with at least one other site), but ordination separated viral communities based on their origin along the thaw gradient. Alpha diversity differed within each site, and increased with depth. While some vOTUs were found in each site through the entire time series, the vast majority of OTUs were ephemeral as they were only detected in a single sample. Hosts for these viruses were *in silico* predicted, with the benefit of co-sampled metagenome-assembled genomes (MAGs), and this revealed 830 MAGs as hosts for

992 vOTUs. Of these MAGs, 790 (95%) have genome encoded functions that include degradation of specific carbon compounds (such as cellulose, fructose, and xylose), implicating viruses in C cycling via infection. In addition, we identified 1,349 instances of 81 unique AMGs in vOTUs from a diversity of metabolic genes, including those involved in central C metabolism and C degradation (e.g. glycoside hydrolase families). Together, these results provide baseline ecological patterns, as well as evidence that soil viruses may impact permafrost C cycling through diverse means.

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).

A Nitrogenase-like Methylthio-alkane Reductase Complex Catalyzes Anaerobic Methane, Ethylene, and Methionine Biosynthesis

Justin A. North,[1] Srividya Murali[1]* (murali.4@osu.edu), Adrienne B. Narrowe,[3] Weili Xiong,[4] Kathryn M. Byerly,[1] Sarah J. Young,[1] Yasuo Yoshikuni,[5] Sean McSweeney,[6] Dale Kreitler,[6] William R. Cannon,[2] Kelly C. Wrighton,[3] Robert L. Hettich,[4] and F. Robert Tabita[1] (former PI, deceased)

[1]Department of Microbiology, The Ohio State University, Columbus, OH; [2]Pacific Northwest National Laboratory, Richland, WA. [3]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO; [4]Chemical Sciences Division, ORNL, Oak Ridge, TN; [5]DOE Joint Genome Institute, Berkeley, CA; [6]NSLS-II, Brookhaven National Laboratory, Upton, NY.

Project Goals: The goal of this project is to identify and characterize the specific enzyme(s) that catalyze anaerobic ethylene synthesis. This is part of a larger project to develop an industrially compatible microbial process to synthesize ethylene in high yields. The specific goals are:

1. Identify the genes and gene products responsible for anaerobic ethylene synthesis.
2. Probe the substrate specificity and metagenomic functional diversity of methylthio-alkane reductases to identify optimal bioproduct generating systems.
3. Characterize the enzymes and the reactions that directly generate anaerobic ethylene.

Abstract Text: Our previous work identified a novel anaerobic microbial pathway (DHAP-Ethylene Shunt) [1] that recycled 5'-methylthioadenosine (MTA) back to methionine with stoichiometric amounts of ethylene produced as a surprising side-product. MTA is a metabolic byproduct of methionine utilization in a multitude of cellular processes. The initial steps of the DHAP-ethylene sequentially converts MTA to dihydroxyacetone phosphate (DHAP) and ethylene precursor (2-methylthio)ethanol (Fig. 1; gray). However, the terminal enzyme(s) responsible for ethylene biogenesis and regeneration of methionine we unknown.

Genes and gene products MarHDK responsible for anaerobic ethylene synthesis: We sought to identify the genes and proteins responsible for ethylene production via proteomics, transcriptomics, and specific gene deletion studies [2]. Ethylene production from (2-methylethio)ethanol is highly regulated by the presence of exogenous sulfate. Therefore, cells were grown under sulfate replete (ethylene suppressing) and sulfate limiting (ethylene inducing) anaerobic growth conditions. Cells were harvested and differential proteome analysis via HPLC-MS/MS was performed to identify proteins that increased in abundance during ethylene inducing conditions. Proteins with the highest increase in abundance during production of ethylene from (2-methylthio)ethanol corresponded to novel nitrogenase-like proteins and previously characterized O-acetyl-L-homoserine sulfhydrylases [2,3] (Fig. 1; rxns 1, 3). This nitrogenase-like complex is named methylthio-alkane reductase, with components MarB and MarHDK based on homology to *bona fide* nitrogenase systems. We further probed the substrate specificity for this novel methylthio-alkane reductase process. Indeed, other small volatile organic sulfur compounds (VOSCs) required the nitrogenase-like gene products to be utilized as a sulfur source by *R. rubrum* for growth and methionine metabolism. Utilization of dimethyl sulfide, the most abundant VOSC in the environment, resulted in stoichiometric production of methane (Fig. 1), and ethylmethyl sulfide led to stoichiometric production of ethane gas [2]. This is the first indication of a nitrogenase-like complex responsible for the reduction of a carbon-sulfur bond, and the first observation of a nitrogenase-like complex involved in sulfur (methionine) metabolism.

Diversity of Organisms with MarHDK: The nitrogenase superfamily is composed of the *bona fide* nitrogenases (groups I-III), which reduce N2 into ammonia, and the Nitrogen Fixation-Like proteins (groups IV-VI), which catalyze a diverse number of reactions of known an unknown function. The methylthio-alkane reductase gene products MarHDK are homologous to nitrogenase proteins NifH (nitrogenase reductase iron protein), and NifDK (catalytic subunits α/β). Methylthio-alkane reductase sequences mined from metagenome databases form a distinct clade (Group IV-C) within the Group IV family of Nitrogen Fixation-Like sequences, and initial tested representatives exhibit activity.

Structure and mechanism of MarHDK: A central part of determining the mechanism of methylthio-alkane reductase is the characterization of its catalytic cofactor and structure of the surrounding protein active site. In collaboration with the JGI FICUS program and the Brookhaven National Laboratory NSLS-II, we are working to solve the structure of the *R. rubrum* methylthio-alkane reductase to understand how ethylene and methane are produced by this system. Initial subunit isolation and purification endeavors reveals that MarH is indeed analogous to NifH and MarDK is analogous to NifDK. This suggests a similar yet distinct structure and mechanism between the two, resulting in their unique respective catalytic functions.

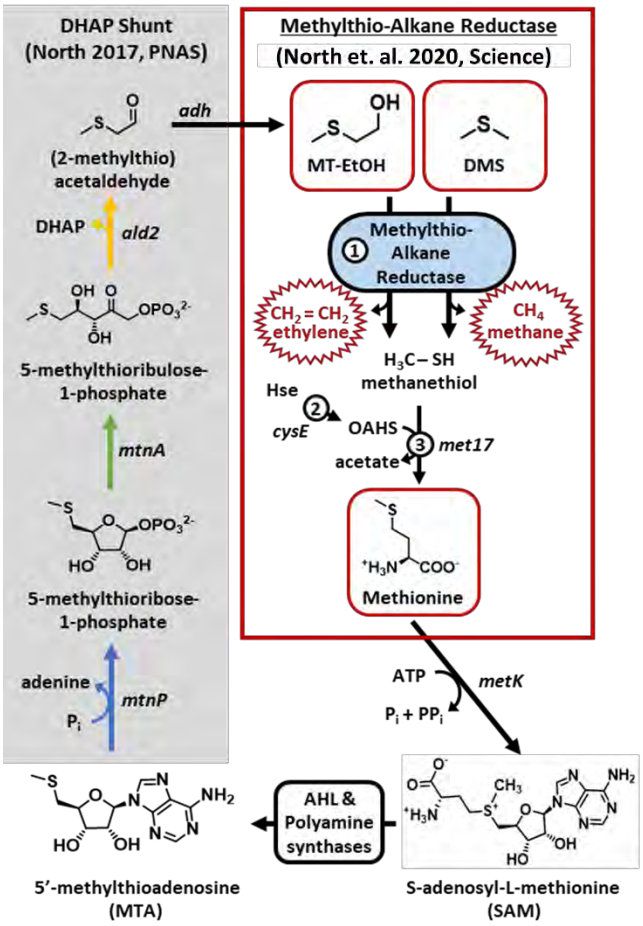


Fig. 1. Full elucidation of the Anaerobic Ethylene Cycle involving a novel nitrogenase-like methylthio-alkane reductases with wide substrate specificity for volatile organic sulfur compounds [2]. (1) Methylthio-alkane reductase (*marHDK*), (2) homoserine acetyltransferase (*cysE)*, (3) acetylhomoserine Sulfhydrylase (*met17*).

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. Erb TJ, Evans BS, Cho K, Warlick BP, Sriram J, Wood BM, Imker HJ, Sweedler JV, Tabita FR, Gerlt JA. A RuBisCO-like protein links SAM metabolism with isoprenoid biosynthesis. *Nat Chem Biol.* **2012,** 8:926–932

Funding Statement: The U.S. DOE, Office of Science, Office of Biological and Environmental Research, Genomic Science Program supported the genomics work (DE-SC0019338 to F.R.T, K.C.W., W.R.C., and J.A.N.) and proteomics work. DNA synthesis and 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.)

Engineering of Enhanced Microbial Anaerobic Ethylene Synthesis Through Predictive Modeling and Metagenomic Functional Gene Discovery

Justin A. North[1]* (north.62@osu.edu), Adrienne B. Narrowe,[2] Sarah J. Young,[1] Kelly C. Wrighton,[2] William R. Cannon,[2] and F. Robert Tabita[1] (former PI, deceased)

[1]Department of Microbiology, The Ohio State University, Columbus, OH; [2]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO; [3]Pacific Northwest National Laboratory, Richland, WA.

Project Goals: Ethylene is the most widely employed organic precursor compound in industry. The potential to impact ethylene formation via the recently discovered microbial "anaerobic ethylene cycle" is promising. This drives the following specific aims:

1. Construct a thermo-kinetic predictive model of microbial metabolism including the the "anaerobic ethylene cycle" to identify hypothesis-driven targets for engineering optimal regulation and metabolic activity.
2. Perform metagenomic analysis of cultured and uncultured organisms, DNA synthesis of candidate genes, and activity screening to identify functional analogs of ethylene cycle genes.

Abstract Text: Ethylene is the highest production industrial platform chemical, required for the manufacturing of plastics, detergents, and solvents. Supplementing current fossil fuel-based ethylene with ethylene produced biologically from renewable sources is desirable to reduce carbon emission. This is feasible using the anerobic ethylene cycle discovered and characterized by our group (Fig. 1A), which is found in photosynthetic and lignocellulosic bacteria [1,2]. Engineering of high yields in industrially tractable bacteria such as *Rhodospirillum rubrum* and *Clostridium cellulolyticum* requires optimization of ethylene cycle pathway kinetics, synthesis of requisite cofactors by supporting pathways, and coordinated gene expression.

Predictive modeling of ethylene production and identifiable bottlenecks: Microbial engineering of ethylene production and subsequent methionine regeneration is highly complex. Synthetic biology engineering endeavors need to address trade-offs between the production of intended targets and the energetic and regulatory requirements of the cell. To specifically address these needs, we have developed hybrid thermodynamic-kinetic model of microbes, including *R. rubrum*, that allows us to characterize energy requirements of native and engineered pathways as well as the regulatory requirements. This model involves approximately 200 metabolic reactions including central carbon metabolism, generic protein, RNA and DNA synthesis reactions, and the ethylene cycle. Currently, key testable constraints identified by the model that limit ethylene production include:

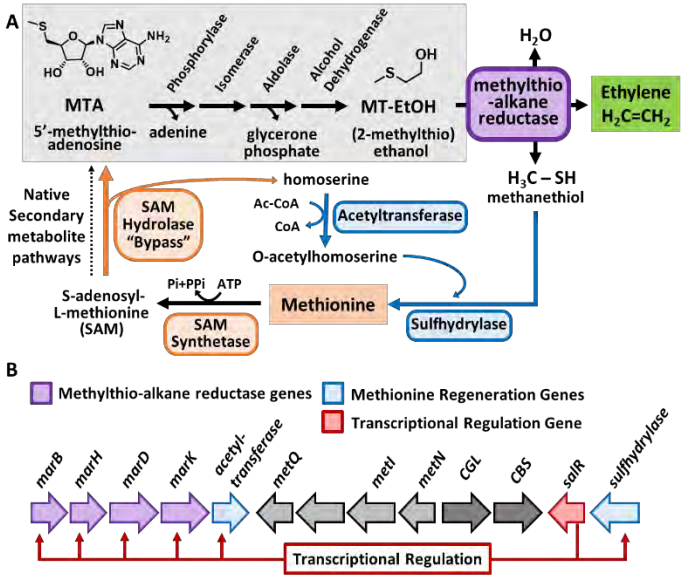


Fig. 1: A) Anaerobic ethylene cycle and B) regulatory network under SalR transcription control.

i. *Ethylene precursory compound synthesis rate:* The ethylene precursor, MTA, metabolized by the DHAP-ethylene shunt (Fig. 1A, gray) is produced natively by multiple pathways. Modeling predicts that increased MTA synthesis rate will increase ethylene production. Initial introduction into *R. rubrum* of the viral enzyme SAM hydrolase (Fig. 1A, bypass) increases ethylene >50-fold.

ii. *Regulation of methionine synthesis:* The model predicts that increased methionine synthesis will increase ethylene production. The methylthio-alkane reductase and methionine regeneration enzymes are under tight transcriptional control by SalR (Fig. 1B) [2]. Replacement of the methylthio-alkane reductase promoter with active promoters increased ethylene yields 10-fold.

Functional diversity of metagenomic ethylene cycle enzymes: Previous studies indicated that aldolase and isomerase orthologous genes could substantially enhance ethylene levels relative to endogenous genes [2]. To scale up the search for ethylene-enhancing orthologs we applied targeted functional metagenomics. Mining of JGI IMG/M genome and metagenome sequence databases for candidate orthologs to the isomerase and aldolase genes (Fig. 2) yielded candidates covering a wide variety of environments including wetlands, forest soils, rhizosphere, and bioreactors. A subset of these genes have been synthesized by the JGI DNA Synthesis Science program and screened via our high throughput *E. coli* lysate activity assay (Fig. 2). Genes for aldolase orthologues with measurable activity in *E. coli* were introduced into the *R. rubrum* aldolase deletion stain. Ethylene was measured and several aldolase orthologues were identified that alone increase ethylene yields 1.3-fold compared to the native enzyme (Fig. 2). Similar approaches will be taken with the isomerase, and combinations of optimal enzymes that increase ethylene yields will be employed.

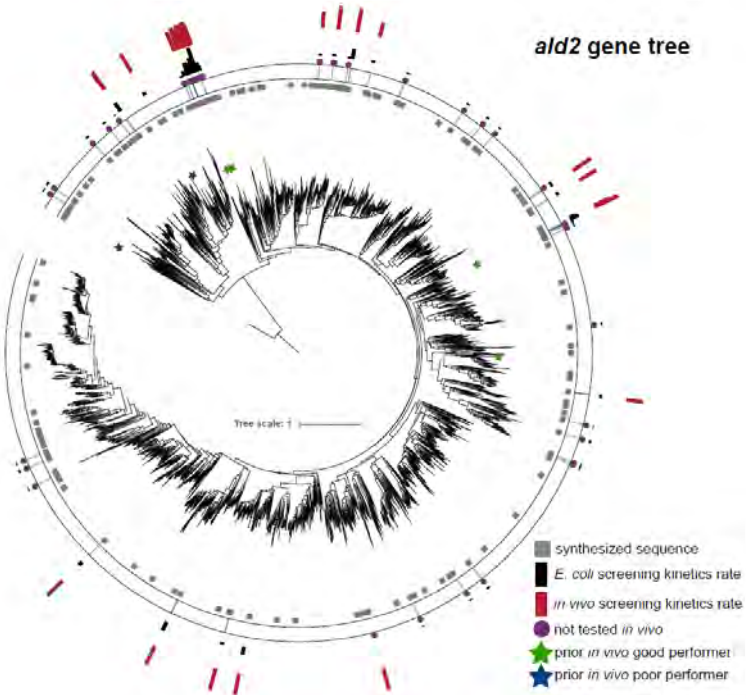


Fig. 2: Functional screen of metagenomic ethylene cycle aldolases

References:

1. 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
2. North JA, Miller AR, Wildenthal JA, Young SJ, Tabita FR. Microbial pathway for anaerobic 5'-methylthioadenosine metabolism coupled to ethylene formation. *PNAS*. **2017**. 114: E10455-E10464
3. Hultqvist JJ, Warsi O, Söderholm A, Knopp M, Eckhard U, Vorontsov E, Selmer M, Andersson DI. A bacteriophage enzyme induces bacterial metabolic perturbation that confers a novel promiscuous function. *Nat Ecol Evol.* **2018.** 2(8):1321-1330

Funding Statement: This work was supported the U.S. DOE, Office of Science, Office of Biological and Environmental Research, Genomic Science Program (DE-SC0019338 to F.R.T, K.C.W., W.R.C., and J.A.N.), and the DOE JGI DNA Synthesis Science Program (505489 to K.C.W, F.R.T., A.B.N., and J.A.N.)

High-Throughput Isolation of Bacteria to Dissect Corrinoid Based Interactions in Soil

Zoila I. Alvarez-Aponte[1] (zoila.alvarez@berkeley.edu), Zachary F. Hallberg,[1] Alexa M. Nicolas,[1] Hans K. Carlson,[1,2] Eoin L. Brodie,[2] Mary K. Firestone,[1] and **Michiko E. Taga[1]**

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

https://tagalab.berkeley.edu

Project Goals: The goal of this project is to gain a comprehensive understanding of microbial metabolic interactions in soil through the study of corrinoids as model metabolites. Corrinoids are a family of structurally diverse cofactors that includes vitamin B_{12}. They are produced by a subset of the bacteria that require them, and thus are shared metabolites. Importantly, bacteria typically can use more than one corrinoid but display distinct preferences for specific structures. For these reasons, we hypothesize that corrinoids are key metabolites that contribute to shaping microbial communities in soil. By using this shared, structurally diverse family of metabolites, we are studying microbial interactions across scales, from the whole community to individual isolates.

Microbial communities in soil comprise much of Earth's microbial diversity and are major drivers of biogeochemical cycles. Interactions between microbes are essential determinants of community structure, function, and dynamics. Disentangling these interactions is crucial for understanding the soil microbiome's global role. However, the physicochemical and taxonomic complexity of the soil microbiome poses a challenge. By focusing on corrinoids as model metabolites, we aim to mechanistically study complex metabolic interactions between soil microbes. Corrinoids are a family of at least 16 structurally diverse cofactors that includes vitamin B_{12}. They are shared nutrients, required by an estimated 86% of all sequenced bacteria but produced by only 36% of bacteria. Importantly, bacteria typically can use more than one corrinoid, but display distinct preferences for specific corrinoid structures. For these reasons, we hypothesize that corrinoids are key nutrients in soil and that corrinoid composition can impact microbial community structure.

By isolating soil microbes on media supplemented with different corrinoids, we will 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 B12 have never been used in growth media, species that prefer them have likely remained uncultured, leading to a pervasive bias across microbiology. To test these hypotheses, we did a high-throughput isolation of soil microbes on media supplemented with one of six corrinoids. We found that adding a corrinoid to the medium results in an increase in isolated soil bacteria. This finding supports our hypothesis that a significant number of soil bacteria are corrinoid-dependent. 158 isolates are currently

being identified by 16S rRNA amplicon sequencing and will be experimentally characterized as corrinoid producers, auxotrophs, or independent. These isolates and knowledge of their corrinoid metabolism will be crucial for future steps in this research aimed at predicting and testing corrinoid-sharing interactions.

This research is funded by the DOE Genomic Sciences Program Grant DE- FOA-00002059 (to MT).

Corrinoids as model nutrients to probe microbial interactions in a soil ecosystem

Zachary F. Hallberg (zhallber@berkeley.edu),[1] Zoila I. Alvarez-Aponte,[1] Alexa M. Nicolas,[1] Hans K. Carlson,[2] Mary K. Firestone,[1] 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 examined/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. We optimized a chemical extraction process for corrinoids from an annual grassland soil that circumvents the ability of soil to adsorb corrinoids. Using this method, we

show that, despite the ability of many soil-derived organisms to produce alternate corrinoids, Vitamin B12 is the predominant corrinoid in bulk soil. We also demonstrate that Vitamin B12 addition to soil enrichment cultures alters their species composition. Unexpectedly, despite the overwhelming presence of Vitamin B12 in this soil, addition of alternate corrinoids elicited marked responses in enrichment culture assembly. This study identifies key corrinoid-limited taxa that are desirable targets to understand corrinoid sharing in natural soil.

This research is funded by the DOE Genomic Sciences Program Grant DE-FOA-00002059 (to MT).

Reliable bioinformatic prediction of cobamide biosynthesis by core biosynthesis genes and taxonomy

Amanda N. Shelton,[1] Alexa M. Nicolas,[2*] Alekhya Govindaraju,[2] Zachary F. Hallberg,[2] Zoila I. Alvarez-Aponte,[2] Kenny C. Mok,[2] and **Michiko E. Taga**[2]

[*] Alexa_Nicolas@berkeley.edu
[1] *Department of Plant Biology, Carnegie Institution for Science*
[2] *Department of Plant & Microbial Biology, University of California, Berkeley, Berkeley, CA, USA.*

https://tagalab.berkeley.edu

Project Goals: The goal of this project is to gain a comprehensive understanding of microbial metabolic interactions in soil through the study of corrinoids as model metabolites. Corrinoids are a family of structurally diverse cofactors that includes vitamin B_{12}. They are produced by a subset of the bacteria that require them, and thus are shared metabolites. Importantly, bacteria typically can use more than one corrinoid but display distinct preferences for specific structures. For these reasons, we hypothesize that corrinoids are key metabolites that contribute to shaping microbial communities in soil. By using this shared, structurally diverse family of metabolites, we are studying microbial interactions across scales, from the whole community to individual isolates.

The biosynthesis of cobamides, such as example Vitamin B12, is unevenly distributed across bacteria and archaea. However, these essential cofactors are required for diverse cellular reactions in a majority of bacteria that cannot biosynthesize them. Therefore, cobamides, as a category of nutrients, provide an important nutrient to study interactions that underlie complex ecosystems, such as the soil microbiome. Recently, comparative genomic studies suggest that cobamide biosynthesis can be predicted by genome sequence alone. We sought to extend these results to Metagenome-Assembled Genomes (MAGs) and test whether there are phylogenetic trends across taxa that can reliably predict whether an organism can synthesize cobamides *de novo* or is an auxotroph. Previous findings suggest three core cobamide biosynthesis genes may be predictive of biosynthesis capacity. However, how well these predictions can estimate cobamide production *in situ* in MAGs is still unknown. We used CAMISIM to simulate reads for a set of seven complete genomes previously predicted to either encode complete cobamide biosynthesis or not. The use of KOfams for the three core cobamide biosynthesis genes as a proxy for the complete pathway is able to recapitulate cobamide biosynthesis predictions for near-complete MAGs, predicted by the presence of 53 single copy genes.

We sought to determine taxonomic trends in cobamide biosynthesis, as a way to predict potential cobamide sharing networks in environments. Of 11,436 genomes from IMG we examined whether taxonomic patterns existed for their previously categorized cobamide production and dependence. We found that for 59 genera out of 181, 75% or more of the species categorized were not capable of complete cobamide production, suggesting that the inability to produce cobamides may be predicted for a subset of genera. Conversely, for 53 genera 75% or more species studied are predicted to be complete cobamide

producers, suggesting that for this subset of genera cobamide predictions may be extended to a species level. Taken together our findings begin to enable predictions for communities' capacity to produce and use cobamides and generate hypotheses for inter-community interactions centered around cobamides.

This research is funded by the DOE Genomic Sciences Program Grant DE- FOA-00002059 (to MT).

Title: Understanding the microbial controls on biogeochemical cycles in permafrost ecosystems

Neslihan Taş[1*], Megan Dillon[1], Yaoming Li[1], Craig Ulrich[1], Yuxin Wu[1], Mary-Cathrine Leewis[2], Mark Waldrop[2], Susannah Sliebner[3], Christopher Chabot[4], Rachel Mackelprang[4] and Hoi-Ying Holman[1]

[1]Lawrence Berkeley National Laboratory, [2] The United States Geological Survey, [3]GFZ German Research Centre for Geosciences Postdam, [4]California State University Northridge

Contact: ntas@lbl.gov

Project Lead Principal Investigator (PI): Neslihan Taş

BER Program: Genomic Science

Project: Early Career Research Project

Project Abstract:

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 soils are one of the world's largest terrestrial carbon storages thus an important focal point for climate change research. With increasing global temperatures, permafrost carbon stores may become available for rapid microbial mineralization and result in increased greenhouse gas (GHG) emissions. Upon permafrost thaw microbial metabolism leads to decomposition of soil organic matter, substantially impacting the cycling of nutrients and significantly affecting the arctic landscape. Permafrost microbiome is a seed bank of mostly novel organisms that have a diverse and broad metabolic potential. In-depth functional characterization of the permafrost microbes is needed to provide a foundation for understanding their responses to thaw. In order to address this gap in our knowledge we performed a pan-Arctic comparative analysis of permafrost metagenomes in which we study biogeography and metabolic functions of permafrost metagenomes assembled genomes (MAGs). This pan-Arctic analysis of permafrost MAGs across multiple locations showed weak correlations to environmental conditions (ice content, topography, continuity, active layer depth, and vegetation) or soil chemistry.

The microbial response to thaw in arctic environments is not uniform and the relationship between permafrost microbiomes and GHG emissions is not well understood. Especially the fate of carbon in deep permafrost, which is currently protected from the warming climate, is uncertain. Following thaw, redistribution of water is a key event that conditions the permafrost for microbial decomposition. We tested the impact of soil moisture availability under microaerophilic and anaerobic conditions via small scale batch experiment. We couple omics (metagenomics and metatranscriptomics) methods with analysis of soil chemistry via synchrotron fourier transform

infrared (sFTIR) spectral imaging at the Berkeley Infrared Structural Biology beamline of the Advanced Light Source (LBNL). Upon thaw variety of organic compounds and metabolites were accumulated. For example, under saturated high-moisture conditions carbohydrates were depleted and soils accumulated aliphatic compounds. This microbial response was tied to the competition between methanogenesis and iron and/or sulfate reduction processes. This project use field observations, laboratory manipulations, and multi-omics approaches to examine how microbial processes, biogeochemical transformations, and hydrology interact during permafrost thaw in different sites in Alaska in order to determine how these factors drive biogeochemical cycles.

Title: A new 15 acre field plantation for CBI harnessing the natural diversity of *Populus trichocarpa* and determining the genetic basis of drought tolerance

Jack H.Bailey-Bale[1,2], Peter H. Freer-Smith[1], Suvočarev K[4] Jin-Gui Chen[3] and Wellington Muchero[3], Stephen DiFazio[5], Jerry (G)A Tuskan[3], Timothy J Tschaplinski[3] **Gail Taylor** gtaylor@ucdavis.edu [1]

[1] Department of Plant Sciences, UC Davis, [2]presenting author [3] Oak Ridge National Laboratory, [4] Department of Land Air and Water Resources, UC Davis, [5] Department of Biology, West Virginia University

Project Goals: *Populus* serves as a promising biomass feedstock for a suite of industrial applications including biofuels conversion. Drought is one of the most important factors limiting cost-effective production of *Populus* biomass. It is imperative to overcome this obstacle to achieve sustainable production of *Populus* biomass. However, drought response in *Populus* is a complex trait requiring the regulation and coordinated interactions of many genes, and identification of genetic networks regulating drought response remains unaccomplished and is urgently needed to inform genetic improvement of *Populus* feedstocks for sustainable biomass production. The goal of this research is to determine the genetic basis of drought tolerance in bioenergy *Populus* enabling tree improvement and the wide-scale deployment of *Populus* for bioenergy in marginal and droughted environments. Fast growing feedstock crops are required for a future bioeconomy where plant-based biofuels, chemicals and biomass for Bioenergy with Carbon Capture and Storage (BECCS) will be utilized as part of a more sustainable, energy secure economy.

Methods: This project has harnessed the natural genetic variation found in wild trees along the Pacific coast of North America. Originating from a range of naturally droughted and wet environments, the population possesses a number of unique genetic and genomic resources to help understand the trees response to water deficit.

Despite the hurdles of 2020, over 17,000 *P. trichocarpa* cuttings were planted in the greenhouse with a very high success rate for cutting survival. After two months growth the cuttings were transferred to a lath house in very close proximity to the field site. The flourishing cuttings were left to acclimate for eleven days before transplantation to the field.

Across a 5 acre drought site and 10 acre control site over 7000 *P. trichocarpa* trees were transplanted in early April 2020 and began to rapidly accumulate biomass. Daily field visits and maintenance were essential for the establishment of the site throughout the season. During this, the first growing season, 2020, trees were all kept “fully irrigated” and at the same time, two Eddy Covariance towers were deployed at the site, one each in “control” and future “drought” plots. These measurements will enable us, in the long-term to quantify the carbon balance of the site, including the potential of these trees for long-term carbon sequestration.

Objectives:

(i) To deploy a large-scale experimental drought trial for up to 1000 unique genotypes of *Populus* equipping the sites with controlled irrigation and drought treatments that are fully automated and monitored.

(ii) To test the hypothesis that a suite of traits identified for drought tolerance in *P. nigra* can be measured in drought and control treatments in the wide germplasm collection of *P. trichocarpa.*

Impacts and benefits: This research will deliver a new 15 acre common garden collection of the CBI *Populus trichocarpa*, that will be open for multiple collaborative opportunities. The replicated and blocked experiment, with a 5 acre plot of selected genotypes subjected to drought, and the addition of Eddy Covariance towers to quantify carbon flux and move towards assessing total carbon balance and carbon sequestration, should be a significant future value for multiple projects.

Publications

Taylor G., *et al* (2019). Sustainable bioenergy for climate mitigation: developing drought tolerant trees and grasses. Annals of Botany, 124: 513-520.

Muchero W. *e al* (2018). Association mapping, transcriptomics, and transient expression and transient expression identify candidate genes mediating plant-pathogen interactions in a tree. *Proc Natl Acad Sci USA* 115: 11573-11578.

This research was funded by the DOE, BER, DE-FOA-0002060, Genomics enables plant biology for determination of gene function.

Title: Identifying traits that underpin the genetic basis of drought tolerance in bioenergy poplar

Marie C. Klein[1,2], Zi Meng[1], Jack H. Bailey-Bale[1], Troy S. Magney[1], Tom N. Buckley[1] Jin-Gui Chen[3] and Wellington Muchero[3], **Gail Taylor** gtaylor@ucdavis.edu [1]

[1] Department of Plant Sciences, UC Davis, [2]presenting author [3] Oak Ridge National Laboratory

Project Goals: *Populus* serves as a promising biomass feedstock for a suite of industrial applications including biofuels conversion. Drought is one of the most important factors limiting cost-effective production of *Populus* biomass. The drought response in *Populus* is a complex trait requiring the regulation and coordinated interactions of many genes, and identification of genetic networks regulating drought response remains unaccomplished and is urgently needed to inform genetic improvement of *Populus* feedstocks for sustainable biomass production. The goal of this research is to determine the genetic basis of drought tolerance in bioenergy *Populus* enabling tree improvement and the wide-scale deployment of *Populus* for bioenergy in marginal and droughted environments. Fast growing feedstock crops are required for a future bioeconomy where plant-based biofuels, chemicals and biomass for Bioenergy with Carbon Capture and Storage (BECCS) will be utilized as part of a more sustainable, energy secure economy.

Methods: We will harness the natural genetic variation found in wild trees, adapted to different amounts of rainfall. Our main study population consists of 1,000 *P. trichocarpa* collected from a range of naturally droughted and wet environments, that already has a number of unique genomic and genetic resources for a tree species, that will be used here.

Drought tolerance, here defined as '*the maintenance of plant biomass production in the face of moderate and persistent drought stress*' is a highly complex trait and we will use high throughput phenotyping, including tower- and UAV-mounted multi-spectral proximal analysis to unravel how drought tolerance varies across this wide population.

Using the latest computational tools and the extensive sequence, re-sequence and RNASeq resources available and new ones to be developed, we will identify multiple genetic loci linked to traits and resolve them through hierarchical network analysis to the level of cis- and trans-acting eQTL.

We have identified a suite of traits of interest, from our own previous research, from the literature and from traits studied in other organisms including (i) **phenology traits**, that are related to seasonality and are unique in trees. We will focus on remotely-sense canopy greenness and function using tower-based mounted cameras (ii) **anatomical traits,** including stomatal density and index, leaf epidermal cell patterning and aspects of hydraulic functioning in wood. Some of these traits are highly heritable and may be of particular value to breeding and selection and (iii) **functional traits,** including water use efficiency (WUE), stomatal conductance and (iv) **performance traits** where we are quantifying the dynamics of tree biomass accumulation. For these traits, we will deploy UAV-based LiDAR and IR thermography.

Objectives:

(ii) To test the hypothesis that a suite of traits identified for drought tolerance in *P. nigra* can be measured in drought and control treatments in the wide germplasm collection of *P. trichocarpa.*

(iii) To use established and novel GWAS model approaches to identify gene loci linked to drought tolerance traits on interest in *P. trichocarpa*.

(**iv)** To undertake comparative analysis of GWAS results for drought tolerance traits in *P. nigra* and *P. trichocarpa.*

Impacts and benefits: This research will deliver new genotypes of *Populus* that are better suited and able to tolerate droughted environments, that are more likely in the future.

Publications

Taylor G., *et al* (2019). Sustainable bioenergy for climate mitigation: developing drought tolerant trees and grasses. Annals of Botany, 124: 513-520.

Muchero W. *e al* (2018). Association mapping, transcriptomics, and transient expression and transient expression identify candidate genes mediating plant-pathogen interactions in a tree. *Proc Natl Acad Sci USA* 115: 11573-11578.

This research was funded by the DOE, BER, DE-FOA-0002060, Genomics enables plant biology for determination of gene function.

Untargeted metabolomics by high resolution LC-MS/MS revealed different metabolic profiles of oaks (*Quercus* spp.)

Nathalia Graf Grachet,[1*] (nathaliagg@arizona.edu), Megan N. Nickerson,[1] Jana M. U'Ren,[1] and **Malak M. Tfaily**[2]

[1]University of Arizona, Tucson

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.

Abstract

Untargeted tandem mass spectrometry by Fourier transform Ion Cyclotron Resonance preceded by liquid chromatography (LC-FTICR-MS/MS) is a high-throughput and sensitive metabolomics technique that yields high mass accuracy. Such experiments produce information-rich datasets that require the use of many software programs of different programming languages for data analysis. This problem is exacerbated by integrating multi-omics data with metabolomics because of the data heterogeneity. In addition, many of the existing state-of-art software were designed for model organisms, and have limited functionality for complex environmental data and non-model organisms, such as soil and plants. Therefore, our goals in this study were two-fold: i) demonstrate the capabilities of LC-FTICR-MS/MS for metabolic profiling, and ii) to develop a data analysis pipeline for a complex and non-model organism, oak (*Quercus* spp.).

North America is one of the centers of oak genetic diversity, having almost 300 species growing across Mexico, USA and Canada. Continental US has approximately 90 of those species growing from coast to coast (USDA, NRCS, 2021). Despite the incredible diversity of oak species and the social and ecological relevance of oak woodlands, the genetic resources of many species are limited (Plomion et al., 2016, 2018; Sork et al., 2016). Therefore, several aspects about oak species biology, and interactions with the environment still remains to be explored. Metabolomics is a fast-emerging technology in systems biology that remains relatively untapped in environmental and forest sciences. In this study, we used a LC-FTICR-MS/MS technique to perform a metabolic profiling of nine oak species, and to demonstrate its use as a tool for discovery of the genetic basis of phenotypic variation.

Metabolite extracts were prepared by Folch extraction (Folch et al., 1957) from healthy, living leaves. Tandem mass spectrometry was collected on a 21 Tesla (21T) Agilent FTICR-MS

equipped with a Waters ultra-performance liquid chromatography system as previously described in (Fudyma et al., 2019). Negative mode raw chromatography data were preprocessed using the R package XCMS (Benton et al., 2010; Smith et al., 2006). Fragmentation spectra (MS/MS or MS2) were searched against the GNPS High-Throughput Dereplication Comprehensive MS/MS Libraries (Wang et al., 2016). The remaining unannotated spectra were annotated *in silico* using Sirius command-line v.4.5 (Dührkop et al., 2019). Downstream multivariate statistical analyses were conducted with features with MS2 that received at least a molecular formula assignment. Genome-metabolome integration was done using MAGI (Erbilgin et al., 2019). The predicted amino acid sequences of the *Q. robur* reference genome (Plomion et al., 2018) were used in order to unveil potential reactions between the metabolites and genes.

One of the challenges of working with non-model organisms is the lack of useful annotation from publicly available MS2 databases. We only identified 19 compounds in GNPS MS/MS libraries. Using SIRIUS, we were able to assign *in silico* molecular formula and potential structure to 916 compounds, but many of these *in silico* predictions remained unknown because structural matches were not found in public databases. In addition, there is a lack of fragmentation spectra libraries generated by ultrahigh resolution mass spectrometers such as the 21T FTICR.

The main differences observed in this dataset were between deciduous and brevideciduous oak species. PCA was performed on the log-transformed MS2 feature intensities, and the groupings of brevideciduous versus deciduous species drove more the variance within the dataset (~35%). This was in agreement with PERMANOVA, which showed that leaf longevity was a significant factor ($R^2 = 0.28$, $p = 0.004$**). Deciduous oaks showed a higher number of organic acids and benzenoids compared to brevideciduous. Brevideciduous species showed a higher number of lipids and lipid-like molecules compared to deciduous species.

Using thermodynamics, Gibbs free energy (GFE) calculation can be used as an indication of the bioavailability of compounds in which lower GFE values indicate more bioavailable compounds such as carbohydrates and sugars (LaRowe & Van Cappellen, 2011). We observed a significant difference in the GFE values of brevideciduous and deciduous oak species (Kruskal-Wallis test, p = 3.5e-05 ***). Brevideciduous species had a higher median GFE value compared to deciduous, which suggests that brevideciduous have more non-polar compounds represented mainly by lipids and lipid-like molecules that could potentially play a role in plant cell wall formation.

The genome-metabolome integration was done using MAGI, and direct associations between amino acids and compounds were found based on known reactions and pathways found in the BioCyc database (Caspi et al., 2018). We leveraged MAGI for a compound-centric view of biochemical reactions between genes and compounds, and we identified seven compounds directly associated with oak genes. Beta-glucogallin (tannin), astragalin (flavonoid-3-o-glycosides), and quercetin (flavonoid, polyphenol) are phytochemical compounds with medicinal properties with potential uses as anti-inflammatory, antioxidant, and antimicrobial (Li et al., 2016; Puppala et al., 2012; Riaz et al., 2018). Uridine diphosphate (UDP) glucuronic acid and UDP galacturonate are both sugar-like molecules involved in the biosynthesis cascade of plant cell wall polymers such as hemicellulose (Mølhøj et al., 2004;

Reboul et al., 2011). Oxidized glutathione (peptide) and prostaglandin-like molecules (lipid fatty acid) are signaling molecules involved in pathways related to stress responses in plants (Mueller, 1998; Rahantaniaina et al., 2013).

Our study has demonstrated the application of LC-FTICR-MS/MS is metabolic profiling of nine different oak species. The profile of MS features were most distinct between brevideciduous and deciduous species rather than oak type, red vs white. Although many compounds remained unknown, we leveraged different software and strategies to structurally annotate unknown compounds, and were able to identify seven compounds with very important biological roles, including at least three phytochemical compounds well-known for their pharmacological applications.

References

Benton, H. P., Want, E. J., & Ebbels, T. M. D. (2010). Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. *Bioinformatics*, *26*(19), 2488–2489. https://doi.org/10.1093/bioinformatics/btq441

Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S., Subhraveti, P., & Karp, P. D. (2018). The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research*, *46*(D1), D633–D639. https://doi.org/10.1093/nar/gkx935

Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., Dorrestein, P. C., Rousu, J., & Böcker, S. (2019). SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, *16*(4), 299–302. https://doi.org/10.1038/s41592-019-0344-8

Erbilgin, O., Rübel, O., Louie, K. B., Trinh, M., Raad, M. de, Wildish, T., Udwary, D., Hoover, C., Deutsch, S., Northen, T. R., & Bowen, B. P. (2019). MAGI: A Method for Metabolite Annotation and Gene Integration. *ACS Chemical Biology*, *14*(4), 704–714. https://doi.org/10.1021/acschembio.8b01107

Folch, J., Lees, M., & Sloane Stanly, G. H. (1957). A simple method for the isolation and purification of total lipides from animal tissues. *The Journal of Biological Chemistry*, *226*(1), 497–509.

Fudyma, J. D., Lyon, J., AminiTabrizi, R., Gieschen, H., Chu, R. K., Hoyt, D. W., Kyle, J. E., Toyoda, J., Tolic, N., Heyman, H. M., Hess, N. J., Metz, T. O., & Tfaily, M. M. (2019). Untargeted metabolomic profiling of Sphagnum fallax reveals novel antimicrobial metabolites. *Plant Direct*, *3*(11), e00179. https://doi.org/10.1002/pld3.179

LaRowe, D. E., & Van Cappellen, P. (2011). Degradation of natural organic matter: A thermodynamic analysis. *Geochimica et Cosmochimica Acta*, *75*(8), 2030–2042. https://doi.org/10.1016/j.gca.2011.01.020

Li, Y., Yao, J., Han, C., Yang, J., Chaudhry, M., Wang, S., Liu, H., & Yin, Y. (2016). Quercetin, Inflammation and Immunity. *Nutrients*, *8*(3), 167. https://doi.org/10.3390/nu8030167

Mølhøj, M., Verma, R., & Reiter, W.-D. (2004). The Biosynthesis of d-Galacturonate in Plants. Functional Cloning and Characterization of a Membrane-Anchored UDP-d-Glucuronate 4-Epimerase from Arabidopsis. *Plant Physiology*, *135*(3), 1221–1230. https://doi.org/10.1104/pp.104.043745

Mueller, M. J. (1998). Radically novel prostaglandins in animals and plants: The isoprostanes. *Chemistry & Biology*, *5*(12), R323–R333. https://doi.org/10.1016/S1074-5521(98)90660-3

Pereira-Leal, J. B., Abreu, I. A., Alabaça, C. S., Almeida, M., Almeida, P., Almeida, T., Amorim, M., Araújo, S., Azevedo, H., Badia, A., Batista, D., Bohn, A., Capote, T., Carrasquinho, I., Chaves, I., Coelho, A., Costa, M., Costa, R., Cravador, A., … Ricardo, C. P. (2014). A comprehensive assessment of the transcriptome of cork oak (Quercus suber) through EST sequencing. *BMC Genomics*, *15*(1), 371. https://doi.org/10.1186/1471-2164-15-371

Plomion, C., Aury, J.-M., Amselem, J., Alaeitabar, T., Barbe, V., Belser, C., Bergès, H., Bodénès, C., Boudet, N., Boury, C., Canaguier, A., Couloux, A., Da Silva, C., Duplessis, S., Ehrenmann, F., Estrada-Mairey, B., Fouteau, S., Francillonne, N., Gaspin, C., … Kremer, A. (2016). Decoding the oak genome: Public release of sequence data, assembly, annotation and publication strategies. *Molecular Ecology Resources*, *16*(1), 254–265. https://doi.org/10.1111/1755-0998.12425

Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillonne, N., Labadie, K.,

Le Provost, G., Lesur, I., Bartholomé, J., Faivre-Rampant, P., Kohler, A., Leplé, J.-C., Chantret, N., Chen, J., Diévart, A., Alaeitabar, T., … Salse, J. (2018). Oak genome reveals facets of long lifespan. *Nature Plants*, *4*(7), 440–452. https://doi.org/10.1038/s41477-018-0172-3

Puppala, M., Ponder, J., Suryanarayana, P., Reddy, G. B., Petrash, J. M., & LaBarbera, D. V. (2012). The Isolation and Characterization of β-Glucogallin as a Novel Aldose Reductase Inhibitor from Emblica officinalis. *PLoS ONE*, *7*(4), e31399. https://doi.org/10.1371/journal.pone.0031399

Rahantaniaina, M.-S., Tuzet, A., Mhamdi, A., & Noctor, G. (2013). Missing links in understanding redox signaling via thiol/disulfide modulation: How is glutathione oxidized in plants? *Frontiers in Plant Science*, *4*. https://doi.org/10.3389/fpls.2013.00477

Reboul, R., Geserick, C., Pabst, M., Frey, B., Wittmann, D., Lütz-Meindl, U., Léonard, R., & Tenhaken, R. (2011). Down-regulation of UDP-glucuronic Acid Biosynthesis Leads to Swollen Plant Cell Walls and Severe Developmental Defects Associated with Changes in Pectic Polysaccharides*. *Journal of Biological Chemistry*, *286*(46), 39982–39992. https://doi.org/10.1074/jbc.M111.255695

Riaz, A., Rasul, A., Hussain, G., Zahoor, M. K., Jabeen, F., Subhani, Z., Younis, T., Ali, M., Sarfraz, I., & Selamoglu, Z. (2018). Astragalin: A Bioactive Phytochemical with Potential Therapeutic Activities. *Advances in Pharmacological Sciences*, *2018*, 1–15. https://doi.org/10.1155/2018/9794625

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*, *78*(3), 779–787. https://doi.org/10.1021/ac051437y

Sork, V. L., Fitz-Gibbon, S. T., Puiu, D., Crepeau, M., Gugger, P. F., Sherman, R., Stevens, K., Langley, C. H., Pellegrini, M., & Salzberg, S. L. (2016). First Draft Assembly and Annotation of the Genome of a California Endemic Oak Quercus lobata Née (Fagaceae). *G3 (Bethesda, Md.)*, *6*(11), 3485–3495. PubMed. https://doi.org/10.1534/g3.116.030411

USDA, NRCS. (2021). The PLANTS Database (http://plants.usda.gov, 11 January 2021). *National Plant Data Team, Greensboro, NC 27401-4901 USA.*

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W.-T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., … Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, *34*(8), 828–837. https://doi.org/10.1038/nbt.3597

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.

Proteomic Analysis of Robust Yarrowia lipolytica Isolates Reveals Key Processes Impacting Sugar Utilization and Lipid Degradation during Growth on Biomass Hydrolysate

Authors: Caleb Walker[1*] (fyz692@vols.utk.edu), Stephanie R. Thompson[2], Richard Giannone[3], Patricia J. Slininger[2], Bruce S. Dien[2], and **Cong T. Trinh**[1] (ctrinh@utk.edu)

Institutions: [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:

Lipid accumulation from biomass hydrolysate sugars (i.e., mainly glucose and xylose) and then subsequent degradation are complex phenotypes making them difficult to control and engineer in *Yarrowia lipolytica.* Yet, the phenotypic diversity of the undomesticated *Y. lipolytica* clade illuminates desirable traits not found in well-characterized laboratory strains[1]. Recently, draft genomes were assembled for five non-conventional *Y. lipolytica* strains with superior xylose utilization, lipid accumulation, and growth on undetoxified biomass hydrolysates in comparison to the conventional laboratory strain CBS7504[2]. Here, we explored the genetic diversity of these strains and identified singleton and unique genes shared by strains exhibiting desirable phenotypes. Next, bioreactor growth characterization using a switchgrass biomass hydrolysate revealed that the unconventional strain YB420 used xylose to support cell growth and maintained lipid levels while the laboratory strain CBS7504 degraded biomass and lipids when xylose was the remaining carbon source in the biomass hydrolysate. Proteomic analysis identified carbohydrate transporters, xylose metabolic enzymes and pentose phosphate pathway proteins stimulated during the xylose uptake stage for both strains after glucose was depleted. Further, we distinguished proteins in lipid metabolism (e.g., lipase, NADPH generation, lipid regulators, beta-oxidation) activated by YB420 (lipid maintenance phenotype) or CBS7504 (lipid degradation phenotype) when xylose was the remaining carbon source in the biomass hydrolysate. This work highlights the importance of studying nonconventional robust isolates to better understand and engineer complex phenotypes such as lipid accumulation from the non-preferred biomass hydrolysate sugars and subsequent degradation or conversion to other products.

References/Publications

1. 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).

2. 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).

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.*

Multi-OMICs Profiling Reveals Key Genes and Cellular Processes Underlying Ionic Liquid Robustness in *Yarrowia lipolytica*

Authors: Caleb Walker[1*] (fyz692@vols.utk.edu), Richard Giannone[2], Alice Dohnalkova[3], Mary Lipton[3], Young-Mo Kim[3], Samuel Purvine[3], Jennifer Kyle[3], Chaevien Clendinen[3], and **Cong T. Trinh**[1] (ctrinh@utk.edu)

Institutions: [1]University of Tennessee, Knoxville; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington.

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: Microbial solvent tolerance is a desirable phenotype for novel strategies in bioconversion of renewable substrates to replace petroleum-derived chemicals and fuels. In comparison to conventional solvents, ionic liquids (ILs) such as 1-ethyl-3-methylimidazolium acetate ([EMIM][OAc]) have emerged as a novel reaction medium with superior results in bioprocessing due to their ability to dissolve a wider range of compounds and their adjustable properties for enzyme stabilization and activation. ILs, however, are toxic to microbial growth (e.g., 1%-5% (v/v) IL) which must be overcome for whole-cell biocatalysis in ILs. Previously, we generated the most IL-tolerant mutant reported to date with robust growth in up to 18% (v/v) [EMIM][OAc] [1,2]. Here, we used multi-omics and morphological characterization to understand the superior IL-tolerance of these *Y. lipolytica* strains. Specifically, we demonstrated a new criterion to predict key genetic targets from dynamic RNA-sequencing that confer IL-tolerance using both single-gene and dual-gene overexpression library enrichment techniques. Additionally, we discovered *Y. lipolytica* reconfigures membrane composition and cellular compartments to tolerate high concentrations of ILs and shed light on proteins and regulatory machinery responsible for these processes. Taken together, our work demonstrates a new approach to identify genetic targets for reverse engineering robustness and highlights overlooked cellular processes underlying robust phenotypes.

References/Publications

1. 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).
2. Walker, C., Ryu, S., Haridas, S., Na, H., Zane, M., Labutti, K., Lipzen, A., Barry, K., Grigoriev, I., Trinh, C.T.*, 2020. Draft Genome Assemblies of Ionic Liquid Resistant *Yarrowia lipolytica* PO1f and its superior evolved strain YlCW001. Microbiol Resour Announce, 9(999): e01356-19.

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.*

Employing Bacterial Microcompartments To Create Privileged Redox Pools for Biofuel Production

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] Bastian Vögeli,[1] Andre G. Archer,[1] Taylor M. Nichols,[1] Keith E.J. Tyo,[1] Niall Mangan,[1] and **Danielle Tullman-Ercek**[1]

[1]Northwestern University, Evanston, Illinois

https://dtelab.northwestern.edu/research/#nanobioreactors

Project Goals: To compartmentalize metabolic pathways along with enzyme cofactor recycling pathways to increase the yield and efficiency of bioproduction processes

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 (eg in organelles in eukaryotes and in bacterial microcompartments in prokaryotes) to solve issues such as intermediate leakage, toxicity, and byproduct formation. Here we propose to deploy compartmentalization as a strategy to overcome a critical roadblock: the requirement for redox cofactor recycling. In traditional systems, redox cofactors are lost to cellular growth and maintenance needs. By compartmentalizing redox cofactors with the biochemical synthesis enzymes, we anticipate increasing the thermodynamic efficiency and preventing the loss of valuable intermediates and cofactors. If successful, it would be the first direct demonstration of this feature of a bacterial microcompartment, and would provide a tool for improving metabolic pathway performance for all enzymes with redox or other cofactors.

With this poster, we will describe results from coupling modeling with experiments to guide experimental design and facilitate the engineering of novel pathways in microcompartments (MCPs). Understanding of the native 1,2-propanediol utilization MCPs will help improve the performance of our target metabolic pathways: 1,3-propanediol production and medium-chain fatty acid synthesis via reverse β-oxidation (rBox) pathway. We used computational simulation to determine the optimal experimental sampling times, creating an assay for pathway function. The mathematical method uses an eigenvector decomposition of the state space model manifold to estimate the sensitivity of experimental timepoint and metabolite sampling to MCP properties. We applied this assay first to the native 1,2-propanediol MCPs. To gain more in-depth characterization of the composition of MCPs and strains expressing MCPs, we used tandem mass tag (TMT) based proteomics on purified MCPs and cell lysates containing MCPs. We identified proteins that showed significant fold-change in relative quantity between purified MCPs and lysates. Additionally, we made headway in the encapsulation of the non-native metabolic pathways. We incorporated cell-free protein synthesis and cell-free metabolic engineering to test function of a partially encapsulated pathway. Butanoic acid and hexanoic acid were produced

when we supplemented MCPs containing a partial pathway with unencapsulated enzymes in an *in vitro* assay.

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.

Modeling and NMR Methods to Probe Spatial Arrangements in Biomolecules: Towards predictive models of plant cell wall structure

Bennett Addison[1,2*], (mailto:mbennett.addison@nrel.gov), Vivek Bharadwaj,[1,2], Anne E. Harman-Ware,[1,2], Yannick Bomble[1], and **Gerald A. Tuskan**[2]

[1]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN

http://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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

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 have developed a new solid state Nuclear Magnetic Resonance (ssNMR) technique to probe polymer-polymer interactions within the secondary cell wall of ^{13}C-enriched poplar wood in great detail [1]. The major ssNMR observables that informed the construction and validation of the SCW atomistic molecular models of poplar are as follows: 1) Roughly 70% of all lignin shares surface contact with xylan; 2) Roughly 70% of acetylated xylan is bound to cellulose in a linearized 2-fold structure; 3) the inter-polymer distance between xylan acetyl groups and the lignin surface is ~0.4 – 0.5 nm; 4) little to no direct lignin/cellulose contact is observed; and 5) there is evidence for preferential lignin/hemicellulose surface interaction in the order

Methoxy > S-Lignin > G-Lignin. Even for a known polymer composition, 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. Moving forward, we will investigate plant SCW superstructures using a combined ssNMR and MD approach for consideration of molecular-level factors that may be associated with cell wall recalcitrance, sustainability metrics, and other plant phenotypes.

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.

Effects of *Claviceps spp./Epicoccum andropogonis* inoculation on switchgrass phenotypic traits

Bochra A. Bahri*[1,2] (bbahri@uga.edu), Morgan Willis[1,2], Suraj Sapkota[1,2], James Buck[1], Thomas H. Pendergast IV[2,3,4], M. Mazarei[5,] G. C. Bergstrom[6], C. N. Stewart, Jr.[5], Ali Missaoui[2,3], Katrien M. Devos[2,3,4], and **Gerald A. Tuskan**[7]

[1]Department of Plant Pathology, University of Georgia, Griffin, GA; [2]Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA; [3]Department of Crop and Soil Sciences, University of Georgia, Athens, GA; [4]Department of Plant Biology, University of Georgia, Athens, GA; [5]Department of Plant Sciences, University of Tennessee, Knoxville, TN; [6]School of Integrative Plant Science, Cornell University, Ithaca, NY; [7]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN

http://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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Abstract: This project is part of the strategic effort to identify, quantify, and evaluate the risk potential of emerging pathogens in switchgrass. This work addresses several under-investigated topics, such as the effects of specific biotic stresses on switchgrass development. The goals of the work presented here are to 1) identify and quantify the causal agent responsible for false smut on switchgrass in Georgia and 2) evaluate the effects of *Claviceps spp./Epicoccum andropogonis* infection on switchgrass phenotypic traits including flowering, seed yield, and biomass production.

Switchgrass is a perennial C4 grass native to North America that has been used for grazing, soil conservation, and as a bioenergy crop. Switchgrass is affected by several diseases that could impact its industrial development. In 2019, black sclerotia with a brain-like appearance were observed on the surface of seeds of accessions from the Center for Bioenergy Innovation (CBI) genome-wide association study (GWAS) panels planted in Watkinsville, GA, Tifton, GA, and Knoxville, TN (Figure 1). The causal agent in all three panels was identified morphologically and molecularly as *Epicoccum andropogonis* independently at the University of Georgia (UGA) and the University of Tenessee (UTK)/Cornell University. *Epicoccum andropogonis* is a mycoparasite responsible for false smut disease and is known to grow saprophytically on the honeydew of *Claviceps* species that cause ergot disease in grasses. To assess the impact of *Claviceps spp./ E. andropogonis* on switchgrass development, cultivar "Alamo" and "Summer" seeds were inoculated with this complex sampled from the Watkinsville, GA field site, and plant development was followed in the greenhouse for 10 months. ANOVA analysis revealed non-significant differences between plants grown from inoculated and non-inoculated seed in plant height at flowering (p=0.867 for Alamo, p=0.0504 for Summer), the number of tillers (p=0.583 for Alamo, p=0.894 for Summer), panicle length (p=0.506 for Alamo, p=0.218 for Summer), and biomass production (p=0.647 for Alamo, p=0.811 for Summer). However, 10-month old inoculated Alamo and non-inoculated Summer genotypes were significantly taller than non-inoculated Alamo (p=0.00744) and inoculated Summer genotypes (p=1.35e-07), respectively. In addition, inoculated Alamo genotypes had a lower number of seeds per tiller (p=0.0135) and a lower number of spikelets per tiller (p=0.0031) than non-inoculated Alamo. No significant difference was observed between inoculated and inoculated Summer genotypes for either trait (p=0.156 and p=0.664, respectively). Furthermore, chi-square tests showed that inoculated Alamo genotypes had a higher percentage of flowering tillers (p=0.00057) than non-inoculated Alamo. Again, no significant difference was observed between inoculated and non-inoculated Summer plants for this trait (p=0.8256). Overall, these preliminary results suggest that *Claviceps spp./Epicoccum andropogonis* inoculation more significantly impacts the flowering process and seed production of switchgrass than its biomass production. Additional experiments with a larger number of genotypes need to be carried out to confirm this pattern. Quantifying *Claviceps spp./Epicoccum andropogonis* levels in the plant tissues (panicle and leaf organs) is also needed to better interpret our results.

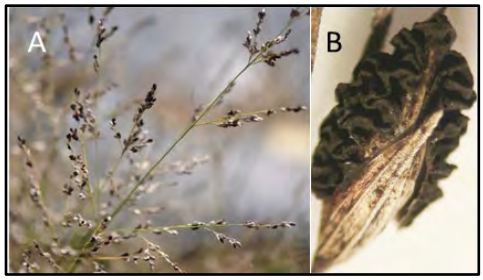


Figure 1: *Panicle (A) and close-up of seed (B) infected with Epicoccum andropogonis.*

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* (andrew.bartling@NREL.gov),[1,2] Michael L. Stone,[2,3] Rebecca J. Hanes,[2,4] Mary J. Biddy,[1,2] Ryan Davis,[1] Jacob S. Kruger,[5] Nicholas E. Thornburg,[1] Yuriy Román-Leshkov[2,3] and Gregg T. Beckham[2,5] and **Gerald A. Tuskan**[2]

[1]Catalytic Carbon Transformation and Scale-Up Center, National Renewable Energy Laboratory, Golden CO, 80401; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN; [3] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge MA; [4]Strategic Energy Analysis Center, National Renewable Energy Laboratory, Golden CO; [5]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden CO

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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by in planta modifications and biological funneling of lignin to value-added chemicals.

Abstract: 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.

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.

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.” Chemical Society Reviews. 2018, **47**, 852-908.

Engineering a Cytochrome P450 System for Oxidative Demethylation of Lignin-Related Aromatics

Alissa Bleem*,[1,2] (alissa.bleem@nrel.gov) Emerald S. Ellis,[3] Daniel J. Hinchen,[4] Lintao Bu,[1] Sam J.B. Mallinson,[1,4] Mark D. Allen,[4] Bennett R. Streit,[3] Melodie M. Machovina,[3,5] Quinlan V. Doolin,[3] William E. Michener,[1] Christopher W. Johnson,[1] Brandon C. Knott,[1] John E. McGeehan,[4] Jennifer L. DuBois,[3] Gregg Beckham[1,2], and **Gerald A. Tuskan**[2]

[1]Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN; [3]Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT; [4]Centre for Enzyme Innovation, University of Portsmouth, Portsmouth, United Kingdom; and [5]Department of Chemistry, University of Illinois Urbana-Champaign, Urbana, IL

cbi.ornl.gov

The Center for Bioenergy Innovation (CBI) is a multidisciplinary center with the vision of *accelerating domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations across the bioenergy supply chain.* The CBI Lignin Valorization Team aims to integrate lignin refining, lignin depolymerization, microbial conversion of lignin-related compounds, and production of new materials from lignin polymers to enable a holistic biorefinery paradigm. The microbial conversion aspect of this effort requires new and improved biological platforms, including bacteria designed to simultaneously catabolize a variety of lignin-related compounds and convert them to valuable products – a process known as "biological funneling".

Abstract: Biological funneling of lignin-related aromatic compounds is a promising approach for valorizing its catalytic depolymerization products. Industrial processes for aromatic bioconversion will require efficient enzymes for key reactions, including demethylation of *O*-methoxy-aryl groups, an essential and often rate-limiting step. The GcoAB cytochrome P450 system comprises a coupled monoxygenase (GcoA) and reductase (GcoB) that catalyzes oxidative demethylation of the *O*-methoxy-aryl group in guaiacol, which serves as the base unit for G lignin. We employed structure-guided protein engineering and detailed biochemical assays to identify mutants of the GcoA monooxygenase that catalyze *O*-demethylation of syringol (the base unit of S lignin) as well as the aromatic aldehydes *o*- and *p*-vanillin. One variant, GcoA-T296S, was utilized for *in vivo* demethylation of *p*-vanillin in *Pseudomonas putida*, an industrially relevant bacterial host. We are also combining structure-guided design with high throughput enzyme evolution screens to identify variants of GcoA that can accept vanillate, a carboxylic acid, as a substrate. This will lay the foundation for a larger effort to compare cytochromes P450 with other enzymatic paradigms for aromatic *O*-demethylation and thereby establish the most efficient strategy for biological funneling.

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.

Thermodynamic analysis of *C. thermocellum* glycolysis using deuterated water (^2H_2O) during high substrate loading fermentations

Melanie M Callaghan*[1,2] (callaghan2@wisc.edu), Tyler B Jacobson[1,2], David M Stevenson[1,2], Daniel G Olson[1,3], Lee R lynd[1,3], Daniel Amador-Noguez[1,2], and **Gerald A. Tuskan**[2]

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

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 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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Clostridium thermocellum is a highly efficient cellulolytic anaerobe bacterium for use in CBP of biomass which can be metabolically engineered to produce C_2 and C_4 alcohols. We continue to work at improving titer, yield and rate for these conversions. Thermodynamics constitutes a key determinant of flux and enzyme efficiency in metabolic networks. A biochemical reaction with a strong thermodynamic driving force will achieve a higher net flux given a fixed amount of enzyme than one closer to equilibrium. Within a pathway, steps closer to equilibrium will be the least enzyme efficient. Thermodynamic analysis can therefore provide unique insights in synthetic pathway design by identifying bottlenecks, pinpointing the enzymes for which changes in activity will have the largest effect on flux, and predicting the most efficient route for product synthesis. Previously, we used ^{2}H-glucose and ^{13}C-glucose as isotope tracers to investigate the *in vivo* reversibility and thermodynamics of the central metabolic networks of *C. thermocellum*, *T. saccharolyticum,* and anaerobically grown *Escherichia coli*. We found that the glycolytic pathway in *C. thermocellum* operates remarkably close to thermodynamic equilibrium, with an overall drop in Gibbs free energy 5-fold lower than that of *T. saccharolyticum* or anaerobically grown *Escherichia coli* [1, 2]. We now hypothesize that the limited thermodynamic driving force of glycolysis in *C. thermocellum* limits ethanol titers in high substrate loading fermentations. To test this hypothesis, we are developing the use of deuterated water (^2H_2O) as a cost-efficient tracer to measure how the thermodynamics of *C. thermocellum*'s glycolytic and fermentative pathways change dynamically during high substrate loading fermentations. Here, we present the initial results of this novel isotope-tracer approach. This work will aid in the construction of accurate metabolic models that incorporate thermodynamic constraints, identify potential bottlenecks, and guide fast rational engineering of microbial networks.

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

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 Resolution Analysis of Recombination Rates in *Populus trichocarpa*

Chanaka Roshan Abeyratne[1], David Macaya-Sanz[1], Gancho Slavov[2], Lee Gunter[3], Kathleen Haiby[4], Richard Shuren[4], Jay Chen[3], Daniel Jacobson[3], Brian Stanton[4], Stephen DiFazio,[1] (Stephen.DiFazio@mail.wvu.edu) and **Gerald A. Tuskan**[3]

[1]West Virginia University, Morgantown; [2]Scion, Rotorua, NZ; [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; and [4]GreenWood Resources Inc., Portland, OR

http://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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

As part of the CBI goals, we have developed *Populus trichocarpa* as a model organism for enhancing feedstock characteristics. A major thrust of the project is to mine natural genetic variation in feedstock-relevant phenotypes using quantitative trait locus mapping in structured and unstructured populations. Accomplishing this goal requires a fine-scale understanding of meiotic recombination, which is one of the fundamental factors determining the outcome of natural and artificial selection. Here we describe an intensive mapping effort using whole genome resequencing of 829 progeny from 49 families to produce genetic maps representing recombination in fourteen parents. We show how recombination varies within and among genomes and between the sexes. This information is essential for understanding patterns of adaptive variation in populations as well as implementation of advanced marker-aided breeding approaches.

High resolution genetic maps for obligate outcrossing perennial trees such as *Populus trichocarpa* (black cottonwood) have been limited due to the high cost of genotyping and the challenge of establishing mapping populations that span multiple inbred generations in long-lived outcrossing trees (Yin, DiFazio, Gunter, Riemenschneider, & Tuskan, 2004). Most existing genetic maps for *Populu*s have been constructed for two parents using their full-sibs or using maternal half-sibs for a limited number of fully informative molecular markers. Our work represents the first attempt at producing multiple linkage maps for both sexes with half-sib families in *P. trichocarpa*. We present seven genetic linkage maps for each sex and all 19 haploid linkage groups that allowed us to ask new questions related to sexual dimorphism as well as intra-specific variation in single generation cross-over (CO) rates across the genome.

We have re-sequenced the genomes of 49 families (N = 829 offspring), corresponding to a full factorial cross of seven females and males each. We have called biallelic SNP/INDELs using

GATK4. Using the pedigree information, we have revealed patterns of segregation in the genome and selected high-quality markers under Mendelian segregation. We have phased and imputed the progeny genomes and recovered the gametic haplotypes of the fourteen parents, allowing us to demarcate CO regions to a finer genomic scale. Using the R package Onemap (Margarido, Souza, & Garcia, 2007), we produced fourteen framework genetic maps that contain 1820 SNP/INDEL markers on average and cover 376.7 Mb of physical length across 19 chromosomes in the Stettler-14 male reference assembly. Resulting marker order is highly collinear with the physical position for all cases other than in LG-XVII for two of the male parents. Interestingly however, we observed considerable variation among individuals for observed recombination rates leading to an analysis of other covariates associated with this variation.

We observed a total of 38,846 CO events within 1658 gametes (829 offspring x 2 haplotypes of each parent) averaging approximately 1.2 CO events per linkage group. We observed male-biased CO rates in *P. trichocarpa* with female and male half-sib families contributing 18,355 and 20,491 COs respectively. Furthermore, we observed a chromosome scale interaction in several linkage groups that attenuated the high CO rate observed in males. However, statistical evidence for a female bias was not observed in any of the linkage groups. We used wavelet analysis to circumvent the issue of having to pre-select a window size to analyze sex-based differences in CO counts. This revealed a chromosome dependent pattern of sex differences at different resolutions.

We present evidence for 125 CO hotspot regions spread across the whole genome. Interestingly, there was minimal overlap of hotspot regions between males and females. We employed the MEME suite of software to identify DNA motif enrichment in 2,054 CO regions that were demarcated to a genomic region of 10 kbps or less and resulting motifs ranking within the top three were all A/T-rich repeat sequences. We identified several other motifs that showed striking similarity to DNA sequence motifs associated with CO's in maize, *Arabidopsis* and tomato as well (Zelkowski, Olson, Wang, & Pawlowski, 2019).

This analysis revealed the genome-wide recombination landscape and patterns of inheritance at a finer scale than has previously been possible in most plants. In *P. trichocarpa* this information could be used to identify Quantitative Trait Loci (QTL) for commercially important phenotypes and along with recombination rate estimates can be used to improve genomic prediction models. Given that *P. trichocarpa* is identified as a promising renewable feedstock for bioenergy and bioproducts, we believe that our findings will be useful for ongoing efforts to accelerate domestication of this and other feedstocks, as well as future studies that investigate broader questions such as evolutionary history, perennial development related to phenology, wood formation, vegetative propagation, and dioecy that cannot be studied using conventional plant model systems such as *Arabidopsis*, rice, or maize.

Margarido, G. R. A., Souza, A. P., & Garcia, A. A. F. (2007). OneMap: Software for genetic mapping in outcrossing species. *Hereditas*, *144*, 78–79.

Yin, T. M., DiFazio, S. P., Gunter, L. E., Riemenschneider, D., & Tuskan, G. A. (2004). Large-scale heterospecific segregation distortion in Populus revealed by a dense genetic map. *Theoretical and Applied Genetics*, *109*(3):451-463.

Zelkowski, M., Olson, M. A., Wang, M., & Pawlowski, W. (2019). Diversity and Determinants of Meiotic Recombination Landscapes. *Trends in Genetics*, *35*(5), 359–370.

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], 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 have not yet been published in this organism. 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*. The CRISPRi system was then used to accumulate more 4.75 times more beta-ketoadipate in shake flask by inducing the knockdown of *pcaIJ*. This initial study demonstrating the usefulness of CRISPRi has laid the foundation for the study of a guide RNA (gRNA) library, constructed through a collaboration with JGI, targeting every coding region in the KT2440 genome and pilot experiments are showing promise that all 87,000 gRNAs will be able to be studied with high resolution.

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.

Simulating landscape-scale impacts of switchgrass nitrogen use efficiency

John L. Field*[1] (John.L.Field@colostate.edu), Michael Udvardi,[2] Keith Paustian,[1] and **Gerald A. Tuskan**[3]

[1]Colorado State University, Fort Collins; [2]Noble Research Institute, Ardmore, Oklahoma; and [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN

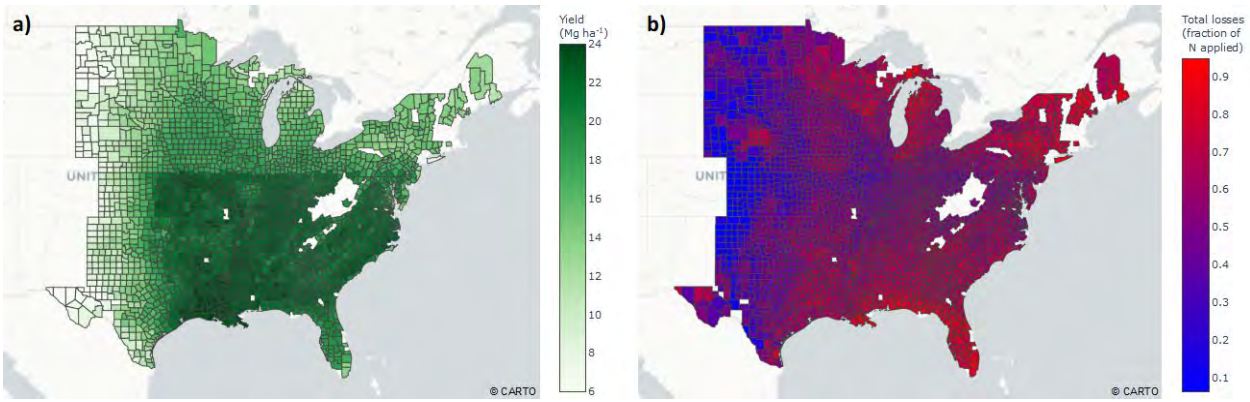
http://cbi.ornl.gov

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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Abstract: Biofuels and bioenergy are being actively developed to support rural development, energy independence, decarbonization, and negative emissions. However, it is unclear whether such technologies can be widely deployed at economically and climatically relevant scales without incurring nitrogen (N) pollution in excess of biogeochemical planetary boundaries (1). Dedicated perennial bioenergy crops such as switchgrass tend to utilize applied N more effectively than annual crops due to their well-developed root zones and ability to translocate N and other nutrients belowground into storage at senescence. Targeting production of such crops on marginal land minimizes costs and impacts to conventional agriculture but may imply greater N fertilizer use (2) and higher losses (e.g., leaching) due to the less favorable soils and climates in such areas.

In this work we explore switchgrass yields and N dynamics using the process-based DayCent biogeochemistry model, which simulates the cycling of carbon, water, and N (including mobilization/immobilization, nitrification/denitrification, and related processes) in agro-ecosystems. DayCent has been previously calibrated to represent commercial switchgrass using data from various field trials across the US, including several on marginal lands (3). Here we use that calibrated model to simulate switchgrass production on abandoned croplands across the eastern US (4). In addition to biomass yield, we track key elements of the system N balance (including major loss pathways such as volatilization, denitrification, and leaching) and calculate simulated nitrogen use efficiency (NUE) metrics. N losses from switchgrass cultivation on marginal lands are expected to vary widely as a function of soil texture and climate, suggesting

the need for careful articulation of regionally-specific N application guidelines. Finally, we will present sensitivity analysis results meant to inform breeding efforts and predict how future advanced switchgrass varieties with reduced N requirements and improved NUE might perform across these heterogeneous real-world agricultural landscapes.



DayCent-simulated a) switchgrass yields, and b) fraction of applied N fertilizer lost through volatilization, leaching, and denitrification. These simulations assume non-irrigated switchgrass cultivation on abandoned croplands, fertilized with 75 kg N ha^{-1} annually, and with lowland ecotypes grown below 40°N latitude and upland above.

References/Publications

1. Heck, V., Gerten, D., Lucht, W., & Popp, A. (2018). Biomass-based negative emissions difficult to reconcile with planetary boundaries. Nature Climate Change, 8(2), 151–155.
2. Mooney, D. F., Roberts, R. K., English, B. C., Tyler, D. D., & Larson, J. A. (2009). Yield and Breakeven Price of 'Alamo' Switchgrass for Biofuels in Tennessee. Agronomy Journal, 101(5), 1234–1242.
3. Field, J. L., Marx, E., Easter, M., Adler, P. R., & Paustian, K. (2016). Ecosystem model parameterization and adaptation for sustainable cellulosic biofuel landscape design. GCB Bioenergy, 8(6), 1106–1123
4. Zumkehr, A., & Campbell, J. E. (2013). Historical U.S. cropland areas and the potential for bioenergy production on abandoned croplands. Environmental Science & Technology, 47(8), 3840–3847.

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.

Disc Milling of Fermented Corn Stover to Increase its Accessibility to Fermentation by *Clostridium thermocellum*

Sanchari Ghosh*[1] (sanchari.th@dartmouth.edu), Evert K. Holwerda[1], Thomas Pschorn[2], Samarthya Bhagia[3], Arthur J. Ragauskas[4], Lee R. Lynd[1], and **Gerald A. Tuskan**

[1]Dartmouth College, Hanover NH; [2]TP Consulting 10117510 Inc., Canada; [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge TN; [4]University of Tennessee, Knoxville, 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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

We are investigating mechanical disruption during fermentation (cotreatment) as an alternative to thermochemical pretreatment to increase the accessibility of lignocellulose (i.e., from poplar or switchgrass) to biological attack. Our overall goal is to explore and test the hypothesis that cotreatment can be an industrially feasible method to enhance carbohydrate solubilization by engineered thermophiles in a consolidated bioprocessing (CBP) configuration. The feasibility of such "C-CBP" requires that a) carbohydrate solubilization be meaningfully increased, b) microorganisms be able to actively ferment in the presence of milling, and c) energy requirements for milling to be sufficiently low

Mechanical disruption during fermentation, cotreatment[1], combined with consolidated bioprocessing offers documented potential for disruptive potential in the cost of lignocellulose conversion[2]. It has previously been shown that *Clostridium thermocellum* fermentation readily proceeds in the presence of continuous ball milling, and that such milling allows total carbohydrate solubilization in excess of 90% of theoretical to be achieved for both woody and herbaceous feedstocks without added enzymes and without thermochemical pretreatment beyond autoclaving[3,4]. Ferment-mill-ferment experiments are reported for fermentation of corn stover and switchgrass by *Clostridium thermocellum* augmented by disc milling. Process variables such as solids loading and disc rotation were investigated, and carbohydrate solubilization was documented. In this first report of the energy consumption for cotreatment, data is presented indicating that the energy requirements for cotreatment via disc milling may be comparable to those for

thermochemical pretreatment, although more remains to be done in order to substantiate this conclusion. The energy requirements for ball milling, however, appear to be prohibitively high due to internal friction.

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.

References:

1. Lynd et al. (2020) U.S. Patent 10,533,194
2. Lynd et al. (2017) Current Opinion in Biotechnol. 45, 202-211. doi:10.116/j.copbio.2017.03.00
3. Holwerda et al. (2019) Biotechnol. Biofuels 12(1). doi: 10.1186/s13068-019-1353-7
4. Balch et al. (2017) Energy Environ. Sci, 10(5), 1252-1261. doi: 10.1039/C6EE03748H

Exploring Catalytic Conditions for C-C Bond Cleavage of Lignin-Based Compounds

Nina Gu*[1,2] (mailto:mnina.gu@nrel.gov), Sean Woodworth[1,2] Rui Katahira[1,2] Gregg Beckham[1,2] and **Gerald A. Tuskan**[1,2]

[1]National Renewable Energy Laboratory, Golden, CO; [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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Lignin depolymerization is a highly desirable approach to generate feedstocks of aromatic monomers, but a significant challenge to efficient depolymerization is the presence of recalcitrant carbon-carbon linkages. The goal of this task is to develop catalytic conditions suitable for carbon-carbon bond cleavage of lignin-derived oligomers to increase the yield of aromatic monomers for downstream processes. High-yielding lignin depolymerization methods would facilitate the conversion of lignin into value-added products.

Efforts to carry out oxidative strategies for carbon-carbon bond cleavage are complicated by the prevalence of phenolic motifs in lignin streams such as RCF oil. One major barrier arising from the presence of phenols is that its antioxidant properties (due to facile hydrogen atom donation) can inhibit radical chain oxidation mechanisms. Additionally, the formation of reactive phenoxyl radicals under oxidative conditions can readily lead to the generation of undesired polymeric tars. The approach utilized in this work to carry out phenol-tolerant oxidation is to employ transition metal catalysts that do not operate via radical chain mechanisms. Based on literature reports of benzylic oxidation of 4-Me-phenols (and related substituted derivatives) to the corresponding aldehyde, the use of such conditions are explored for the aerobic oxidation of RCF oil model compounds. In particular, substrate models featuring a bibenzyl backbone akin to β-1 and β-5 linkages are utilized, in addition to propylphenol compounds as models of β-β linkages. Efforts to couple this oxidation chemistry with various carbon-carbon cleavage methods are discussed herein.

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.

Field experiments of switchgrass GWAS population for increased sustainability: nitrogen-use efficiency and rust disease resistance

Yaping Xu[1,2], Cristiano Piasecki[1,2], Ben Wolfe[1,2], Reginald J. Millwood[1,2], Mitra Mazarei*[1,2] (mmazarei@utk.edu), C. Neal Stewart Jr.[1,2], and **Gerald A. Tuskan**[2]

[1]Department of Plant Sciences, University of Tennessee, Knoxville, TN; [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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Switchgrass (*Panicum virgatum* L.) is a perennial C4 grass that is widely considered as a leading candidate for bioenergy production. Its natural traits, including high biomass production, wide adaptation, and low agronomic input requirements, make it a highly desirable bioenergy feedstock. Increased productivity and sustainability of plant feedstocks in bioenergy crops are key factors for biofuel production. Factors affecting switchgrass quality and performance can be broadly attributed to plant genetics and the growing environment, signifying the importance of performing field studies for successful establishment and subsequent sustainability of feedstocks. To this end, a switchgrass genome-wide association study (GWAS) common garden was established in Knoxville, Tennessee containing 330 diverse accessions. A total of four replicated plots (two replicates per low and moderate nitrogen fertility treatment) were established for use in a nitrogen-use efficiency (NUE) study. The switchgrass GWAS field was evaluated manually for various phenotypic and sustainability traits including plant height, dry biomass production, NUE, and progression of rust disease over the growing season. Our results showed wide variation among the GWAS panel. Subsequently, top GWAS accessions representative of high performing genotypes (e.g., with $\geq$ 2.9 m plant height, $\geq$ 4 kg dry biomass/plant, and $\leq$ 35% of rust disease severity) were selected. Furthermore, we developed a drone-based remote sensing modeling approach for switchgrass high-throughput phenotyping and biomass yield in the field. Here, we expanded our study to develop this system for automated assessment of sustainability traits of the field-grown switchgrass. For that, several drone flights over the GWAS field during the growing season were performed. Drone-based analysis was initiated using a multispectral camera and various predictive algorithms to estimate greenness/chlorophyll (an indicator for nitrogen status in plant as well as general health of plant) through photogrammetric image analyses. Our established GWAS field

enables identification of superior genotypes and candidate genes associated with high biomass yield and sustainability in long-lived perennial 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.

Using Iterative Random Forest to Predict the Progeny of Crosses in *Populus trichocarpa*

Jonathon Romero*,[1,2] (romerojc@ornl.gov), David Kainer,[1], Ashley Cliff,[1,2], Daniel Jacobson,[1,2] and **Gerald A. Tuskan**[1]

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN; [2]Bredesen Center, University of Tennessee, Knoxville, 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 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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

The goal of this task was to develop a method that uses Explainable Artificial Intelligence (X-AI) to assist in the Genomic Selection of *Populus trichocarpa* thus improving desired positive phenotypes such as yield and pathogen resistance in a manner that takes fewer generations than traditional Genomic Selection techniques.

Iterative Random Forest (iRF) [1,2] is a machine learning model with multiple uses in the field of Genomics. We have shown that iRF, when used in conjunction with *in silico* breeding software (SBVB) [3], can accurately predict the real outcome of crossing two parents together by training the model on a GWAS population that does not contain the parents. This method takes advantage of the data collected from a GWAS population, but by using iRF, the method is able to incorporate non-additive effects that it detects in the population to aid in model accuracy. Once the iRF model has been trained for the phenotype of interest, the sequenced genomes of potential parents are synthetically crossed to produce virtual full-sibling families. The virtual genomes of each progeny are then used as input to the iRF model to predict the distribution of the phenotypes expected from the progeny of each parental cross. The rank order of these families proved to be a highly accurate representation of the rankings of the actual families they represent when grown in a common garden. This method selected crosses from a large diversity of parents instead of from only the top ranked males and females that a traditional Genomic Selection technique would use. This achieves an important goal in genomic selection, which is to keep as much genetic diversity as possible

while improving a selected phenotype. This method has been validated in *Populus trichocarpa* by accurately predicting the ranking of height of *in silico* progeny and comparing to real progeny crossed from the same parents.

References:

1. Sumanta Basu, Karl Kumbier, James B. Brown, Bin Yu, Iterative random forests, *Proceedings of the National Academy of Sciences* Feb 2018, 115 (8) 1943-1948; DOI: 10.1073/pnas.1711236115
2. Cliff A, Romero J, Kainer D, Walker A, Furches A, Jacobson D. A High-Performance Computing Implementation of Iterative Random Forest for the Creation of Predictive Expression Networks. *Genes* (Basel). 2019 Dec 2;10(12):996. doi:10.3390/genes10120996. PMID: 31810264; PMCID: PMC6947651.
3. Pérez-Enciso M, Forneris N, de Los Campos G, Legarra A "Evaluating Sequence-Based Genomic Prediction with an Efficient New Simulator." *Genetics* 2017; 205(2):939-953 https://doi.org/10.1534/genetics.116.194878

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.

Identifying growth-coupled genetic loci using integrated metabolic modeling in populus

Debolina Sarkar,*[1] (debolina@psu.edu), Jin Zhang[2], Kai Feng[2], David Kainer[2], Timothy Tschaplinski[2], Dan Jacobson[2], Wellington Muchero[2], Costas D. Maranas[1] and **Gerald A. Tuskan**[2]

[1]Department of Chemical Engineering, Pennsylvania State University, University Park, PA; [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 will address 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 specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Designing feedstock transgenics with a combination of desired characteristics has been a focus of plant studies aimed at improving growth yield, feedstock composition, and nutrient-use efficiency. Although a small fraction of commercially available genetically modified crops present 'stacked' traits, only a handful of products have been developed that modify two or more genes. In this work, we identify enzymatic genes and their optimal combinations that accord a growth benefit in *P. trichocarpa* by integrated metabolic modeling. Previously, a large number of single nucleotide polymorphisms (SNP) have been identified in populations of *Populus trichocarpa* [1]. A previously developed SNP role assignment tool ('**SNPeffect**'[2]) was first used to identify functional SNPs and their effect on pathway flux. SNPeffect constructs mechanistic genotype-to-phenotype relationships by integrating the biochemical knowledge encoded in metabolic models with all available omics, phenotype, and sequence data. Inferred SNP activities were used to constrain corresponding gene activities, and then predict growth-coupled genetic interventions. As expected, our results indicate that growth is a complex polygenic trait governed by carbon and energy partitioning. The predicted intervention sets are associated with experimentally characterized growth-determining genes and also suggest putative ones. Predicted genes belong to pathways such as amino-acid metabolism, nucleotide biosynthesis, and cell wall and lignin biosynthesis, in line with breeding strategies that target pathways governing carbon allocation, and cellular ATP and reductant production. We envision this framework to be used in future for targeting additional agronomically-relevant traits such as robust growth yield, nutrient-use efficiency, and poplar feedstock composition. Beyond the design of stacked transgenics, we expected this work to have extensions in genetically assisted breeding of improved non-transgenic lines. This work contributes to the understanding of the molecular mechanisms governing growth and plant productivity, in order to further exploit them and develop a sustainable bioenergy feedstock.

References:

1. Tuskan, G. A., Muchero, W., Tschaplinski, T. J. & Ragauskas, A. J. Population-level approaches reveal novel aspects of lignin biosynthesis, content, composition and structure. *Current Opinion in Biotechnology* vol. 56 250–257 (2019).
2. Sarkar, D. & Maranas, C. D. SNPeffect: identifying functional roles of SNPs using metabolic networks. *Plant J.* **103**, 512–531 (2020).

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.

Support for the Poplar GWAS dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER) via the Bioenergy Science Center (BESC) under Contract No. DE-PS02-06ER64304. The Poplar GWAS Project used resources of the Oak Ridge Leadership Computing Facility and the Compute and Data Environment for Science at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Systems Metabolic Engineering of *Clostridium thermocellum* for Direct Conversion of Cellulosic Biomass to Designer C4-derived Esters

Hyeongmin Seo*[1,3], (hseo5@vols.utk.edu) Jong-Won Lee[1,2], Sergio Garcia[1,3], Niantai Li[1,3], Cong T. Trinh,[1,2,3] and **Gerald A. Tuskan**[3]

[1]Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN, [2]Bredesen Center for Interdisciplinary Research and Graduate Education, The University of Tennessee, Knoxville, TN, [3]Center of Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

The C4-derived esters (e.g., isobutyl acetate, butyl acetate, isobutyl isobutyrate) are industrially important chemicals with versatile applications as fuels, flavors, fragrances, and solvents. Biologically, these esters are synthesized by condensation of either (iso)butanol with acyl-CoAs or alcohols with (iso)butyl-CoA using alcohol acyltransferases (AATs). Consolidated bioprocessing using efficient and robust biocatalysts, capable of direct conversion of lignocellulosic biomass to esters at high efficiency, offers a viable solution to sustainable bioeconomy. While the cellulolytic, thermophilic *Clostridium thermocellum* is a promising CBP microorganism due to its robust metabolism to degrade lignocellulosic biomass, it cannot natively produce esters. The critical challenges for optimal ester biosynthesis in *C. thermocellum* are to rewire its metabolism and harness enzymes and pathways that are efficient, robust, and compatible with the host. The aim of our project is to enable systems metabolic engineering of *C. thermocellum* for direct conversion of cellulosic biomass to designer C4-derived esters. Through bioprospecting and model-guided protein engineering, we first repurposed chloramphenicol acetyl transferases (CATs) to function as AATs that are efficient, robust, and compatible with *C. thermocellum* for designer bioester synthesis.[1-2] To eliminate the endogenous ester degradation, we next performed genome mining and enzyme characterization to identify and disrupt two critical carbohydrate esterases (CEs) from the genome of *C. thermocellum* while not affecting its robust capability of cellulose degradation.[3] In combination with gene expression and fermentation optimization, we demonstrated the first-generation of engineered *C. thermocellum* capable of producing 1 g/L and 0.34 g/L of C4-derived esters with and without isobutanol supplemented cellulose fermentation, respectively. To enable biosystem designs of *C. thermocellum*, we have built the genome-scale metabolic model[4-6] and used it to design the modular (chassis) cells that can be assembled with exchangeable production modules for optimal

biosynthesis of designer bioesters with minimal requirement of strain optimization cycles.[7-11] Future studies are to characterize and elucidate the modular design of *C. thermocellum* for the target ester biosynthesis.

References/Publications

1. Seo, H.; Lee, J.-W.; Garcia, S.; Trinh, C. T., Single mutation at a highly conserved region of chloramphenicol acetyltransferase enables isobutyl acetate production directly from cellulose by Clostridium thermocellum at elevated temperatures. *Biotechnology for biofuels* **2019,** *12* (1), 245.
2. Seo, H.; Lee, J.-W.; Giannone, R. J.; Dunlap, N. J.; Trinh, C. T., Repurposing chloramphenicol acetyltransferase for a robust and efficient designer ester biosynthesis platform. *bioRxiv* **2020**, 2020.11.04.368696.
3. Seo, H.; Nicely, P. N.; Trinh, C. T., Endogenous carbohydrate esterases of Clostridium thermocellum are identified and disrupted for enhanced isobutyl acetate production from cellulose. *Biotechnology and Bioengineering* **2020,** *117*, 2223-2236.
4. Thompson, R. A., Layton, D.S., Guss, A.M., Olson, D.G., Lynd, L.R., Trinh, C.T., Elucidating Central Metabolic Redox Obstacles Hindering Ethanol Production in *Clostridium thermocellum*. *Metab Eng* **2015,** *32*, 207-219.
5. Thompson, R. A.; Dahal, S.; Garcia, S.; Nookaew, I.; Trinh, C. T., Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of Clostridium thermocellum DSM 1313 implementing an adjustable cellulosome. *Biotechnology for Biofuels* **2016,** *9* (1), 194.
6. Thompson, R. A.; Trinh, C. T., Linking overflow metabolism and growth cessation in Clostridium thermocellum DSM1313 during high cellulose loading fermentations. *Biotechnology and bioengineering* **2017,** *114* (11), 2592-2604.
7. Trinh, C. T., Liu, Y., Conner, D, Rational Design of Efficient Modular Cells. *Metab Eng* **2015,** *32*, 220-231.
8. Garcia, S.; Trinh, C. T., Multiobjective strain design: A framework for modular cell engineering. *Metabolic engineering* **2019,** *51*, 110-120.
9. Garcia, S.; Trinh, C. T., Comparison of Multi-objective Evolutionary Algorithms to Solve the Modular Cell Design Problem for Novel Biocatalysis. *Processes* **2019,** *7* (6), 361.
10. Garcia, S.; Trinh, C. T., Harnessing natural modularity of metabolism with goal attainment optimization to design a modular chassis cell for production of diverse chemicals. *ACS Synthetic Biology* **2020,** *9* (7), 1665–1681.
11. Garcia, S.; Thompson, A. R.; Giannone, R. J.; Dash, S.; Maranas, C.; Trinh, C. T., Development of a genome-scale metabolic model of Clostridium thermocellum and its applications for integration of multi-omics datasets and computational strain design. *Frontiers Bioeng Biotechnol* **2020,** *8* (772), 1-16.

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 neat poplar lignin oil to jet-range aromatic hydrocarbons with molybdenum carbide

Michael L. Stone* (mlstone@mit.edu),[1,3] William P. Mounfield III,[1,3] Ana Morais,[2,3] Earl Christensen,[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 Renewable Energy Laboratory, National Bioenergy Center, Golden CO; [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, 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 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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

The aviation industry requires sustainable aviation fuels (SAF) capable of reducing greenhouse gas emissions while satisfying strict safety and quality standards. Lignin is a promising renewable feedstock for the production of aromatic hydrocarbons, the missing fraction needed to achieve 100% SAF. The use of lignin in SAF hinges on reducing oxygen content while limiting ring hydrogenation and maximizing yields of C8-C20 hydrocarbons. Herein, we utilize molybdenum carbide (Mo_2C) catalysts to hydrodeoxygenate lignin oil produced via reductive catalytic fractionation (RCF) of untreated poplar. We designed a 3-phase trickle-bed reactor that generates steady-state partial-conversion kinetic data to analyze catalyst activity while deoxygenating complex lignin feeds, concluding that surface oxidation is the key catalyst limitation while processing neat lignin oil. At 350°C, complete deoxygenation is achieved in a single-pass at steady-state, with 94.2% selectivity of monomeric products to propylbenzene and methylpropylbenzene and 70.8 C-mol% recovery of whole oil. While achieving high monomer recovery, single-pass reactions at 350°C have low recovery of dimers and larger oligomers. We hypothesize this is due to high reactivity of oxygenated oligomers toward unwanted side reactions at 350°C. To recover larger oligomers, a multi-pass reaction was performed in which 50% of oxygen was removed at 300°C, an additional 25% oxygen was removed at 325°C and the remaining oxygen was removed in a third pass at 350°C. This multi-pass experiment resulted in an oil containing 49.5 wt.% 1-ring aromatics, 25.6 wt.% 2-ring aromatics, and 14.1 wt.%

cycloalkanes. Deoxygenation of neat RCF oil corresponded to an increase in carbon content from 65.4 to 88.7 mass %, a decrease in oxygen content from 26.8 to 0.7 mass %, an increase of lower heating value from 21.73 to 39.99 MJ/kg, and a decrease in viscosity at 40°C from 231 to 1.04 cP.

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.

Title: A Rapid Semi-Automated Phenotyping System to Capture the Highly Useful Diameter at Breast Height In *Populus trichocarpa*

Authors: Jared Streich*, (ju0@ornl.gov), Hari Chhetri,[3] Mirko Pavicic,[4] Daniel Jacobson,

and **Gerald Tuskan**

Institutions: [1]Department of Biosciences, Oak Ridge National Laboratory, Oak Ridge TN

Website: https://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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Abstract text: The goal of this project is to reduce the complexity and difficulties associated with capturing accurate precise phenotypes (traits) and encoding the resulting data into easily useable formats for bioinformatic analysis. Here we discuss the implementation of a novel scientific instrument that logs a sample identity via a radio frequency tag on the specimen and records image information. This novel device and more like it will reduce human error and rapidly increase phenotyping efforts.

Diameter at breast height (DBH) is an important phenotype in tree breeding, often correlating with many other phenotypic traits such as height and biomass which are important for biofuel production. In the post genomic revolution, phenotyping has become a major bottleneck in scientific research. Manual measurements are prone to human error as is recording or transcribing to electronic records. Therefore, we have developed a novel hand-held semi-automated DBH camera that captures sample IDs through encoded RFID tags and rapid highly calibrated time-stamped measurements through image processing.

We assessed how this instrument compared to manual phenotyping in a common garden in East Tennessee. Field studies often have many hundreds to thousands of samples, for example, some CBI field sites have up to 4,500 trees. As such, scalable, accurate phenotyping is very important. Speed trials of two people with a manual tape measure and manual clipboard measured 14 trees in 20 minutes, where one person with the DBH Camera captured 60 tree DBH measurements in the same amount of time, or 180 trees in 60 minutes. This 4.3x increase in efficiency means two

people can phenotype 4,500 trees in 12.5 hours (≈25 person hours) with minimal human error, versus 107.14 hours using two people or 214.28 person hours. Thus, the new device requires only 23.34% of the original labor effort, not including the arduous manual typing of sample names and measurements. In contrast, the DBH camera data is automatically captured and produces formats common for analysis software making it near instantly accessible. Thus, such handheld automated phenotyping devices enable the same number of people to capture a wider array of phenotypes with more accuracy and precision in smaller time-windows. This reduces the associated costs for fieldwork campaigns meaning more traits can be used for crop improvement methods.

Funding statement: 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.

Engineering Plant Cell Wall Polysaccharide *O*-Acetyltransferases with Altered Specificity

Hsin-Tzu Wang*[1,2] (hw19911@uga.edu), Vivek S. Bharadwaj[3], Erica T. Prates[4], Manesh Shah[4], Jeong-Yeh Yang[1], Digantkumar Gopaldas Chapla[1, 2], Maria J. Peña[1], Daniel A. Jacobson[4], Kelley W. Moremen[1,2], William S. York[1,2], Yannick J. Bomble[3], Breeanna R. Urbanowicz[1,2], and **Gerald A Tuskan**[4]

[1]Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602; [2]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602; [3]Bioscience Center, National Renewable Energy Laboratory, Golden, Colorado 80401; [4]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Biosciences Division, Oak Ridge, Tennessee 37831

cbi.ornl.gov

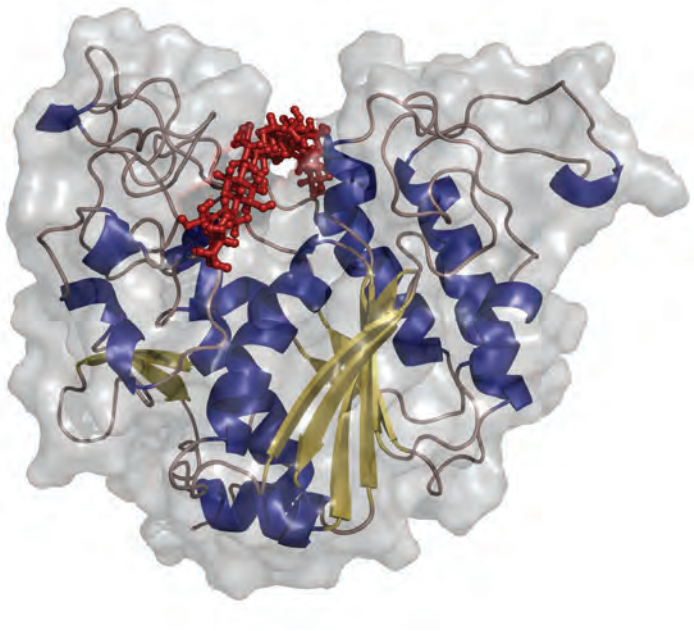
The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI has identified key barriers for the current bioeconomy in (1) high-yielding, robust 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.

We aim to modify lignocellulosic feedstocks with adjusted acetylation levels towards reducing cell wall recalcitrance via the understanding and control of key acetylation enzymes. *O*-Acetylation is a universal modification of non-cellulosic polysaccharides in most bioenergy feedstocks, which is an obstacle for saccharification and fermentation in the process of biomass conversion into bioproducts. Our goal in this project is to understand the molecular basis of the catalytic activity of *O*-acetyltransferases from plants that are used as pilot organisms in CBI, such as *Populus* and *Arabidopsis*.

The level of acetylation of the hemicelluloses that constitute plant cell walls can have significant impacts on lignocellulosic biomass-based biofuel production. *O*-Acetyl groups play a major role in polymer-polymer interactions in the cell wall, such as those formed between xylan-lignin and xylan-cellulose, and also sterically hinder hydrolases from accessing their polysaccharide substrates. In addition, *O*-acetyl groups are one of the most abundant substituents of the plant cell wall components, particularly xylans and mannans; thus, the acetate released during both pretreatment and deconstruction also inhibits downstream microbial fermentation processes. The enzymes involved in *O*-acetylation of hemicelluloses belong to a plant-specific Trichome Birefringence-Like (TBL) protein family. Understanding the interactions between TBL proteins

and their substrates could drive the development of engineered enzymes and modified feedstocks for improved conversion into fuels and products. In our research, we used molecular docking and molecular dynamics simulations to analyze enzyme-substrate interactions of the *Arabidopsis thaliana* TBL protein xylan *O*-acetyltransferase 1 (XOAT1/TBL29) with xylan. The results point out the key amino acid residues involved in substrate-binding at the active site, and kinetic analysis of the resulting XOAT1 mutant variants substantiates their crucial role, providing us an opportunity to potentially tune acetylation levels in biomass through genetic engineering (see figure). Additionally, we employed high-throughput computational modeling of the TBL family in *Populus* to identify the molecular determinants of substrate specificity. This computational and experimental workflow suggests a new strategy of using predicted models to fine-tune the biochemical properties of the plant cell wall.

Funding of this research was provided by the Center for Bioenergy Innovation (CBI), and from the U.S. Department of Energy (DOE) Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science.



XOAT1 is at the heart of xylan synthesis. The structure of XOAT1 reveals a putative substrate binding groove shown here with a xylodecaose (red stick representation) bound in MD simulations.

Lunin, Vladimir V., et al. "Molecular mechanism of polysaccharide acetylation by the Arabidopsis xylan O-acetyltransferase XOAT1." *The Plant Cell* 32.7 (2020): 2367-2382.

Genome-wide association studies of drought stress and water use efficiency related traits in switchgrass

Yongqin Wang[1,2] (yqwang@noble.org), Hari B. Chhetri[2], Weihong Dong[1,2], David Kainer[2], Larry M. York[1,2], Malay C. Saha[1,2], Michael K. Udvardi[1,2], Daniel Jacobson[2], Yun Kang[1,2], and **Gerald A. Tuskan**[2]

[1]Noble Research Institute, LLC. Ardmore, Ok; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

Website: https://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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Abstract: Switchgrass (*Panicum virgatum*) is a promising feedstock for biofuels, but periodic drought often limits its productivity, especially in marginal lands where it is likely to be planted. To identify causal alleles controlling biomass yield and persistence under drought stress and screen for switchgrass genotypes with superior biomass production, water use efficiency (WUE) and drought tolerance, genome-wide association studies (GWAS) were performed using a switchgrass GWAS panel of 299 genotypes. Phenotyping experiments were conducted under greenhouse conditions with five biological replicates (two tillers per replicate) each for drought-stress 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 architecture. Large genotypic variations were observed in all the traits characterized. GWAS analyses have been performed on leaf area, specific leaf weight, leaf cuticular wax and well-watered shoot biomass, and multiple significant associative markers were identified. Top candidate genes and quantitative trait loci (QTL) will be further evaluated to facilitate subsequent genomic selection, maker-assisted breeding programs and biotechnology strategies to enhance sustainability in switchgrass production.

Funding statement: *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.*

Physiological responses of *Populus trichocarpa* genotypes to drought

Marvin Wright (mwright7@mix.wvu.edu)[1,3], Roshan Abeyratne[1,3], Stephen DiFazio[1,3], Jonathan Cumming,[1,2,3] and **Gerald A. Tuskan**[3]

Department of Biology, West Virginia University, Morgantown, WV[1]; Department of Natural Sciences, University of Maryland Eastern Shore, Princess Anne, MD[2]; and Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN[3]

http://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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Abstract: *Populus trichocarpa* (black cottonwood) is a species commonly used as a plantation crop for biofuel production. Although extensive variation exists in growth in response to water limitation, the physiological underpinnings of this variation remain largely unknown. We established two studies to investigate water-limitation physiology in *P. trichocarpa*: a common garden study in Boardman, OR where 60 genotypes (ranked as either high or low performers based on growth data between droughted and well-watered groups) were analyzed; and a greenhouse study in Morgantown, WV where a subset of 20 of the genotypes were evaluated. In Boardman, osmolality ($p=0.0018$), stem water potential ($p<0.0001$), and specific leaf area ($p=0.0047$) differed between droughted and well-watered groups, and there was genotypic variation in osmolality ($p=0.0002$), SPAD ($p<0.0001$), stomatal conductance ($p=0.0045$), and specific leaf area ($p=0.0131$). There were no differences, however, between the high and low performing ranked groups. In Morgantown, stem water potential ($p<0.0001$), stomatal conductance ($p<0.0001$), photosynthesis ($p<0.0001$), and water use efficiency ($p<0.0001$) all changed during the dry-down period, and there were genotypic differences in SPAD ($p<0.0001$), stem water potential ($p=0.0102$), stomatal conductance ($p=0.0005$), photosynthesis ($p<.0001$), and water use efficiency ($p=0.0002$). The high performers had higher SPAD values ($p=0.0207$), more negative stem water potentials ($p=0.0327$), lower stomatal conductance ($p=0.0262$), and reduced photosynthesis ($p<0.0001$) compared to low performers. Thus, the physiological factors that appear to establish the field performance rankings are those that allow the tree to uptake and retain as much water as possible. This means that those genotypes that can more efficiently use water when available, and minimize loss when water is limiting, will be more successful as a plantation crop species.

Funding statement: *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.*

Application of Base Editing Technology in Poplar

Guoliang Yuan*[1] (mailto:yuang@ornl.gov), Haiwei Lu[2], Md. Mahmudul Hassan[2], Tao Yao[1], Wellington Muchero[1], Jin-Gui Chen[1], Xiaohan Yang[1], and **Gerald A. Tuskan**[1]

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory; [2]Biosciences Division, Oak Ridge National Laboratory

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 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 specialty biofuels (C4 alcohols, C6 esters and hydrocarbons) using CBP at high rates, titers and yield in combination with cotreatment, pretreatment or catalytic upgrading. CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Poplar (*Populus* spp.) is an important perennial woody bioenergy crop. In CBI, we seek to develop new capabilities for genetic improvement of poplar as a sustainable source of biomass for biofuels production. To achieve this goal, a host of new technologies are indispensable for accelerating the domestication of poplar plants. Base editing that directly generates precise point mutations is one of the most recent advances in the area of genome engineering. Although base editing has been used in multiple plant species, its application in poplar has not been well established yet. Here, we applied base editing technology in hybrid poplar clone 717 (*P. tremula* × *P. alba*) and in *P. deltoides* WV94, aiming to establish a robust system for the high-precision genome engineering in poplar. Our previous genome-wide association study (GWAS) and RNA-seq in poplar identified one transcription factor that was associated with disease resistance (Muchero et al., 2018). Another transcription factor was recognized as a key regulator in response to heat stress in plants but it's role in poplar has not been studied yet (Charng et al., 2007). To validate the functions of these two transcription factors, we generated loss-of-function mutants in these poplar lines by introducing pre-mature stop codons in the protein-coding sequences using a base editor. So far, the desired point mutation has been detected in genetically modified hybrid poplar 717 plants, indicating that the base editor functioned well in that poplar line. Also, base-editing constructs have been engineered into *P. deltoides* WV94, and transgenic plants are being characterized for desired point mutations in the target genes. The impact of the mutations mediated by base-editing on phenotypic

traits (e.g., disease resistance, heat tolerance) will be evaluated in the genetically-modified poplar plants in the future.

References

Charng YY, Liu HC, Liu NY, Chi WT, Wang CN, Chang SH, Wang TT (2007) A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in Arabidopsis. Plant Physiol **143:** 251-262

Muchero W, Sondreli KL, Chen JG, Urbanowicz BR, Zhang J, Singan V, Yang Y, Brueggeman RS, Franco-Coronado J, Abraham N, Yang JY, Moremen KW, Weisberg AJ, Chang JH, Lindquist E, Barry K, Ranjan P, Jawdy S, Schmutz J, Tuskan GA, LeBoldus JM (2018) Association mapping, transcriptomics, and transient expression identify candidate genes mediating plant-pathogen interactions in a tree. Proc Natl Acad Sci U S A **115:** 11573-11578

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.

Characterizing Thiamine Diphosphate Dependent Enzymes for Promiscuous C-C Bond Formation Catalysis

Tracey Dinh[1]*(tracey.dinh@u.northwestern.edu), 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: Increasing recognition of non-canonical enzyme activity has revealed potential problems for heterologous expression; however, understanding the potential cell burden due to promiscuous enzyme activity remains a challenge. Toward this end, our team seeks to develop cheminformatics tools that predict enzyme substrate promiscuity and the resulting metabolomic consequences. Merging both experimental and computational screening approaches, this work aims to comprehensively characterize the catalytic landscape of thiamine diphosphate (ThDP)-dependent enzymes and prospect novel promiscuous transformations. We developed a high throughput activity assay to screen carboligase enzymes against diverse sets of α-keto acid substrates. Activity data from this screen will be used to build support vector machine classifiers for in silico reaction screening. Comprehensive activity landscapes developed from predictive models will provide insight into substrate selectivity and will facilitate enzyme active site engineering. Robust activity models for each carboligase enzyme will allow us to predict promiscuous activity on genome-scale models of host organisms such as *E. coli*.

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 C4 bioenergy 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 are: 1. Assembly and 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:
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, dicots). 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. This project leverages 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. 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 a overlaps between these sets established based on sequence

homology criteria. A manuscript describing the curation process and preliminary characterization of Deep Green proteins is in preparation, including 460 members from Chlamydomonas, and . Under Objective 3 (assembly of reverse genetic resources for Chlamydomonas Deep Green Proteins) we have identified pre-existing CLiP library(1) insertions for 345 mutants, and for the remaining 115 we have adapted an efficient genome editing procedure (2) that uses CRISPR-Cas9 and a barcoded selectable marker cassette to generate tagged mutants. 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 shared 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 with *de novo* structure predictions obtained using the I-TASSER software and the Rosetta Abinitio Relax module. Together, these data will help guide researchers in investigating the contribution of conserved unknown proteins to diverse aspects of photosynthetic 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:

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

Kranthi Varala[1]* (kvarala@purdue.edu), Rajeev Ranjan[1], Ying Li[1], and Karen Hudson[2]
[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 regulator network inference [1] from public RNA-Seq data (Aim 1) identified the top predicted regulators of seed lipid biosynthesis. This list included four known regulators of this process in the top 10 predicted TFs and many novel TFs that are predicted to have a strong influence on seed lipid biosynthesis. We have identified and collected mutant lines in most of the

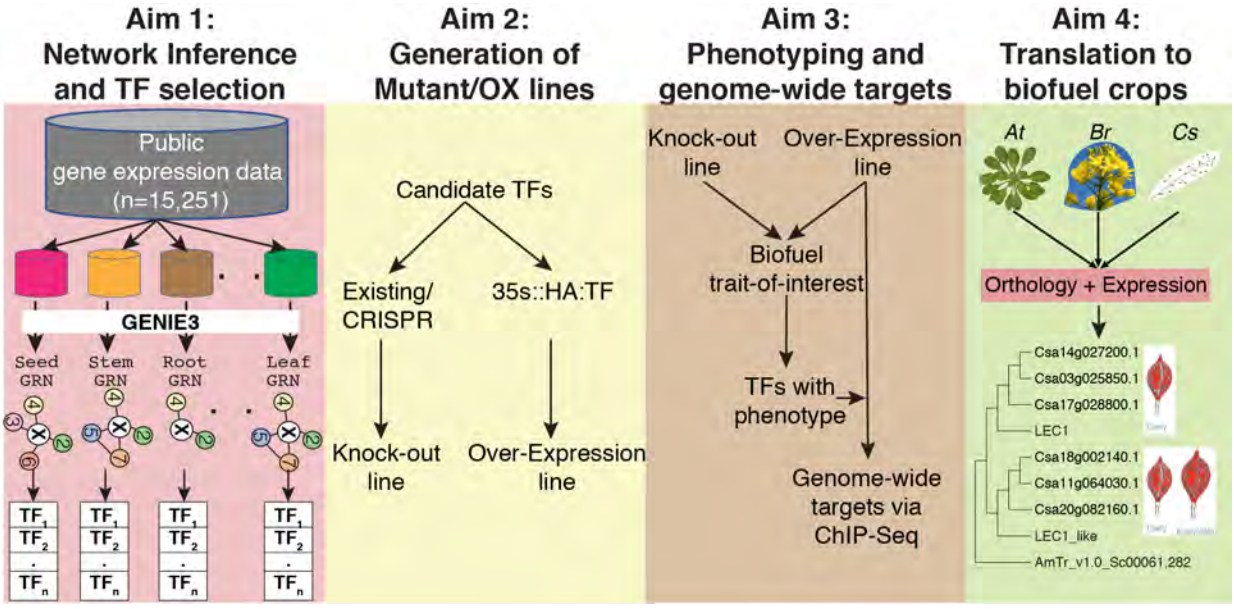


Table 1 Over-expression lines for each candidate TF. All TFs were conjugated with a HA tag to enable ChIP-Seq assays to identify global targets of TF binding.

Gene ID	Gene Name	35S pro: CDS-HA	NapA pro: CDS-HA	Native pro: Genomic-HA
At5G65640	bHLH93	Done	Done	Done
At5G65410	HB25	Done		Done
At5G04760	Div2			Done
AT5G08520	SRM1	Done	Done	Done
At3G28910	MYB30	Done	Done	Done
At2G46590	DAG2	Done		Done
At1G25330	CESTA	Done		Done
At5G65100	EIN3	Done		Done
At5G10030	TGA4	Done	Done	Done
At3G16940	CAMTA6			
At3G60030	SPL12			Done
At3G57390	AGL18			Done
At3G27785	MYB118			Done
At5G40360	MYB115	Done	Done	Done
At3G54320	WRI 1	Done		Done

novel TFs and are in the process of phenotyping them (see below) to detect changes in seed lipid profile. 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. In all three constructs the TF was tagged with a HA epitope to allow downstream ChIP-Seq studies (see Table 1). In addition to the proposed lipid profiling of mature seeds from the mutant and over-expression lines, seed size and shape were assayed and compared to the wild-type mature seeds.

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 submitted for peer review and is currently under revision.

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 (under revision)

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.

Meta-Transcriptomic Network Analysis of Inter-Species Interactions in a Cyanobacterial-Methanotroph Coculture

Ryan McClure[1], Alexander Beliaev[1*] (alex.beliaev@pnnl.gov), Pavlo Bohutskyi[1], Eric Hill[1], Marina Kalyuzhnaya[2], and **Jin Wang[3] (PI)**

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA; [2]San Diego State University, San Diego, CA; [3]Auburn University, Auburn, AL;

Project Goals: In nature, microbial communities use highly efficient mechanisms to recover energy and capture carbon from both CH_4 and CO_2 through coupling of methane oxidation and oxygenic photosynthesis. However, harnessing the metabolic potential of phototroph-methanotroph consortia for biotechnology applications requires systems-level understanding of interspecies interactions and how these interactions affect the overall consortium dynamics and output. In this study we applied meta-transcriptomic network analysis to a model *Methylomicrobium alkaliphilum* 20Z - *Cyanobacterium stanieri* HL-69 coculture to identify the potential points of metabolic interactions between phototrophic and methanotroph bacteria. We expect that the knowledge gained from this work may be generally applicable to other cross-feeding consortia, and the tools developed can be adapted to study the interactions and dynamics of other multi-organism platforms.

Abstract: To reveal the specific points of interactions between *M. alkaliphilum* 20Z and *C. stanieri* HL-69, the coculture was grown under a range of environmental conditions that included variations in temperature, pH, dissolved O_2, nutrient concentrations and energy (*i.e.*, light) inputs under steady-state and batch growth regimes. Species-specific RNA sequencing was carried out to determine the transcriptomic response of each species. Global analysis of transcriptome patterns showed that batch conditions induced a broad range of response while bioreactor conditions were much more similar. Within batch conditions, temperature and nutrient type induced the largest changes in transcriptional response in *M. alkaliphilum* 20Z, while alterations in pH and dissolved O_2 induced response specifically in *C. stanieri* HL-69, respectively.

We next pooled all transcriptomic data into a gene co-expression network to examine the interspecies edges linking *M. alkaliphilum* 20Z and *C. stanieri* HL-69 subnetworks (Fig. 1). These points of transcriptional coordination can be used to identify which pathways and processes are likely linked between the species as well as to infer instances of cross-species interaction. Within the entire coculture network comprised of 20,000 total edges (links describing coordinated expression of two transcripts), 2,393 of them were found to be interspecies (linking a gene in *M. alkaliphilum* 20Z with one in *C. stanieri* HL-69). This subnetwork included 1,182 *M. alkaliphilum* 20Z genes and 1,022 *C. stanieri* HL-69 genes. Further examination of the cross-species subnetwork revealed that genes enriched in amino acid metabolism, energy metabolism, cell structure and transport occupied positions of high centrality indicating that these processes are likely to be critical for co-culture growth. Notably, the distribution of cross-species edges in the coculture network varied significantly as a function of cultivation regimes; we found that the number of cross-species edges was ~ 4-fold higher under steady-state conditions compared to batch cultures. This suggests that coordination between species is much higher under steady state vs. batch growth conditions. We attribute this discrepancy to the variation in specific growth phases of *M. alkaliphilum* 20Z *C. stanieri* HL-69 in batch cultures as the both populations in the coculture growing in steady state was

synchronized. Functional analysis of cross-species edges enriched in a specific function showed that carbohydrate metabolism of *M. alkaliphilum* 20Z was positively correlated with transport processes in *C. stanieri* HL-69 as were amino acid metabolism and translation, respectively. These experiments are among the first to build a co-expression network linking genes across species in a phototroph-methanotroph system and the knowledge gained will be critical not only in understanding interactions specific to these two functional guilds but also in building paradigms that describe fundamental coordination of metabolism and exchange of nutrients that are common to a large number of bacterial communities involved in the cycling of CH_4 and CO_2.

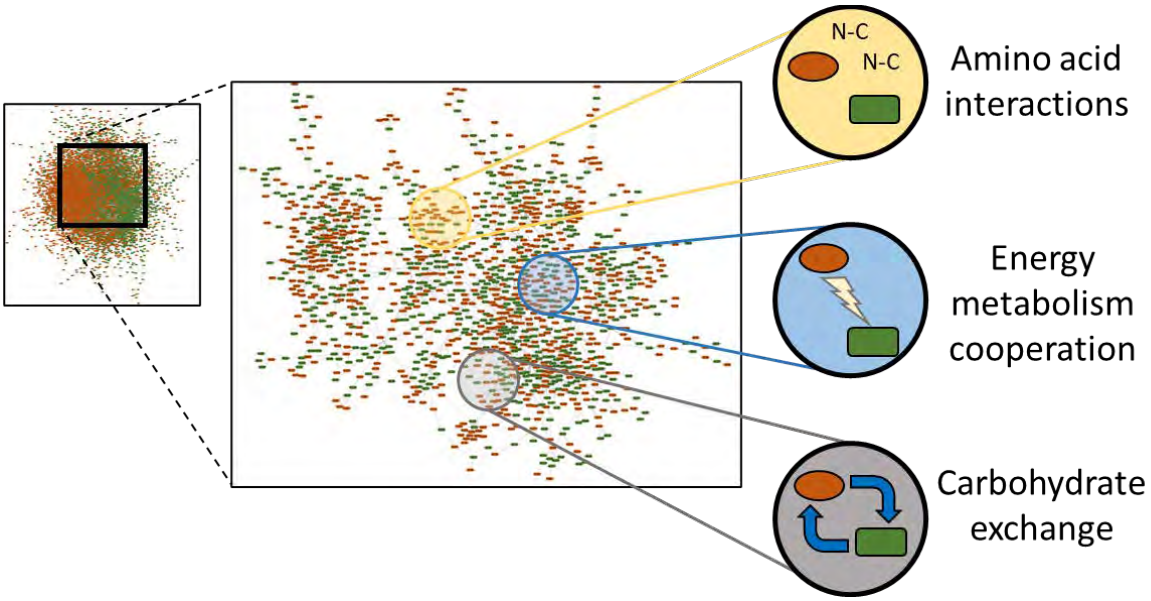


Figure 1. Meta-transcriptome network of *M. alkaliphilum* 20Z and *C. stanieri* HL-69 coculture. Green rectangles represent *C. stanieri* HL-69 genes and orange rectangles represent *M. alkaliphilum* 20Z genes. Lines indicate points of coordinated expression between genes with several pathways that are strongly linked (highlighted).

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

Harnessing Methanotroph-Photoautotroph Interactions for Biogas Conversion

Kiumars Badr[1], Matthew Hilliard[1], Q. Peter He[1], Marina Kalyuzhnaya[2], Alexander S Beliaev[3], and **Jin Wang**[1*] (wang@auburn.edu)

[1]Auburn University, Auburn, AL; [2]San Diego State University, San Diego, CA; [3]Pacific Northwest National Laboratory, Richland, WA;

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: Biogas derived from organic waste streams through anaerobic digestion has immense potential as a renewable feedstock for producing high-density liquid fuels and commodity chemicals. However, effective and economical biogas utilization has been challenging due to the presence of contaminants. In this project, we have clearly demonstrated that M-P cocultures offer a flexible platform for highly efficient biological CH_4-CO_2 co-utilization and enable significantly enhanced cell growth of both species in a model M-P coculture[2-5]. Beyond the designed interspecies exchange of in situ produced O_2 and CO_2, we hypothesize that there exist other emergent metabolic exchanges that enabled the observed enhanced coculture growth. Through designed experiments and a semi-structured kinetic modeling approach, we successfully validated our hypothesis and quantified the effect of emergent interactions within the model coculture on cell growth, without knowing what the emergent metabolites exchanges are.

Genome-Scale Metabolic Modeling of the methanotroph-photoautotroph coculture

Once the existence of the unknown emergent interactions is confirmed, the next question is how to postulate/identify the potential metabolites that are exchanged within the M-P coculture. It is very challenging to answer this question via experimental approach alone, as the key metabolite being exchanged may not be detectable – the metabolites being produced by one species could be completely and immediately consumed by the other. To address this challenge, we explore a fully structured modeling approach, i.e., genome-scale metabolic modeling for the coculture.

Genome-scale metabolic models (GEMs) represent extensive knowledge bases of microorganisms and provide a platform for model simulations and integrative analysis of different sources of data, including various omics data. In addition, it offers a convenient and powerful tool to generate and test various hypotheses regarding "metabolic links" within the M-P coculture. Using *Methylomicrobium buryatense - Arthrosipira platensis* as the model coculture, we have developed the very first GEM model for the M-P coculture using SteadyCom. The coculture GEM was validated by comparing the model predicted individual growth rates with experimental measurement, while using the cross-membrane fluxes for major substrates uptake as constraints.

The coculture GEM is able to predict the change in species abundance and consequently the population ratio of the coculture in response to changes in defined nutrient and carbon substrate to the model. In addition, the coculture GEM is able to predict the "metabolic links" within the coculture. For the model M-P coculture, our GEM predicted 19 metabolites being exchanged between the methanotroph and photoautotroph. The predicted metabolic exchanges include organic acids, amino acids, as well as key metabolite involved in central carbon metabolism. These model predicted metabolic links offer the specific targets to be tested via designed experiments. Currently, we are conducting experiments to collect meta-transcriptomic profiles of the coculture to validate these model predicted exchanges.

A novel circulating coculture biofilm photobioreactor(CCBP): addressing the engineering challenges associated with the M-P coculture-based biotechnology for biogas conversion

Despite the many advantages offered by the M-P coculture for biogas conversion, several long standing technical and engineering challenges have to be addressed before the coculture-based biotechnology can be successfully commercialized. Specifically, mass transfer resistance associated with gas phase substrate and light attenuation in liquid severely limit the achievable cell density and scale up of the coculture-based biotechnology. In addition, the high energy cost associated with biomass harvesting is another major challenge.

To address these long standing challenges, we proposed to explore biofilm-based cultivation, and have developed a patent pending circulating coculture biofilm photobioreactor (CCBP) (USPTO patent application # 16,934,766, filed on July 21, 2020 [1]). The CCBP is designed to effectively address the challenges associate with gas substrate, light attenuation, and biomass harvesting, by offering the following major advantages. (1) By exposing the coculture biofilm directly to the gas phase, the mass transfer resistance from the gas phase to the cells can be significantly reduced, which eliminates the energy intensive agitation or aeration process. (2) The vertically arranged biofilm not only provides high biomass production area with low land footprint requirement, which allows effective scale up without light availability limitation, but also enables sun light dilution and reduce light inhibition. (3) Biomass harvesting can be achieved easily through a retractable scraping blade, therefore significantly reduce the energy and cost required for biomass harvesting. (4) Extracellular polymeric substance (EPS) and emergent properties of the biofilm could enhance the tolerance of the M-P coculture to culture stress and inhibitors.

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 P200A150074.

Publications:

1. 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;
2. Badr K., Whalen W., He Q.P. & Wang J. (2020), Fast and Easy Quantitative Characterization of Methanotroph-Photoautotroph Cocultures, ***Biotechnology and Bioengineering***, in press, https://doi.org/10.1002/bit.27603;
3. 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;
4. Badr K., Hilliard M., Roberts N., He Q.P. and Wang J. (2019), Photoautotroph-Methanotroph Coculture – A Flexible Platform for Efficient Biological CO_2-CH_4 Co-utilization, ***IFAC-PapersOnLine***, 52 (1), 916-921;
5. 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.

Microbiome transfer and synthetic community approaches for determining the genetic and environmental factors underlying mutualism within a *Sphagnum* peatmoss system

David J. Weston[1*] (westondj@ornl.gov), Travis J. Lawrence[1], Alyssa Carrell[1], Tatjana Živković[1], Sara Jawdy[1], Dale Pelletier[1], Jon Shaw[2], Gustaf Granath[3], Adam Healey[4], Jeremy Schmutz[4]

[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 diazotrophs 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 importance of 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 germ-free plants, and exposed the new hosts to temperature stress. Although warming decreased plant photosynthesis and growth in germ-free plants, the addition of a microbiome from a thermal origin that matched the experimental temperature completely restored plants to their pre-warming growth rates. Metagenome and metatranscriptome analyses revealed that warming altered microbial community structure, including the composition of key cyanobacteria symbionts, in a manner that induced the plant heat shock response, especially the Hsp70 family and jasmonic acid production. The plant heat shock response could be induced even without warming, suggesting that the warming-origin microbiome provided the host plant with thermal preconditioning. Together, our findings show that the microbiome can transmit thermotolerant phenotypes to host plants, providing a valuable strategy for rapidly responding to environmental change.

In an effort to understand how broad these finding extend to other warming sites, we performed a microbiome transfer study using 5 donor *Sphagnum* species across global peatland warming sites spanning Sweden, France, Iceland and USA (SPRUCE stie). The microbiomes were transferred onto three germ-free *Sphagnum* recipients in the laboratory and exposed to a range of experimental temperatures. Growth rate of *Sphagnum* recipients was recorded over 4 weeks. Enhanced growth rates of recipient *Sphagnum* was measured when cultured with microbiome from a matching donor *Sphagnum* species and was 48% and 252% greater than the maximum growth rate of the non-matched and germ-free *Sphagnum* host-microbiome pairs, respectively. Normalized growth rates were maximized for matched host-donor pairs when the treatment temperature was similar to the origin temperature. Ongoing metagenome and metatranscriptome analyses are being conducted to determine if the microbiome influence on host HSP70s and jasmonic acid is conserved across these diverse conditions.

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 chemicals

Lauren A. Riley[1,2], Neely Wood[1,3], Adam M. Guss[1,2], **Janet Westpheling (janwest@uga.edu)[3]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]University of Tennessee, Knoxville; [3]University of Georgia, Athens

The native ability to condense acetyl-CoA to efficiently generate C4 to C8 compounds makes *Megasphaera elsdenii* a compelling platform for the production of fuels and chemicals from lactate and plant carbohydrates. *M. elsdenii* produces organic acids as fermentation products from lactic acid and glucose, including butyric (4 carbon), valeric (5 cargon), hexanoic (6 carbon), and in some cases octanoic (8 carbon) acids as major fermentation products, likely via a chain elongation pathway using acetyl-CoA. As the carbon chain length increases, fuel properties improve. Energy density increases and hygroscopicity decreases. Virtually nothing is known about the metabolic pathways in *M. elsdenii* that result in organic acid formation beyond predictions based on genome annotation. We developed the first methods of DNA transformation of *Megasphaera elsdenii*, developed selectable markers and chromosomal deletion methods opening this organism to advanced physiological studies and bioengineering. We used these tools to identify genes involved in valeric acid production and to engineer butanol production from lactic acid. We generated a metabolic reconstruction of glucose and lactic acid conversion to hexanoic acid using the DOE KBase platform. We continue to develop genetic tools to enable rapid and complex construction of strains capable of producing long chain carbon molecules at high yield and high titer.

This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Energy Frontier Research Centers program under Award Number DE-SC0019401.

Developing the yeast *Kluyveromyces marxianus* as a thermotolerant bioproduction host

Ian Wheeldon,[1*] (wheeldon@ucr.edu), Xuye Lang[1], Mengwan Li[1], Pamela Besada-Lombana[2], Danielle Bever-Sneary[2], Tami McTaggart[2], Nancy Da Silva[2], Dipankar Baisya[1], Stefano Lenoardi[1]

[1]University of California, Riverside; and [2]University of California, Irvine

Project Goals: This systems and synthetic biology project seeks to understand and engineer the native stress tolerance phenotypes of the yeast *Kluyveromyces marxianus* with the goal of developing a new synthetic biology chassis for fuel and chemical production.

The non-conventional yeast *Kluyveromyces marxianus* is one of the fastest growing eukaryotes, is thermotolerant to temperatures upward of 50°C, and has the capacity to assimilate a wide range of C_5 and C_6 sugars. These traits make *K. marxianus* an attractive host for the industrial production of biochemicals. However, in comparison to the common yeast synthetic biology chassis, *Saccharomyces cerevisiae*, there is a clear lack of functional genomic and synthetic biology tools. In the initial stages of this project, we were successful in creating a new synthetic biology toolbox for pathway engineering, including standardized promoters with variable expression [1], CRISPR-mediated markerless integration [2], and strategies for rapid pathway refactoring. We have also demonstrated the utility of these tools by engineering enhanced production of triacetic acid lactone (TAL) and 2-phenolethanol. Here we report on the success of our pathway engineering efforts as well as describe our current efforts in creating a genome-wide CRISPR-Cas9 system for functional genomics.

With respect to pathway engineering, we have enhanced production of the native metabolite, 2-phenolethanol, by refactoring the Shikimate pathway [2]. A 27-member strain library that varied the expression of Shikimate pathway genes ARO4, ARO7, and PHA2 demonstrated that flux to 2-phenolethanol was limited by ARO4 expression. Alleviating tyrosine inhibition to ARO4 and increasing expression level resulted in significant improvements in titer. Building on the refactored Shikimate pathway, a titer of nearly 2 g/L of 2-phenolethanol was achieved by eliminating acetyltransferase activity to limit conversion to 2-phenolethyl acetate, overexpressing the key Ehrlich pathway gene ARO10, and by culturing in a small-scale bioreactor with fed-batch operation. After initial studies demonstrated the promise of *K. marxianus* for high-level TAL synthesis [3], we engineered the central carbon pathways using our high-efficiency PolII-based CRISPR system. We redirected carbon flux via native gene knockouts and heterologous gene integrations, demonstrated the importance of metal ion availability, and evaluated enzyme colocalization in the mitochondria to increase efficiency. The introduction of a heterologous carbon saving pathway increased TAL synthesis by 2-4 fold, reaching titers in excess of 1.3 g/L.

In collaboration with the DNA Synthesis team at JGI, we have created a genome-wide CRISPR-Cas9 functional genomics screening tool for *K. marxianus*. We first designed eight single guide RNAs (sgRNAs) for every gene in the genome; four guides targeted the coding region, while another four guides targeted the promoter region. In total, more than 35,000 guides were designed, synthesized, and cloned into a vector for sgRNA expression. Building on our past work in the non-conventional yeast *Yarrowia lipolytica* [4], we are conducting a genome-wide

cutting efficiency experiment to quantify the ability of each guide in the library to make a double stranded break in the genome. To do so, we have eliminated the ability of *K. marxianus* to repair DNA through non-homologous end-joining, the dominant repair mechanism in this and other non-conventional yeast. These genome-wide cutting efficiency experiments are currently ongoing, but we have developed the analysis workflow to successfully translate the outcomes of these experiments into a machine learning-based algorithm to predict guide cutting efficiency. Our method of quantifying Cas9 cutting efficiency across the genome generates tens of thousands of data points relating CRISPR guide sequence, genome structure (*i.e.*, chromatin structure via ATAC-seq and nucleosome occupancy via MNase-seq), and Cas9 activity. From the Yarrowia data set, our deep-learning convolutional neural network (CNN) can predict CRISPR guide sequences that result in high Cas9 activity at a desired locus. Validation experiments are on-going, but initial results suggest that we can predict sgRNAs for Cas9 that lead to a strong cutting efficiency 95% of the time (based on 19 successful guides out of 20 tested). Current work is extending this analysis to *K. marxianus* Cas9 experiments and to Cpf1-based cutting in Yarrowia.

References

1. Lang, X., et al., *Developing a broad-range promoter set for metabolic engineering in the thermotolerant yeast Kluyveromyces marxianus.* Metabolic Engineering Communications, 2020. **11**: p. e00145.
2. Li, M., et al., *CRISPR-mediated multigene integration enables Shikimate pathway refactoring for enhanced 2-phenylethanol biosynthesis in Kluyveromyces marxianus.* Biotechnology for Biofuels, 2021. **14**(1): p. 3.
3. McTaggart, T.L., et al., *Synthesis of polyketides from low cost substrates by the thermotolerant yeast Kluyveromyces marxianus.* Biotechnology and Bioengineering, 2019. **116**(7): p. 1721-1730.
4. Schwartz, C., et al., *Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast Yarrowia lipolytica.* Metabolic Engineering, 2019. **55**: p. 102-110.

This work was supported by DOE DE-SC0019093

FATHMM: Frameshift Aware Translated Hidden Markov Models

Genevieve Krause,[1]* (genevieve.krause@umontana.edu),
Travis Wheeler[1]

[1]University of Montana, Missoula.

Project Repository: https://github.com/TravisWheelerLab/hmmer/tree/frameshift

Project Goals

Frameshifts can occur in protein coding regions of DNA sequences due to natural processes such as pseudogenization, in which a gene that has fallen out of use acquires mutations in the normal course of replication and repair. They can also result from errors made during DNA sequencing, particularly when using newer long-read sequencers [1]. Improving the annotation of frameshifted sequences is therefore an important step in improving the annotation of both highly decayed pseudogenes [2], and microbial metagenomics dataset that increasingly rely on long read sequencers for assembly [3]. In pursuit of this goal, we have created FATHMM, a sequence similarity search tool that produces accurate translated alignments between protein profile hidden Markov models and DNA sequences containing frameshifts.

Abstract

High-quality annotation of DNA typically relies on sequence alignment to produce evidence of evolutionary relationships between sequences. Annotation of protein coding DNA can be achieved through translated alignment, in which the DNA is aligned to known proteins by translating codons into amino acids. In DNA that contains errors, the correct open reading frame can be obscured by frameshift-inducing insertions or deletions, preventing accurate translation (Fig 1). By explicitly modeling frameshifts within codons, FATHMM allows translated alignments to accommodate these errors without changing the translation of the subsequent codons. This is achieved through the use of a frameshift aware hidden Markov model (Fig 2), supported by dynamic programing algorithms that allow for variable length codons. The result is significant improvement in the sensitivity and specificity of translated alignments for frameshifted DNA (Fig 3), with only a moderate increase in overall alignment run time (Tbl 1).

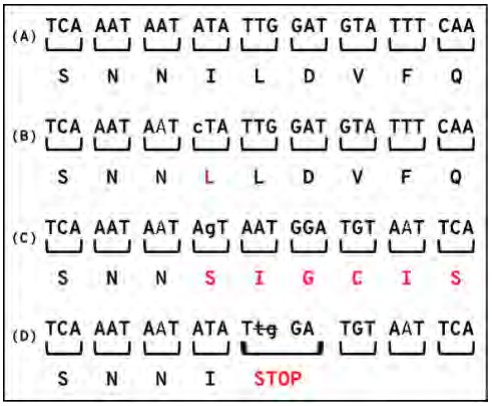


Figure 1. Indels break open reading frames.

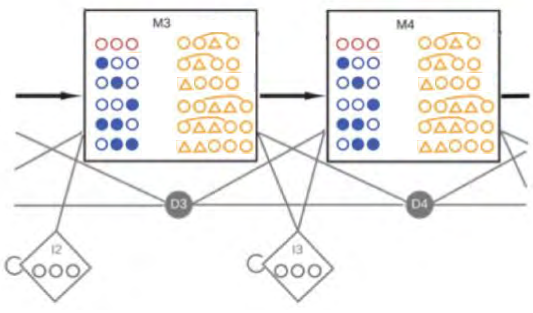


Figure 2. Portion of Frameshift-aware profile hidden Markov model.

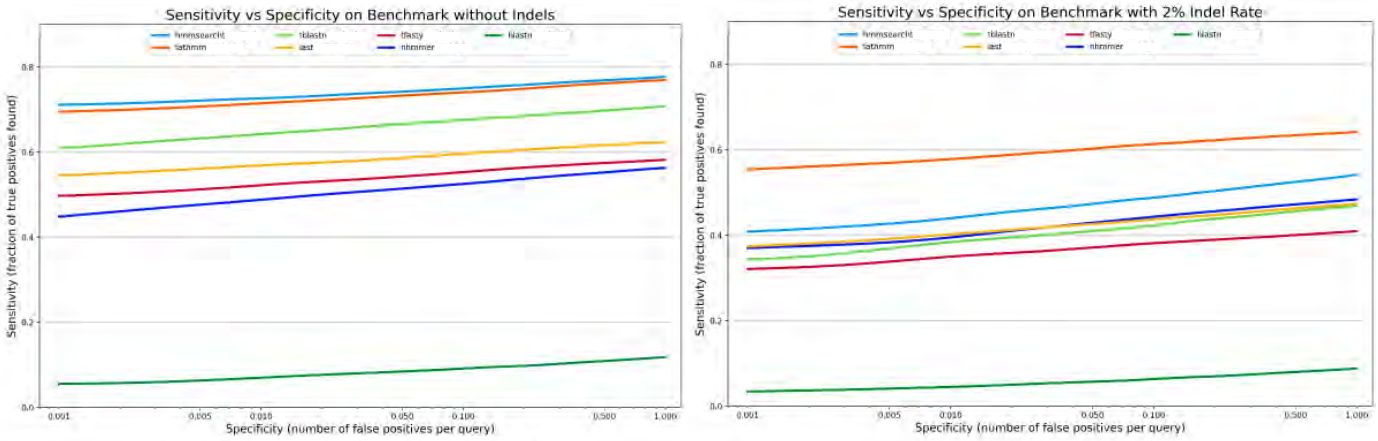


Figure 3. Sensitivity as a function of false-positive annotation for various translated search tools, on benchmarks without (left) and with (right) frameshift-inducing insertions and deletions (indels). FATHMM (orange) is much more sensitive when indels are present, but not prone to error when indels are not present.

Benchmark Runtime (Hours)							
Indel Rate	FATHMM	LAST	TFASTY	hmmsearcht	tBLASTn	nhmmer	BLASTn
0%	68.90	7.43	20.96	11.91	5.95	29.33	5.33
2%	66.65	6.41	17.09	8.89	3.62	26.16	9.02

Table 1. Runtimes of various tools tested in the frameshift benchmark from Fig 3.

References

1. Watson, M., & Warr, A. (2019). Errors in long-read assemblies can critically affect protein prediction, *37*(February), 124–126.

2. Cheetham, S. W., Faulkner, G. J., & Dinger, M. E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. *Nature Reviews Genetics*, *21*(3), 191–201. https://doi.org/10.1038/s41576-019-0196-1

3. Bharti, R., & Grimm, D. G. (2019). Current challenges and best-practice protocols for microbiome analysis, *00*(October), 1–16. https://doi.org/10.1093/bib/bbz155

Funding statement.
This work was supported by NIH grant R15 GM123487 and DOE grant DE-SC0021216.

Experimental pyrocosms demonstrate key features of the autecology of post-fire fungi.

Thomas D. Bruns (pogon@berkeley.edu) [1], Akiko Carver[1], Kyra Stillman[1], Catharine A. Adams[1], **Thea Whitman**[2]
[1]Dept. Plant and Microbial Biology University of California, Berkeley, [2] Department of Soil Science,University of Wisconsin, Madison

Project Goals: Pyrophilous fungi fruit prolifically after fire and are generally restricted to post-fire environments (Peterson 1970, Hughes et al 2020). To increase our understanding of the behavior of these fungi we have used experimental units termed "pyrocosms" to control soil heating and incubation conditions. Here we used soils from multiple forest, meadow, and agricultural sites to test the generality of the *Pyronema* response and we varied incubation conditions to see if other pyrophilous fungi could be stimulated. These manipulations were designed to test the following four hypotheses:

H 1 *Pyronema* inoculum is widespread in forest soils;
H 2 *Pyronema* inoculum is more abundant in forest soils than adjacent meadow soils;
H 3 more genera of pyrophilous fungi will develop with longer incubation times;
H 4 prolonged chilling of soil to simulate winter will stimulate additional pyrophilous fungi.

Background:
To understand the interactions of pyrophilous fungi with post-fire soil carbon we previously isolated a set of common pyrophilous fungi, sequenced them, and assembled their genomes (Steindorff et al 2021, Steindorff et al. poster– this conference), and we are studying their gene expression under controlled conditions (Whitman et al. talk – this conference). However, the factors that allow these fungi to dominant post-fire soils is poorly known. Here we used pyrocosms to subject soil samples to experimental fire in which soil heating and fuel composition was held constant, while incubation conditions involved either ambient temperatures or a simulated winter (vernalization) followed by 32 weeks of incubation. Development of fungal communities was monitored by sampling at 7 time points, followed by high-throughput sequencing of the ITS region. The results were used to test the four hypotheses listed in the goals.

Results
1) *Pyronema domesticum* was found to be widespread in conifer forest and adjacent meadow soils and responded rapidly to burning. *Pyronema* appears to be more abundant in forest compared to meadow soils. In Sierra and Cascade soil samples it accounted for over 70% of all sequences one to two weeks after burning , while in the central California coastal forest sample it accounted for a lesser, but still impressive 20% .
2) Known pyrophilous fungi that developed in some units were: *Pyronema spp*. (3 OTUs), *Pholiota aff highlandensis* (1), *Geopyxis spp.* (2), *Lyophyllum aff anthracophila* (1), *Neurospora aff. tericola* (9); however, with the exception of *Pyronema* and *Lyophyllum*, sequence abundances were < 1% at all sample times.
3) Some taxa not known to be pyrophilous responded positively to fire.
4) *Lyophyllum aff. anthracophila*, a known pyrophilous fungus, responded strongly and positively to vernalization of soil post-fire after 16 weeks of incubation and was restricted to forest soils .

Conclusions: Hypothesis (H_1) is based on the assumption that pyrophilous fungi reside in these soils between fire events as spores or sclerotia. Our results now support this hypothesis, as the rapid, predictable response of *Pyronema* is consistent with a stimulation of resident inoculum. Dispersal seems unlikely to explain the pattern, since the species rarely fruits in unburned habitats and the response was too rapid, consistent, and sample-specific to be accounted for by rare dispersal events. **Hypothesis (H_2)** that pyrophilous fungi, such as *Pyronema spp*., would be more abundant in forest vs meadow soils is based on the idea that higher course-fuel load in forests create hotter temperatures at greater depths, which have created larger inoculum pools. Our results with both *Lyophyllum* and *Pyronema* support this view. Our initial pyrocosm study resulted in only a strong response by *Pyronema spp.* (Bruns et al 2020). We predicted that longer **incubation time (H_3)** and the **simulated winter (H_4)** would result in growth of additional known pyrophilous fungi. The response of *Lyophyllum aff anthracophila* to both these factors demonstrated the validity of these hypotheses, but in a rather restricted way, as other known pyrophilous fungi, although present, did not exhibit the large increases seen with *Pyronema* or *Lyophyllum*. However, several unknown fungi responded in ways that suggest they may be unrecognized pyrophilous fungi. Our results show that additional pyrophilous fungi are present in the samples, and we are optimistic that further alteration of simple factors in the pyrocosm

environment will stimulate abundant growth of some of these additional pyrophilous fungi. Overall, the post-fire soil environment is relatively simple and likely to be highly predictable at multiple scales. Variation in fuel loads within a forest are predicted to create undulating depths of thermo-chemically defined zones at the landscape scale (Bruns et al 2020). Timing and severity of fires, though variable, are predictable in aggregate by the fire regimes associated with the plant community type and climate. Correlation between these known variables and autecology of pyrophilous fungi is highly likely, and the pyrocosm approach provides a tool to dissect their niche spaces.

References

Hughes, KW et al. 2020. **Mycologia** 112: 677-698

Bruns TD et al. 2020. **PLOS ONE** 15(3):e0222691. https://doi.org/10.1371/journal.pone.0222691

Petersen PM. Danish Fireplace Fungi. **Dansk Botanisk Arkiv.** 1970;27(3):1-97

Steindorff, AS et al. 2021. Environ Microbiol, 23: 99-109. https://doi.org/10.1111/1462-2920.15273

Funding statement:

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. 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: Life After Fire: Microbial Community Dynamics Over Time in Fire-Affected Soils

Authors: Monika S. Fischer[1]*(monikaf@berkeley.edu), Neem J. Patel[1]* (neempatel@berkeley.edu), Phillip de Lorimier[1], **Matthew F. Traxler[1]**

Institutions: [1]Plant & Microbial Biology, University of California, Berkeley, CA.

Website: https://traxlerlab.berkeley.edu/research

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 pyrolyzed organic matter, 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 natural aspect of many forest ecosystems that dramatically effects both community composition and nutrient cycling dynamics. Importantly, wildfires are increasing in frequency and severity, and this trend is projected to continue due to climate change. Bacteria and fungi are the first-responders within days-to-weeks after fire. As pioneer species, these organisms are poised to affect the entire trajectory and outcome of post-fire succession and ecosystem recovery. Previous work has demonstrated a clear successional pattern in soil fungal communities that occurs within roughly the first year following fire (1). Furthermore, fungi in the genus Pyronema are the first to dramatically dominate the post-fire fungal community, and *Pyronema domesticum* can breakdown the complex aromatic compounds that are characteristic of charcoal and pyrolyzed organic matter (1, 2). We produced a high-resolution dataset describing post-fire succession of both bacteria and fungi at a fine temporal scale following prescribed fires at an experimental pine forest (UC's Blodgett Forest). We collected triplicate soil samples at least once/month for 17 months in four different plots: high-intensity burn, moderate-intensity burn, low-intensity burn, and a no-burn control. In all plots we collected at least one time-point prior to treating the plot with fire. From these soil samples we extracted total DNA, amplified the 16S region of bacteria or the ITS region of fungi, and then sequenced these community amplicons. Here we present our preliminary analysis of this community DNA sequencing data. We show the structure of the microbial communities in each plot, and how these communities change over time in response to fire, soil pH, and seasonal weather patterns. These data provide critical information about the initial stages of post-fire recovery, in particular, the bacterial component of post-fire succession. Additionally, these data point toward key

organisms that warrant further investigation into the specific role they play during post-fire succession.

References/Publications

1. Bruns TD, Chung JA, Carver AA, Glassman SI. A simple pyrocosm for studying soil microbial response to fire reveals a rapid, massive response by Pyronema species. *PLoS ONE*. 2020;15: e0222691. doi:10.1371/journal.pone.0222691
2. Fischer MS, Stark FG, Berry T, Zeba N, Whitman T, Traxler MF. Pyrolyzed substrates induce aromatic compound metabolism in the post-fire fungus, *Pyronema domesticum*. *advanced stage of preparation for submission*.

Funding statement:

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

Bridging Scales: Conceptualizing microbe-climate links in wetland ecosystems

Pieter Candry[1] (pcandry@uw.edu), Zach Flinkstrom[1], Yi Xiong[2], Sam Bryson[1], Juliet Johnston[3], Britt Abrahamson[1], Ludmila Chistoserdova[1], Chongle Pan[2], Xavier Mayali[3], David Stahl[1], **Mari Winkler**[1]

[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 predicting their response to climate change induced stressors such as drought and saltwater intrusion is of prime importance. This project aims to link wetland microbial activity 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 plat-like, 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.

Abstract: Wetlands play a key role in the global carbon cycle by storing 20-30% of all soil carbon in 4-8% of land surface. At the same time, they are responsible for 40-50% of all biogenic CH_4 emissions, showing the large impact these ecosystems and their microbial community have on global climate processes. The nutrient density (specifically carbon and nitrogen) of these systems also means climate-induced changes in wetlands can result in strong feedbacks, either positive or negative, to climate change. For instance, estuarine wetlands could suffer from seawater intrusion as sea levels rise at increasing rates, while inland wetlands may suffer from higher and more frequent disturbances in water table levels as dry seasons become longer and dryer, while wet seasons become shorter and wetter. Predicting how these changes will affect the microbial community – and by extension GHG fluxes and the global climate – remains challenging, and there is currently no coherent approach to connect wetland microbes to global-scale climate processes. Here, we present a conceptual model of multi-scale climate-microbe interactions and propose approaches to connect this to molecular data of wetland microbial communities, experimental and mathematical models to better connect wetland microbes to climate-scale processes.

As a first step, we developed a conceptual model of wetland microbial conversions. This model holistically integrates carbon cycling with other key biogeochemical nutrient cycles (i.e. N, Mn, Fe, S), using the redox sequence as a spatial factor controlling microbial metabolisms in wetland systems. Furthermore, not only heterotrophic conversions driven by carbon are considered, but we also include an exhaustive analysis of demonstrated chemolithotrophic metabolisms, since these can result in cryptic nutrient cycles (e.g. Fe, Mn & S-cycles) driving further carbon breakdown. This conceptual model for metabolic versatility in wetland systems drove a meta-analysis of wetland metagenomes. We analyzed publicly available data from the JGI IMG/M database to connect the functional potential of wetland microbiomes to spatial organization within the wetland soil and across wetland types. These analyses showed that communities diverging at the phylogenetic level coalesce at the functional level, both for general functionality and when focusing on genomic potential for different redox reactions driving wetland nutrient cycles. The

remaining community divergence on the functional level could then be linked to physicochemical parameters in wetland ecosystems, e.g. sulfate in coastal wetlands. Future analyses will complement these public datasets with our own datasets, both from wetland sediments and lab-scale experimental models for wetland ecosystems, to better connect community structure at the functional level to environmental conditions and GHG fluxes.

Building our metabolic conceptual model of wetland microbial processes around a consistently observed spatial structure (i.e. the redox sequence), we can connect the actions of microbes to centimeter and meter-scale processes. An overview of intermediate scale processes connecting wetland microbes and climate-scale processes was developed, taking into consideration sediment aggregate structure, microbe-root interactions, wetland plant communities, water drainage, local geography and local weather patterns. It is these processes that control the microbial processes responsible for carbon and nitrogen capture, GHG fluxes from (or to) wetland ecosystems, and ultimately connect back to global climate processes. This overview was in turn used to develop wetland-specific microbe-climate feedback loops that integrate impacts of climate change on these multiple scales (Figure 1). Put together, we have developed a coherent multi-scale framework to link wetland microbes to global climate processes, a key step towards developing and testing multi-scale experimental and mathematical models.

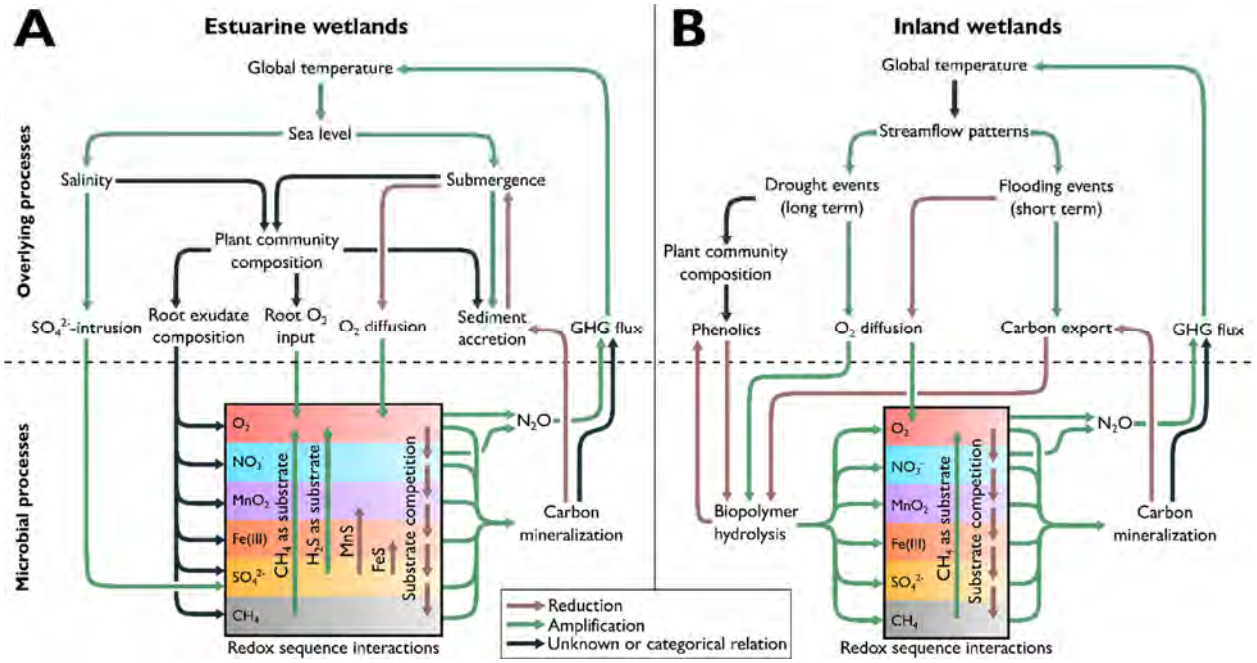


Figure 1. Feedback loops arising from multi-scale interactions in wetland systems. The dashed line separates processes driven by the microbial community (below) from larger-scale processes (above). Panel A shows feedback loops for estuarine wetlands, where rising sea levels are a major driver for changes in GHG fluxes. Panel B shows feedback loops for freshwater wetlands where changes in streamflow patterns due to climate change will affect GHG fluxes.

This work was supported by the U.S. Department of Energy's Office of Biological and Environmental Research Genomic Science Program (Grant #DE-SC0020356); *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.*

Hydrogel Beads to Encapsulate Sediment Microbes as a Strategy to Quantify Climate Impacts on Microscale Biogeochemical Activity

Juliet Johnston (johnston50@llnl.gov)[1], Bruce Godfrey[2], Congwang Ye[1], Pieter Candry[2], Britt Abrahamson[2], Sam Bryson[2], David Stahl[2], Ludmila Chistoserdova[2], Xavier Mayali[1], **Mari Winkler**[2]

[1]Physical and Life Science Directorate, Lawrence Livermore National Laboratory, Livermore, CA.
[2]Department of Civil and Environmental Engineering, University of Washington, Seattle, WA.

Project Goals: Wetlands capture and release large amounts of greenhouse gases (CO_2, CH_4 and N_2O) and predicting their response to climate change induced stressors such as drought and saltwater intrusion is of prime importance. This project aims to link wetland microbial activity 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 plat-like, 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.

Abstract: Microbial communities are the driving force for carbon degradation and nutrient cycling throughout wetland ecosystems and are expected to experience significant disruption due to climate change. For instance, estuarine wetlands are expected to suffer from increased saltwater intrusion while freshwater lacustrine wetlands may experience more frequent and more severe drought events over the coming decades. Accurately predicting the impacts of climate change on these microbial communities necessitates a mechanistic understanding of micro-scale processes and how these scale-up to ecosystem-scale fluxes. To develop such a multi-scale approach, this project aims to connect microbial processes occurring at the μm-scale to community organization in a single sediment aggregate (0.5-1mm in size) and to the nutrient flux within the soil column (meter scale). These results could then inform the quantification of greenhouse gas release and capture from wetlands (km scale) and their impact on global climates (Mm scale) by mathematical modelling. This abstract focuses on building an experimental framework to develop a sediment model system that mimics diffusion limiting conditions under controlled and hence, experimentally manipulable conditions. We will work with novel polymer based synthetic sediment-like aggregates that encapsulate native microbial wetland communities. These PEGDMA (polyethylene glycol dimethacrylate) based hydrogel beads maintain the complex spatial organization of sediment microbial communities while facilitating downstream molecular analysis due to simplifications in both matrix and community complexity. PEGDMA beads are not biodegradable, allowing entrapment of plant like material into the bead interior and hence the assembly of a sediment model system that can mirror carbon availability of real wetland systems. A special feature of these hydrogel beads is that they can be produced in various sizes to match the soil compositions from coarse granules of several mm down to fine sand smaller than 1 mm. This provides us with the opportunity to design the proper beads to match each individual study's requirements, as well as investigate the role bead size, or sediment aggregate size, plays in microbial community structure and nutrient cycling.

An initial step towards the application of hydrogel entrapment to study complex soil microbial communities is the validation of the experimental methodology and the subsequent challenges in data analysis. Important milestones for developing this experimental approach include assessing

the activity of microbes entrapped in the hydrogel and evaluating the compatibility of hydrogel encapsulated microbes with downstream 'omics'-based and stable isotope probing (SIP) analyses such as NanoSIMS and proteomic-SIP. A particular challenge for the latter will be identification of cross-feeding between potential synergistic partnerships that occur in complex communities such as wetland soils. In order to validate the technology prior to encapsulating complex wetland communities we tested the approach with two simplified biofilm communities. In the first hydrogel experiment, we paired a complete ammonium oxidizing bacterium (Comammox, *Nitrospira inopinata*) with an enriched anaerobic ammonium oxidizing (Anammox, *Candidatus* Brocadia anammoxidans) community. *Nitrospira inopinata* localized on the aerobic bead periphery to provide *Ca.* Brocadia anammoxidans with nitrite in the inner anaerobic core to form N_2. An unexpected result was the complete removal of nitrogen as Anammox produces some nitrate in its anabolic pathway. We hypothesize that Anammox and Comammox form a mutualistic relationship under anaerobic conditions in which Comammox utilizes formate and nitrate (supplied by anammox) to perform anaerobic nitrate reduction to nitrite, which is then removed by anammox to N_2, leading to complete nitrogen removal. This hypothesis was also supported with *N. inopinata* batch testing and observations of spatial organization using FISH (fluorescent in-situ hybridization) (Figure 1). The second synthetic community under investigation is an Anammox community that has been shown to support a significant heterotrophic community. This partnership has exhibited stability when maintained within bioreactors and without the addition of organic carbon sources for more than one year. This suggests that the autotrophic Anammox bacteria are the primary producers of the system, supplying carbon to the diverse heterotrophic organisms for secondary production. By supplying ^{13}C-bicarbonate over an extended incubation period (~1 month) and collecting samples at multiple time-points for proteomic SIP analysis, we will identify how autotrophs fuel the heterotrophic community. We are scaling up the successes from simple communities entrapped in hydrogels towards more complex microbial cross-feeding dynamics and relationships for this study on climate change. Therefore, we have developed protocols to extract and immobilize live cells from freshwater and saltwater wetland sediments at high cell densities while drastically reducing humic matter that potentially inhibits molecular analysis. Overall, our hydrogel beads show promise for downstream analytical applications such as meta-omics for cellular functions and activities as well as NanoSIMS for spatial-analysis and nutrient uptake in complex wetland environments.

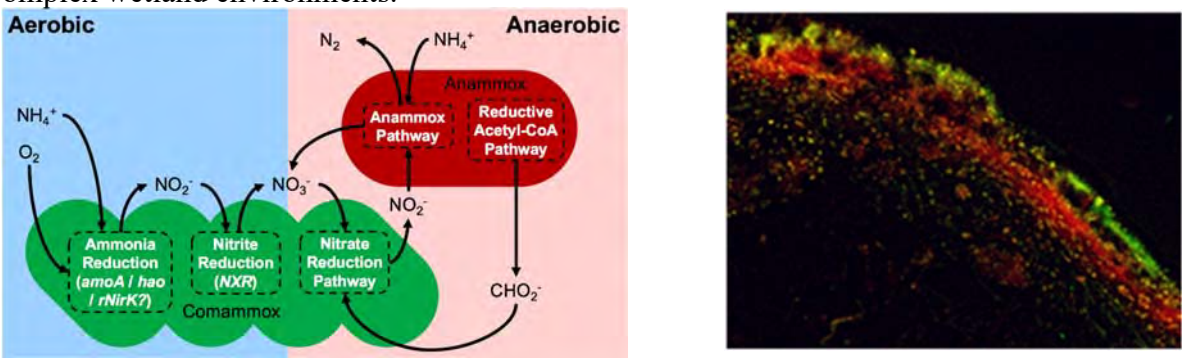


Figure 1: *Left:* Experimentally confirmed pathways of Comammox. Formate is hypothesized to be supplied by Anammox. *Right:* Hydrogel bead confirms partnership of Comammox (green) and Anammox (red).

This work was supported by the U.S. Department of Energy's Office of Biological and Environmental Research Genomic Science Program (Grant #DE-SC0020356); 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.

Metabolic Modeling of Synthetic Estuarine Wetlands Microbial Communities in response to Climate Change

Dongyu Wang[1], Britt Abrahamson[2], Xuanyu Tao[1], Neil Q. Wofford[1], Aifen Zhou[1], David A. Stahl[2], Jizhong Zhou[1], Michael J. McInerney[1], Ralph S. Tanner[1], Xavier Mayali[3], **Mari Winkler[2]**, Chongle Pan[1*] (cpan@ou.edu)

[1]University of Oklahoma, Norman, OK
[2]University of Washington, Seattle, WA
[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 predicting their response to climate change induced stressors such as drought and saltwater intrusion is of prime importance. This project aims to link wetland microbial activity 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 plat-like, 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.

Abstract:

Microbial communities in estuarine wetlands work in concert to degrade lignocellulosic biomass anaerobically to CH_4, CO_2, and H_2S. The greenhouse gas (GHG) emissions from estuarine wetland ecosystems will be severely perturbed by increased seawater intrusion caused by climate change. It remains a challenge to accurately model and predict ecosystem feedbacks to such perturbations. The altered GHG emissions from estuarine wetlands may in turn provide positive or negative feedbacks to global-scale climate change. A mechanistic understanding of metabolic networks of interaction in estuarine wetland communities is critically needed to elucidate the impact of seawater intrusion on these processes. We hypothesize that seawater intrusion of estuarine wetlands will reduce, but not eliminate, methane emissions and increase H_2S and CO_2 emissions, due to the increased availability of sulfate.

Lignocellulosic biomass in estuarine wetlands is degraded by four major functional guilds of microorganisms: cellulolytic bacteria, hydrogenotrophic methanogens, acetoclastic methanogens, and sulfate reducing bacteria. These functional guilds maintain extensive metabolic interactions with one another, including the cross-feeding of carbon from cellulolytic bacteria to the other guilds, and substrate competition between methanogens and sulfate reducers. These metabolic interactions will be perturbed by the increased sulfate availability from seawater intrusion. Sulfate respiration is energetically more favorable than methanogenesis and therefore sea level rise could promote sulfate reduction at the expense of methanogenesis and accelerate the overall carbon degradation.

Many natural wetland communities have been characterized by genotyping and meta-omics analyses. While these studies revealed the taxonomic composition and the key metabolic processes, they do not provide insights on the fate of carbon within the community due to

challenges in measuring and modeling carbon fluxes in a highly complex microbial community. There is hence an urgent need to develop a simplified mechanistic metabolic model that mimics these complex wetland communities with reduced complexity. Synthetic communities consist of defined microorganisms that represent key known metabolic processes that are representative of interactions occurring in natural communities and provide model systems of reduced complexity to study the interplays between those metabolic processes. A synthetic community also allows the integration of experimentally confirmed metabolic pathways into metabolic flux balance models that can help to better predict the carbon flux in natural ecosystems and these models can be very powerful to generate hypotheses for natural ecosystems that can be experimentally tested.

In this study, we designed a synthetic community comprised of representative members of the four major functional guilds of estuarine wetland communities that are responsible for cellulose hydrolysis, hydrogenotrophic methanogenesis, acetoclastic methanogenesis, and sulfate reduction (Figure 1). Several potential representatives were tested from each of the four functional guilds in a synthetic community, considering how they interacted *in vitro*. The most optimal consortium included the cellulolytic bacterium *Clostridium cellulolyticum*, the hydrogenotrophic methanogen *Methanospirillum Hungatei*, the acetoclastic methanogen *Methanosaeta concilii*, and the sulfate reducing bacterium *Desulfovibrio vulgaris Hildenborough*. The quad-culture was compared with a series of tri-cultures and bi-cultures of its constituents. The quad-culture was more productive in methane production than the tri-cultures and bi-cultures, demonstrating the synergy exerted by the community to conversion of cellulose to methane. The addition of sulfate to the synthetic community promoted sulfate reduction and reduced the methane production. Flux-balance models were constructed for each bacterium to simulate the carbon fluxes in the synthetic community. This provides a foundation for the development of mechanistic metabolic models of the estuarine wetland communities. We will develop and validate the metabolic models using genomics, proteomics, stable isotope probing, and gas emission data. The metabolic models will allow more accurate prediction of future GHG emissions from estuarine wetlands under climate change.

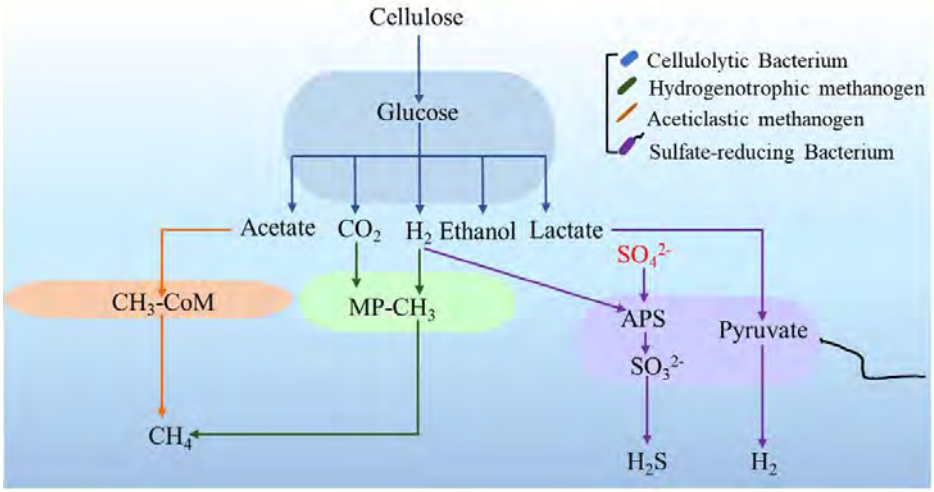


Figure1. Scheme of carbon metabolism pathway in the synthetic community

Funding statement.

This work was supported by the U.S. Department of Energy's Office of Biological and Environmental Research Genomic Science Program (Grant #DE-SC0020356); 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.

From Genomes to Methane Emission: Targeting Critical Knowledge Gaps in Wetland Soils

Adrienne Narrowe[1*] (adrienne.narrowe@colostate.edu), Bridget McGivern, Tim Morin[4], Jorge Villa Betancur[2], Pengfei Liu, Rebecca Daly[1], Mikayla Borton[1], David Hoyt[3], Christopher Miller[6], William Riley[5], Trent Northen[5], Malak Tfaily, Gil Bohrer[2], **Kelly Wrighton**[1]

[1]Colorado State University, Fort Collins, CO; [2]The Ohio State University, Columbus, OH; Pacific Northwest National Laboratory, Richland, WA; [4]State University of New York College of Environmental Science and Forestry, Syracuse, NY; [5]Lawrence Berkeley National Laboratory, Berkeley, CA; [6]University of Colorado Denver, Denver, CO

Project Goals:

This early career research project interrogates carbon cycling in soils from a freshwater, coastal wetland adjacent to Lake Erie. Due to climatically driven increases in water levels in Lake Erie, the wetland water table has increased by ~3-6 ft over the sampling period, with hydrological inundation serving as a natural, model system for climatic induced shifts on soil biogeochemical cycling. With a sampling that incorporated time-series and highly spatially resolved sampling over this hydrological perturbation, we use multiple microbial genomics methods to resolve microbial community metabolic activity over years, across and within seasons, and along centimeter depth to meter land coverage gradients. Leveraging the multi-omics data in conjunction with geochemical, metabolomic, and greenhouse gas measurements, we provide unprecedented insight into how hydrological perturbations impact methane flux from a coastal, freshwater wetland.

Abstract:

Wetlands are the largest natural source of atmospheric methane, a potent greenhouse gas. Being able to predict the state and the changes microbial methane metabolism in wetland soils is critical to accurate modeling of methane flux. During our sampling over a three-year period, the wetland has experienced a shift in hydrologic state, with mud flats that were seasonally exposed now submerged under 3-6 ft underwater and where the water depth in prior open water channels increased by up to 2 ft. In this project we quantify how climatically induced hydrological disturbances change soil redox, organic carbon content and concentration, microbial decomposition, and methane cycling across these flooded conditions, offering a new chemical and microbial resolution of the controllers of the soil microbial methane cycle and how these are impacted by hydrologic perturbations caused by rising lake levels and modified storm events.

We first verified this hydrologic perturbation decreased the redox in historical season mud-flat and open water surface soils with bulk porewater dissolved oxygen levels dropping from oxygenic (>65 μM) to anoxic (< 2μM) in the first 0-5 cm of soil. Regardless of historical land coverage (open water, seasonal mud-flat), previously anoxic deeper soils (25-35 cm) remained anoxic. Consequently, we hypothesized that microbial carbon cycling in surface soils, especially those seasonally exposed, will be more impacted than in deeper soils which are better insulated from climatically driven perturbations. Here using multi-omics and in situ methane production data we address the impacts of shifting redox on microbial carbon cycling by initially targeting microbial metabolism at two critical stages of the carbon cycle that are historically thought to be regulated soil oxygen concentrations: plant polymer degradation and methanogenesis.

In oxygen-depleted soils, it is widely asserted that polyphenols inhibit microbial activity through cellular toxicity, substrate binding, and microbial enzyme inhibition, thus acting to “lock” soil carbon cycling. Given the need for a tractable substrate, we first used anoxic wetland soil reactors amended with and without a chemically defined polyphenol to test this hypothesis. Challenging the idea that polyphenols are not bioavailable under anoxia, we provided time-series metabolite and gene expression evidence for polyphenol depolymerization, resulting in monomer accumulation and catabolism, and further phenolic acid metabolism. Our findings also indicated that polyphenol amendment was not universally toxic, instead selectively stimulating a subset of microorganisms. Furthermore, this data provided a new view of microbial-polyphenol interactions, where these compounds did not restrict, but enhanced, overall soil carbon cycling increasing the production of methanogenic substrates (formate and acetate) relative non-amended controls. Collectively, our high-resolution multi-omic results provide the first demonstration that soil microbiota subverted the polyphenol lock, to sustain carbon depolymerization, and downstream carbon processing even in anoxic soils.

To better track carbon cycling and methane production/consumption in these soils in situ, we developed the MUCC (Meta'omics to Understand Climate Change) wetland genome database. This database was composed of over 17,000 genomes and contains a catalog annotated genomic content of over 2,500 dereplicated quality microbial genomes. Although soil methanogens and methanotrophs have been cultivated for decades, our genome recovery approach resulted in 88 methanogen genomes representing all methanogenic pathways. These include 3 novel families, 17 novel genera and for some taxa, the first genomes identified in a wetland environment, illuminating the phylogenetic and metabolic diversity harbored in terrestrial ecosystems. Over this three-year period, we have generated an unprecedented wetlands metatranscriptome of 115 spatially and temporally distributed samples which we have used better understand how soil microbial metabolism is impacted by flooding. We link this metabolic activity to *in situ* porewater methane measurements, methane chamber flux measurements, geochemistry, and NMR, LC-MS, and FT-ICR metabolite data to provide a highly resolved view of the methane cycling organisms and the upstream and downstream processes impacting the resulting methane emissions.

First, we used NMR-metabolites to survey the distribution of methanogenic substrates across the wetland, with flooding we observed increases in methanogen substrates of acetate and methanol only in the surface soils. Concomitant with increased availability of these substrates, we observed gene activity from methanogens utilizing acetate and methanol in these surfaces increased 4.9 and 1.6-fold respectively with flooding. This expanded substrate availability and concentrations is likely responsible for the nearly 5-fold increase in methane production reported in the surface soils with flooding. Surprisingly, while we expected the shift to anoxia would decrease aerobic methanotrophy in surface soils, gene expression data indicated aerobic methanotrophs metabolized methane using very low oxygen concentrations, and we also demonstrated first soil gene expression data suggesting nitrate enabled methanotrophy in soil systems. Given that methanol utilizing methanogens and anaerobic methanotrophs are currently not accounted for in climate models, our findings provide important data to be used in updating biogeochemical models of terrestrial methane emissions. Despite these redox and soil carbon changes, using co-expression gene network analysis we show the membership of carbon cycling communities does not change, but their activity is enhanced with flooding.

In summary, while our findings show redox shifts associated with flooding, the microbial communities and metabolic networks are largely maintained over years despite overwhelming environmental changes. We find that anoxia does not short cut microbial carbon cycling as thought, but instead previously underappreciated microbial carbon depolymerization increases the concentration of methanogen metabolites. This increased availability of substrates, coupled to anoxia, resulted in significant increases in methane production in surface soils with flooding, regardless of historical land coverage type. Our findings demonstrate methane metabolic circuitry encoded across microbial genomes may be resilient drying/rewetting/flooding pressure, but the flux through these networks may result in greater in situ production of methane. Collectively, our results highlight the soil microbial metabolisms influencing the terrestrial microbial methane cycle under climatic induced shifts, thereby offering direction for increased realism in predictive process-oriented models of methane flux in wetland soils.

References/Publications:

1. Angle JC, Morin TH, Solden LM, Narrowe AB, Smith GJ, Borton MA, Rey-Sanchez C, Daly RA, Mirfenderesgi G, Hoyt DW, Riley WJ, Miller CS, Bohrer G, Wrighton KC. 2017. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. Nature Communications 8:1567.
2. Narrowe AB, Angle JC, Daly RA, Stefanik KC, Wrighton KC, Miller CS. 2017. High-resolution sequencing reveals unexplored archaeal diversity in freshwater wetland soils. Environmental Microbiology 19:2192-2209.
3. Narrowe AB, Borton MA, Hoyt DW, Smith GJ, Daly RA, Angle JC, Eder EK, Wong AR, Wolfe RA, Pappas A, Bohrer G, Miller CS, Wrighton KC. 2019. Uncovering the Diversity and Activity of Methylotrophic Methanogens in Freshwater Wetland Soils. mSystems 4:e00320-19.
4. Smith GJ, Angle JC, Solden LM, Borton MA, Morin TH, Daly RA, Johnston MD, Stefanik KC, Wolfe R, Gil B, Wrighton KC. 2018. Members of the Genus *Methylobacter* Are Inferred To Account for the Majority of Aerobic Methane Oxidation in Oxic Soils from a Freshwater Wetland. mBio 9:e00815-18.

Funding Statement: This material is based upon an Early Career Award to KCW from the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0019746

Advancing the molecular understanding of growth in algae and plants

Feiyang Ma[1], Patrice A. Salomé[1], Marina Linardic[1], Shawn Cokus[1], Sabeeha S. Merchant[1,2,3], Matteo Pellegrini[1], Siobhan A. Braybrook[1]* (siobhanb@ucla.edu), and **Todd O. Yeates**[1]

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 algal and plant genomics. Green algae and plants are primary producers whose growth generates lipid and carbohydrate-based macromolecules for the production of biofuels and bioproducts. We have been working to advance our molecular understanding of algal and plant growth through the development and application of advanced 'omic' technologies. Our team is pioneering new techniques for scRNAseq in model algal species and advancing genome and transcriptome studies in non-model algal species. Our biological focus centers on optimizing growth in plants and algae in a changing environment.

Abstract:

The green unicellular alga *Chlamydomonas reinhardtii* has been a subject of study in the group and we have previously described how the bulk transcriptome responds to environmental changes. Here, we have applied single-cell RNA sequencing (scRNA-seq) to probe the heterogeneity of Chlamydomonas cell populations under three environments and in two genotypes differing in the presence of a cell wall. We demonstrate that single algal cells with and without cell walls can be used for scRNA-seq, offering the possibility to sample algae communities in the wild and the laboratory. We further show that single cells could be successfully separated into non-overlapping cell clusters according to their growth conditions; cells exposed to iron or nitrogen deficiency were easily distinguished despite a shared tendency to arrest cell division to economize resources. While these results mirrored our bulk RNA-seq results, they also revealed inherent heterogeneity that correlated with circadian responses.

Marine macroalgae also undergo growth in variable environmental conditions. We have produced a de novo transcriptome describing desiccation in lab and field conditions for *Fucus*, or rockweed. Since macroalgae are an emerging molecular system, our de novo transcriptome represents an incredible resource towards understanding novel desiccation tolerance mechanisms; our goal is to further dissect the molecular responses to desiccation in marine macroalgae and to apply this knowledge to improving drought tolerance in land plants. In parallel to this project, we are developing a draft genome for *Fucus* utilizing Illumina and Nanopore sequencing; this would be the third macroalgal genome available.

Future directions. Within the DOE IGP at UCLA we are working to accelerate the development of innovative genomics and structural biology tools for bioenergy-relevant plants. We are

developing technologies in two model species: the model grass, *Brachypodium distachyon*, and the model eudicot *Thlaspi arvense,* or pennycress. This approach allows for rapid technology development within our existing facilities which we can transfer to other DOE researchers for deployment in crops. Our plan will interrogate cell wall structure and function in these models, as it relates to growth. We will leverage our experience in scRNA-seq to develop methods that maintain tissue-context spatial information. Our program will examine the *in vivo* structural details of cell wall biosynthesis enzymes and their subcellular localization in cell-type-specific contexts. *In vivo* structures for cell wall modifying enzymes, within the cell wall, will be revealed as well as how such proteins interact with each other and their target carbohydrates under relevant physiological conditions.

This work is supported by DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-FC02-02ER63421.

Development and Deployment of Enabling Technologies at the UCLA-DOE Institute

Jose A. Rodriguez,[1] David Eisenberg,[1] Marcus Gallagher-Jones[1*](marcusgj@chem.ucla.edu) Logan Richards,[1] Ambarneil Saha,[1] Matthew Agdanowski,[1] Roger Castells-Graells,[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, where we are advancing the understanding of plant and microbial biosystems, their genomics and molecular biology. Our team is pioneering new enabling capabilities that facilitate the discovery of molecular structural features affecting protein function and specificity, to better our understanding of bioenergy crops and microbes. These capabilities span the broad areas of X-ray diffraction, electron microscopy, and micro-electron diffraction (MicroED), 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. We have developed the first example of a working molecular scaffold that can image such small proteins.

Abstract:

Our efforts in imaging science and protein characterization bridge a number of technological areas to address pressing problems in protein structure and function.

Breakthroughs in cryo-EM –
Numerous technical advances have made cryo-EM an attractive method for atomic structure determination. Cryo-EM is ideally suited for very large structures; symmetrical structures like viruses are especially amenable. However, problems of low-signal-to-noise in imaging small proteins makes it practically impossible to determine structures smaller than about 50 kDa, leaving a great many cellular proteins and enzymes (and nucleic acid molecules) outside the reach of this important structural technique. 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. A 3.8 Å resolution for a 26 kDa protein provides the leading benchmark for cryo-EM scaffolding (Fig. 1). Our ongoing efforts aim to further rigidify designed scaffolds, apply them to microbial and plant protein targets, and further extend the scaffolding approaches to nucleic acid molecules.

Enabling micro-ED methods -
A broad array of atomic structures has now been determined by MicroED; they include naturally occurring peptides, synthetic protein fragments and peptide-based natural products. However, *de*

novo structure determination by MicroED remains problematic for all but ideal crystals. Automated, fragment-based approaches to structure determination eliminate the need for atomic resolution diffraction, instead enforcing stereochemical constraints through libraries of small model fragments. We have demonstrated the application of fragment-based phasing on various macromolecular structures including some for which all traditional phasing methods have failed (Fig. 2). During refinement of MicroED structures, potential maps can reveal a wealth of typically untapped information about charged states. This information remains out of reach largely because of inadequate parameterizations of ionic electron scattering factors. To rectify this, we have developed a public web server 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. We are exploring its application to a number of published cryo–EM structures that previously neglected charge, which plays an important role in electron scattering.

Universal tools for structure validation -

The veracity of validation efforts by the broad structure community has been vital in creating a robust PDB. Similar efforts are now developing in connection with structures determined by cryo-EM. Among the leading concerns is the issue of accurate determination of absolute length scale – typically embodied by refinement of the EM detector 'pixel size'. Discrepancies in this determination lead to structures suffering from overall stretching or compression, in some cases approaching 1%. This problem recalls parallel challenges addressed in earlier days of x-ray crystallography. By revisiting our earlier treatments of that problem, now in the context of cryo-EM, we have developed algorithms useful to the broad x-ray and EM communities for reliable post-facto analysis of absolute length scale in refined structures. A systematic study highlights the need for vigilance in calibrating widely used experimental facilities.

Collectively, the enabling capabilities we are developing will broadly facilitate the determination and refinement of unknown macromolecular structures with importance for bioenergy.

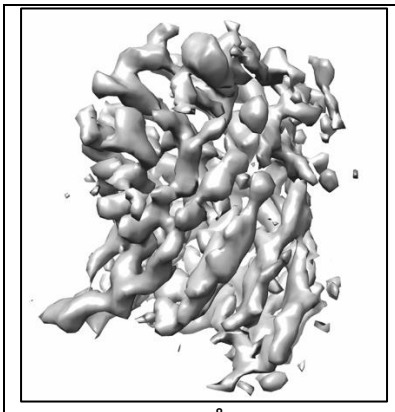


Figure 1. A 3.8 Å cryo-EM map of a small protein (26 kDa GFP).

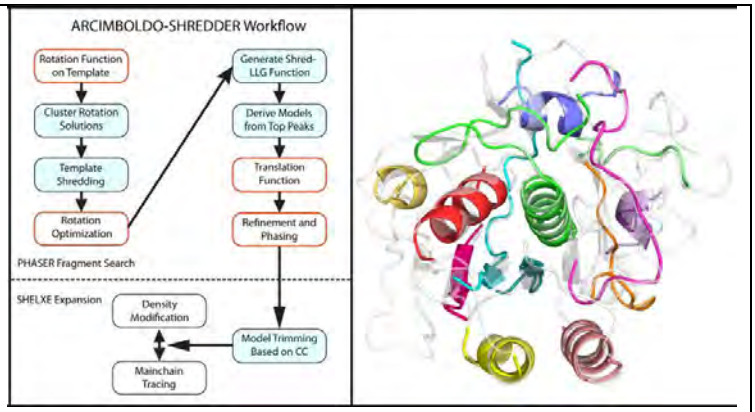


Figure 2. An automated fragment-based phasing pipeline for structure determination from electron diffraction datasets.

References

1. M.C. Thompson, T.O. Yeates, J.A. Rodriguez (2020). Advances in methods for atomic resolution macromolecular structure determination. https://f1000research.com/articles/9-667

This work is supported by DOE Office of Science, BER grant DE-FC02-02ER63421.

Title: Metabolism in Microbial Communities and the Associated Biochemistry of Polymer Deconstruction

Authors: John Muroski[1], Janine Fu[1], Erin Natale[1], Allen Takayesu[1], Brendan Mahoney[1], Orlando E. Martinez[1], Marco Morselli[1], Matteo Pellegrini[1], Rachel R. Ogorzalek Loo[1], Katherine Chou[2], Robert Clubb[1], Joseph E. Loo[1] and Robert Gunsalus[1]* (robg@microbio.ucla.edu) and **Todd Yeats[1]**

Institutions: [1]UCLA DOE Institute, University of California, Los Angeles, Los Angeles, CA; [2] National Renewable Energy Laboratory, Bio Science Center, Golden, CO.

Project Goals:

Our UCLA DOE Institute microbiology project area goals are to employ a coordinated set of molecular and *in silico* approaches to examine model microbial communities and their component parts to better understand the complex interactions and 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. With other UCLA DOE Institute members, we are initiating development of next-gen omics methods to interrogate environmental and genomic interactions between multiple organisms within a microbial community, and where possible, test these proposed functions.

Abstract text:

Major activities within the UCLA-DOE Institute in the past year deal with the investigation microbial metabolic processes central to global carbon cycling and biofuel production. We are employing molecular and *in silico*-based approaches to analyze anaerobic plant biomass degradation pathways in model syntrophic and fermentation-based microbial systems.

Abstract text:

Elucidation of syntrophic microbial pathways for metabolism of model substrates.

Proteomic and informatic studies were performed on defined microbial communities to elucidate how model environmental fatty acid substrates are metabolized into their gaseous end products. All key pathway enzymes were identified in the syntrophic bacterium, *Syntrophomonas wolfei* for C4-C8 chain length oxidations along with the supporting electron transport reactions to generate hydrogen gas and acetate. Analogous proteomic co-culture studies with the *Methanospirillum hungatei* community partner reveal a conserved carbon dioxide reduction pathway reliant on a core hydrogenase and formate dehydrogenase. Recombinant studies of five key enzymes of the carbon oxidation pathway in *Syntrophomonas wolfei* were performed by cloning, expression and purification to better define the biochemical basis for the thermodynamic limiting steps occurring during syntrophic cell growth. Protein structures have also been generated for two of these core syntrophic pathway enzymes for subsequent modeling.

Acyl-lysine modification of syntrophic pathway proteins.

Proteomic and mass spectrometry methods were also performed to characterize protein post-translational modifications of the above carbon and electron transfer pathways to decipher their relationship with the metabolism of syntrophic microbial communities. Acyl-lysine modifications affecting protein function are among the most striking findings from the proteome analysis of the model syntrophic bacteria. The abundance of such acylations in *Syntrophus aciditrophicus* and *Syntrophomonas wolfei* is sufficient to circumvent the pan-specific antibody enrichment steps (and their biases) typically employed in acylation studies. By analyzing the identity of the modified proteins, the sites of modification, and the composition of the modifications, we glimpse at how thermodynamically-challenged organisms can employ reversible catalysis to survive, balancing chemical degradation, synthesis, and excess reducing equivalents. Another feature of our studies is the development of mass spectrometry methods based on immonium ions that confidently identify acyl modifications in proteomes.

Cellulosome assembly and display in cellulolytic anaerobic bacteria.

In complementary microbial studies we are investigating how highly cellulolytic anaerobic bacteria 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 that control gene expression by sensing extracellular polymers. Here we present our progress toward understanding the function of RsgI9, a novel anti-σ factor whose function has yet to be determined. Using a combination of NMR spectroscopy, X-ray crystallography and biochemical assays, we have discovered that RsgI9's ectodomain comprises S1C protease-like and NTF2-like domains that cooperatively bind to cellulose. Here we present the atomic structures of RsgI9's ectodomain and solution-state characterization using small-angle X-ray scattering (SAXS), and our ongoing work to define RsgI9 regulates cellulase gene transcription. We hypothesize that cellulolytic anaerobic bacteria assemble and display cellulosomes using a conserved molecular pathway. We are therefore 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.

Our activities interface with other UCLA DOE Institute technical and plant research groups to complement research activities here and at the DOE National Laboratories and Bioenergy Research Centers in translational science applications.

References/Publications:

1. Muroski JM, Fu JY, Nguyen HH, Ogorzalek Loo RR, and Loo JA. “Leveraging immonium ions for targeting acyl-lysine modifications in proteomic datasets.” Proteomics, in press. (https://doi.org/10.1002/pmic.202000111)

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.

Automated knowledge harvesting from literature text, tables, and figures using natural language processing and machine learning

Shinjae Yoo,[1,*] (sjyoo@bnl.gov), Ian Blaby[2], Sean McCorkle[1], Gilchan Park[1], and Carlos Soto[1]

[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. Since the cost (financial and time) of investigations seeking to capitalize on genome-enabled organisms by biological redesign to meet BER goals, 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, and 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. Here, **we focus on three subproblems**, their associated challenges, and our approaches, methods, and results in addressing them in this work: 1) ML for **processing non-machine-readable documents**, 2) NLP for **identifying protein entities and relationships between them in the text**, and 3) ML for **automated information extraction from tables and figures**.

Most published scientific works are available as PDFs – either as scans of old printed manuscripts or as digitally-sourced documents. These are readily accessible by human readers, but unfortunately cannot be processed automatically by computers. The publications' titles, authors, and abstracts may be indexed for digital search, but it remains relatively uncommon in most scientific fields to publish full-text articles in machine-readable formats. To alleviate this limitation, we developed a document processing pipeline that leverages ML techniques originally designed for object detection to visually segment the salient regions of a PDF article. The ML method was trained to recognize and isolate figures, tables, captions, main body text, and other document components for downstream processing by further specialized techniques. Our method achieved 80% - 94% detection accuracy on major region classes after training on a relatively small 100-document novel annotated dataset.

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. This effort includes ongoing annotation of a novel dataset for this purpose, which currently has over 400 entries and which we expect to quickly grow to over 1000. Our model currently achieves over 85% accuracy in identifying protein-protein interactions in the text.

Finally, 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 ability to identify data plots of interest and automatically extract the data values they contain. For this purpose, we are developing a two-stage detection model and a novel value extraction model. Both of these efforts are in early to intermediate stages but we have already demonstrated up to 88% accuracy in table identification, 45% raw table structure recognition accuracy (before post-processing), and 73% sub-figure detection accuracy.

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.

Rapid Strain Phenotyping and Metabolic Flux Analysis to Accelerate Engineering of Microorganisms.

Berkley Ellis[1-5*] (Berkley.ellis@vanderbilt.edu), Piyoosh Babele[6,7], Joshua Abraham[8], Brian F. Pfleger[8], Jamey D. Young[6,7], **John A. McLean**[1-5]

[1]Department of Chemistry, Vanderbilt University, Nashville, TN USA; 2Center for Innovative Technology, Vanderbilt University, Nashville, TN; [3]Institute of Chemical Biology, Vanderbilt University, Nashville, TN; [4]Institute for Integrative Biosystems Research and Education, Vanderbilt University, Nashville, TN; [5]Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN; [6]Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN,; [7]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; [8]Department of Chemical and Biological Engineering, University of Wisconsin, Madison, WI

Project Goals: This project aims to comprehensively enhance microbial production of commodity chemicals. While portions of this project are directed towards increasing titers by engineering pathways and manipulating the circadian clock, this work is focused on the technologies evaluating strain production. We describe the development and application of analytical methods measuring metabolism that comprehensively characterize the outcomes of specific editing events and elucidate targets for further optimization. By increasing throughput and molecular breadth of analytical measurements assessing phenotypes, we aim to accelerate engineering and ultimately commercial utility of bacterial constructs.

Reading and editing DNA sequences are no longer the rate-limiting steps in microbial strain engineering, rather it is characterizing the metabolic outcomes of a specific genetic editing event.[1,2] To address this issue, we have developed a desorption electrospray ionization- imaging mass spectrometry (DESI-IMS) method that simultaneously samples various strains and biosynthetic products. Our DESI-IMS workflow is performed under ambient conditions with minimal sample preparation. The sole sample preparation step is adhering a membrane scaffold to a glass slide on which microorganisms and their metabolites are retained. By preparing many strains (samples) in a single step, we facilitate a robust interrogation of relative metabolite abundances and reduce the time and error associated with sample preparation by alleviating the need for pipetting and extractions. We demonstrate the inherent multiplexing capabilities of IMS by simultaneously characterizing various Escherichia coli strains engineered for free fatty acid (FFA) production and their respective biosynthetic products.

Using the developed workflow, we can phenotype engineered strains on the basis of biosynthetic products and also across the measured metabolome via untargeted IMS acquisitions. This is a necessary innovation in analytical measurements of phenotypes, because annotating global metabolism in addition to targeted metabolites typically require additional experiments.[2] To enable comprehensive analyses, we developed an unbiased data analytics workflow using unsupervised segmentation. With this workflow, we establish the ability to phenotype various engineered E. coli strains and differentiate them on the basis of comprehensive metabolomic measurements. We demonstrate the advantages of both untargeted acquisitions and unsupervised data analytics by characterizing secondary fatty acid production and providing insight into the

prevailing biology of microorganisms via variations in membrane lipid saturation. In sum, we establish that the developed workflow can accelerate synthetic biology strategies with applications in directed evolution, functional genomics, and metabolic flux studies.

Work within this project has also been directed towards developing novel FFA producing cyanobacteria, which fix atmospheric CO_2 and does not compete with arable land. We developed strains of Synechococcus sp. PCC7002 that produce up to 860 mg/L octanoic acid (C8) titers. Our future plans are to determine the metabolic bottlenecks inhibiting higher FFA production. Stable isotope (13C) based nonstationary metabolic flux analysis (INST-MFA) analyses of FFA producing Synechococcus PCC 7002 strains will help us to identify the bottleneck pathway for further genetic modifications to improve titers.[3] We will also use the developed DESI-IMS method to quickly determine high performing fatty acid strains and identify favorable mutations.

Our studies demonstrate the importance of DESI-IMS and INST-MFA. The development of these technologies has implications in the field of synthetic biology by providing rapid analytical readouts on metabolic production and efficiency. Further, future work incorporating MFA within DESI-IMS workflows presents a technology capable of rapid annotation of metabolism which might be crucial for the engineering of metabolically efficient microorganisms for industrial applications.

References

1. Garst, A. D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic, and genome engineering. Nat. Biotechnol. 35, 45-55 (2017).
2. Industrialization of biology: a roadmap to accelerate the advanced manufacturing of chemicals. Washington, D.C.: The National Academies Press. xvi. (2015)
3. Jazmin, L. J. et al. Isotopically nonstationary 13C flux analysis of cyanobacterial isobutylaldehyde production. Metab. Eng. 42, 9-18 (2017).

This work was supported by Department of Energy award DE-SC0019404

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.
- **References** can be **Publications** instead, if needed. Use any common style for these citations.

Mitigating guanidine toxicity and manipulating circadian clock for enhanced ethylene production in engineered cyanobacteria

Bo Wang[1,2*](bo.wang.2@vanderbilt.edu), Yao Xu[3], Xin Wang[4,5], Joshua S. Yuan[4], **Carl H. Johnson**[3], **Jianping Yu**[2], **Jamey D. Young**[1,6]

[1] Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN 37235, USA. [2] Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401, USA. [3] Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA. [4] Synthetic and Systems Biology Innovation Hub, Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX 77843, USA. [5] Current address: Department of Microbiology, Miami University, Oxford, OH 45056, USA. [6] Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37235, USA.

We are developing cyanobacteria as optimized cell factories for producing biofuels and other renewable biochemicals. Our *long-term goal* is to develop technologies that can be used to optimize cyanobacteria and other microbes for producing renewable chemicals at commercially feasible rates and yields. One objective of our project is to achieve sustained production of ethylene in the engineered *efe*-expressing cyanobacterium *Synechococcus elongatus* through removal of the toxic byproduct guanidine. A second objective is to manipulate the biological clock in the engineered ethylene-producing and guanidine-degrading strains to further enhance and stabilize photosynthetic production of ethylene.

Due to biotechnological interests in developing an alternative pathway for renewable production of ethylene, which is the most highly produced organic compound in the petro-chemical industry, the *efe* gene (encoding the ethylene-forming enzyme; EFE) from *Pseudomonas syringae* (a plant pathogen) has been introduced into a variety of microbial species[1]. Some hosts, *e.g.*, *Pseudomonas putida* KT2440 and the cyanobacterium *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis* 6803), have been able to accommodate stable, high-level expression of EFE and thereby sustain enhanced production of ethylene[2]. Other species, such as cyanobacterium *Synechococcus elongatus* PCC 7942 (hereafter *Synechococcus* 7942), however, have not been able to tolerate high-level expression of EFE, and the recombinant strains suffered severe growth inhibition[3, 4] that was rescued by spontaneous chromosomal mutations that abolished the expression of functional EFE and therefore ethylene formation[3, 4].

In this study, we show that accumulation of guanidine (a byproduct of biological ethylene production) significantly inhibits the growth of cyanobacterial cells, and destabilizes their genome in response to recombinant EFE expression. We found that Sll1077, previously annotated as an agmatinase in the arginase superfamily in *Synechocystis* 6803[5, 6], is more likely a "guanidinase", because it degrades guanidine rather than agmatine to urea. This result is consistent with the finding that there is a conserved sequence motif of the guanidine riboswitch upstream of the *sll1077* ORF. *Synechococcus* 7942 lacks a homologous enzyme in its genome and is unable to mitigate guanidine toxicity. Heterologous expression of Sll1077 in a recombinant *Synechococcus* 7942 strain confers the ability to degrade guanidine into non-toxic urea. Co-expression of Sll1077 and EFE in *Synechococcus* 7942 stabilizes the genome of the resultant strain and leads to sustained production of ethylene from light and CO_2.

Next, we sought to further enhance the bioethylene production through manipulating the biological clock. The cyanobacterial circadian clock exerts control over global gene regulation and oscillations of intracellular metabolism. Our previous studies have proven that overexpression of a key clock positive regulator could constantly enhance expression of both endogenous and exogenous genes in *Synechococcus* 7942[7]. We recently found that mutations of clock negative regulators and deletion of some clock output pathway players could even further enhance expression of genes of interest. Our preliminary results showed that ethylene production in the engineered *Synechococcus* 7942 strain is under regulation of the circadian clock. Based on these findings, we have generated a series of engineered *Synechococcus* 7942 strains, and further characterization is underway. If successful, manipulation of the circadian clock genes could be a promising approach for enhancing production of biofuels and chemicals in cyanobacteria.

References

1. C. Eckert, W. Xu, W. Xiong, S. Lynch, J. Ungerer, L. Tao, R. Gill, P. C. Maness and J. Yu, *Biotechnol Biofuels*, 2014, **7**, 33.
2. B. Wang, C. Eckert, P. C. Maness and J. Yu, *ACS Synth Biol*, 2018, **7**, 276-286.
3. M. Sakai, Ogawa, T., Matsuoka, M., Fukuda, H., *Journal of Fermentation and Bioengineering*, 1997, **84**, 434-443.
4. K. Takahama, M. Matsuoka, K. Nagahama and T. Ogawa, *J Biosci Bioeng*, 2003, **95**, 302-305.
5. A. Sekowska, A. Danchin and J. L. Risler, *Microbiology (Reading)*, 2000, **146 (Pt 8)**, 1815-1828.
6. J. W. Nelson, R. M. Atilho, M. E. Sherlock, R. B. Stockbridge and R. R. Breaker, *Mol. Cell*, 2017, **65**, 220-230.
7. Y. Xu, P. D. Weyman, M. Umetani, J. Xiong, X. Qin, Q. Xu, H. Iwasaki and C. H. Johnson, *Current biology : CB*, 2013, **23**, 2365-2374.

This work was supported by Department of Energy award DE-SC0019404.

Development of untargeted metabolomics approaches to study bacterial-fungal co-cultures

Amy Zheng[1], Baltazar Zuniga[1], Fabio Palmieri[2], Simone Lupini[3], Berkley Ellis[4], Debora F. Rodrigues[3], Pilar Junier[2], Patrick Chain[5], **Jamey Young[1,6]**

[1]Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee; [2]Laboratory of Microbiology, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland; [3]Civil and Environmental engineering, University of Houston, Houston, Texas; [4]Chemistry, Vanderbilt University, Nashville, Tennessee; [5]Biosciences Division, Los Alamos National Lab, Los Alamos, New Mexico; [6] Molecular Physiology and Biophysics, Vanderbilt University, Nashville Tennessee

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 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 to understand these species-dependent interactions. In the future, we plan to compare metabolism of multiple different co-culture pairs and identify the impact of growth conditions on their interactions using both Gas Chromatography-Mass Spectroscopy (GC-MS) and high-resolution Liquid Chromatography-Mass Spectroscopy (LC-MS) or Desorption Electrospray Ionization-Imaging Mass Spectrometry (DESI-IMS) untargeted metabolomics platforms.

We are developing a data analysis workflow for annotating biologically relevant compounds detected in our samples. We aim to solve two major problems. First, this workflow will be able to process and quantify mass spectra detected using multiple platforms (DESI-IMS, GC-MS, and LC-MS). Second, it will identify conserved metabolites by simultaneously cross-referencing putative targets against multiple spectral libraries, metabolic pathway databases and organismal metabolomes. To the best of our knowledge, no other software possesses these functionalities. This approach will streamline our workflow by eliminating biologically irrelevant targets and aid data interpretation by identifying pathways in our organisms of interest impacted by co-culture conditions. Additionally, the unique ability to simultaneously search multiple related metabolomes is ideal for identifying biologically relevant metabolites in experiments like ours that utilize environmentally derived organisms that may have uncertain phylogeny.

To analyze spatial profiling of extracellular metabolites, we are developing a metabolomic analysis pipeline using DESI-IMS, which uses an electrospray mechanism to ionize metabolites from surfaces under ambient conditions. These methods are being applied to study the interactions of the oxalogenic fungus *Aspergilllus niger* (*A. niger*) with the bacteria *Pseudomonas putida* (non-oxalotrophic*)* and *Cupriavidus oxalaticus* (*oxalotrophic*). Our hypothesis is that the interactions between this fungal and these bacterial strains are dependent

on the spatial distribution of oxalic acid. produced by the fungus and consumed only by the oxalotrophic bacterium. This interaction depends primary on the effect of this low molecular organic acid on the pH of the growth medium. *A. niger* produces oxalic acid, which lowers the pH of the medium. *C. oxalaticus* metabolizes oxalic acid and thereby antagonizes *A. niger* growth by raising the local pH in the vicinity of the bacterial colonies. In contrast, *P. putida* coexists with *A. niger* without altering the pH of the medium. We have developed and are currently testing a DESI-IMS strategy to spatially profile the concentration of oxalic acid and other metabolites by growing 2D co-cultures on semi-permeable membranes placed atop agar plates. Using this approach, we expect to investigate the nature of the metabolic interactions between these species.

By developing new software tools and experimental pipelines, we expect to better understand the microbial metabolic processes that occur at the interface of bacterial-fungal co-cultures.

This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59T.

Title: Integration of Experimental and Modeling Approaches to Understand, Predict, and Modulate Rhizosphere Processes for Improved Bioenergy Crop Productivity

Authors: Oriane Moyne[1], Amelie Gaudin[2], Yuntao Hu[3] Trent Northen[3], John Vogel[3], **Karsten Zengler[1]**

Institutions: [1]University of California, San Diego, [2]University of California, Davis, [3]Lawrence Berkeley National Laboratory,

Project Goals: Short statement of 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*. The results will be used to predict plant mutants and microbial amendments, which improve low-input biomass production and validate these predictions in lab and field studies. We will determine the mechanistic basis of dynamic exudate exchange in the grass rhizosphere. A specific focus will be 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 nitrogen use efficiency (NUE). To do this we will use stable isotope dilution mass spectrometry methods to quantify nitrogen (N) flow between plants and microbes and use plant mutants to identify plant genes that influence exudation. Furthermore, we will develop two predictive genome-scale rhizosphere models (*S. bicolor* and *B. distachyon*) that are grounded in lab and field data. These models will be used to predict plant growth and exudate metabolism to maximize N flux to the plant as a function of light and N availability over a diurnal cycle. Finally, we will use these models to design genetic and microbial interventions that are tailored to increase plant biomass production under N limited conditions by optimizing plant-microbe interactions. We will also design and test the effects of amendments added at different times of the day and night to increase plant productivity by altering the microbiome to alleviate N limitation. The most favorable and robust approaches will be tested under marginal field conditions.

Abstract text: To avoid competition with food crops and maximize their economic and environmental benefits, bioenergy crops should be grown on marginal soils with minimal inputs, especially inputs of energy intensive synthetic N fertilizer. Thus, bioenergy crops are more reliant on microbial-driven organic N for nutrition than food crops. Root exudates are thought to play a critical role in recruiting and maintaining beneficial microbes including those that make N available to the plant. Up to 25% of all photosynthetically fixed carbon is released into the rhizosphere through root exudates. These exudates provide a critical source of nutrition for diverse microorganisms and are composed of many compounds including organic acids, amino acids, and mono- and polysaccharides. However, it is unclear if this exchange of exudates for plant benefits holds during the night, when the plants are not photosynthesizing. Our preliminary data suggest that exudate composition shows strong circadian rhythms. 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. Predictive models are necessary to understand the complexity of these processes and optimize sustainable bioenergy systems. We hypothesize that exudate and bacterial community engineering will enable structuring of rhizosphere communities that enhance N supply and/or reduce microbial competition for organic N thereby increasing plant NUE.

Funding statement: This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under Award DE-SC0021234,

Improved Biofuel Production through Discovery and Engineering of Terpene Metabolism in Switchgrass

Kira Tiedge[1]*(kjtiedge@ucdavis.edu), Andrew Muchlinski[1]* (amuchlinski@ucdavis.edu), Yuxuan Chen, Danielle Davisson, **Philipp Zerbe**[1]

[1]University of California, Davis

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 terpenoids 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 terpenoids. This project merges genome-wide enzyme discovery with comparative –omics, protein structural and plant microbiome studies to define the biosynthesis and stress-defensive functions of switchgrass (*Panicum virgatum*) terpenoid metabolism. These insights will be combined with the development of genome editing tools to design plants with desirable terpene blends for improved biofuel production on marginal lands.

Abstract:

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 study reports the genome-wide discovery and functional characterization of terpenoid-metabolic genes, enzymes and pathways toward a diverse range of diterpenoids in the bioenergy crop switchgrass (*Panicum virgatum*). Biochemical analysis of several dozen enzymes revealed a modular metabolic network producing a diverse array of diterpenoid metabolites several of which perhaps uniquely occurring in switchgrass. Structure-guided protein mutagenesis of select enzymes identified active site determinants that may resemble neo-functionalization events that occurred during diversification of the switchgrass diTPS family and offer resources for engineering of bioactive metabolites. Combined transcriptomics and metabolomics studies identified multiple enzyme products in switchgrass root and leaf tissue with, albeit moderate, accumulation in response to oxidative and drought stress, thus supporting a possible physiological functions in the plants adaptation to abiotic stressors. Cultivation and analysis of switchgrass in small controlled environment devices (EcoFABs) now enabled a precise analysis of the role of switchgrass root diterpenoids in plant-environment interactions.

References:

1. Tiedge K, Muchlinski A, Zerbe P (2020) Genomics-enabled analysis of specialized metabolism in bioenergy crops: current progress and challenges. *Synth Biol* 5:ysaa005.

2. 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.
3. Muchlinski A, Chen X, Lovell JT, Köllner TG, Pelot KA, Zerbe P, Ruggiero M, Callaway L 3rd, Laliberte S, Chen F, Tholl D (2019) Biosynthesis and Emission of Stress-Induced Volatile Terpenes in Roots and Leaves of Switchgrass (Panicum virgatum L.). *Front Plant Sci*. 10:1144.

Funding:

The authors' work has been supported with funds from U.S. Department of Energy (DOE) Early Career Research Program (DE-SC0019178, to P.Z.); DOE Joint Genome Institute (JGI) DNA Synthesis Science Program (grant #2568, to P.Z.); German Research Foundation (DFG) Research Fellowship (TI 1075/1-1, to K.T.).

Leveraging comparative population genomics to dissect the mechanisms of multi-stress tolerance of *I. orientalis*

Yusuke Sasaki[1,2](yusukesasaki@lbl.gov), Ping-Hung Hsieh[1,2], Jing Ke[1,2], Zhying Zhao[1,2], Zong-Yen Wu[1,2], Andrei Stecca[1,2], Sajeet Haridas[1,2], Zia Fatma[3], Huimin Zhao[3], and **Yasuo Yoshikuni**[1,2]

[1]Department of Energy Joint Genome Institute, [2]Lawrence Berkeley National Laboratory, [3]University of Illinois at Urbana-Champaign

Project Goals: Dissect phenotypic variations of 170 *I. orientalis* strains under major stress conditions through statistical analysis

Abstract:

Issatchenkia orientalis has a distinguished capability of tolerance to multiple environmental factors (e.g., low pH, heat, and chemicals consisting of lignocellulosic components) and antifungal drugs (e.g., azoles, echinocandins, and polyenes). These unique features have attracted attention to *I. orientalis* to be utilized as a next generation industrial chassis capable of producing biofuels and bioproducts from crude lignocellulosic hydrolysates. Large-scale population genomic studies will provide comprehensive views of *I. orientalis* genetic diversity, population structure, and genotype-phenotype associations that open further opportunities such as, rational strain engineering and evolutionary lineage analysis of the yeast. In contrast to the previous population genomics study of *I. orientalis* that sequenced 32 isolates (Douglass et al., 2018), we performed deep coverage genome sequencing of 170 isolates and assessed their phenotypic traits under 58 conditions. Genome sequencing of 170 *I. orientalis* isolates collected from various habitats identified 305,405 single nucleotide polymorphism (SNPs), 16,177 insertions and deletions (INDELs), and other genetic variations (e.g., ploidy, gene copy number, and pan-genome). Phenotypes under major stress conditions (9,338 variations) were quantitatively assessed by measuring cell growth changes with an automation-based high-throughput assay. We have completed the phenotypic assays under 58 conditions and collected 9,338 traits. Correlation analysis of the data shows that significant correlations with several types of azole drugs. Clustering and community detections allows to categorize each condition and strain based on the phenotypic traits, which provides a new insight of strain classification. As our future study, genome-wide association study (GWAS) allows to capture genetic variants shaping the architecture of the phenotypic traits and genetic engineering of *I. orientalis* for validation of the variants identified by GWAS proves the concept of our strategy.

References:

A.P. Douglass, B. Offei, S. Braun-Galleani, A.Y. Coughlan, A.A.R. Martos, R.A. Ortiz Merino, K.P. Byrne, K.H. Wolfe, "Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: one species, four names" PLoS Pathog., 14 (2018), p. e1007138

Funding Statement:

This work was funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-FOA-0001650. 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.

Metabolic regulation in *Issatchenkia orientalis* revealed by integrative omics

Yihui Shen[1]* (yihuis@princeton.edu), Hoang Dinh[2], Catherine Call[1], Rolf-Peter Ryseck[1], Zia Fatma[3], Vinh Tran[3], Heide Baron[4], Patrick Suthers[2], Huimin Zhao[3], Martin Wühr[4], Costas Maranas[2], **Joshua Rabinowitz[1]**

[1]Department of Chemistry and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ
[2]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA
[3]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL
[4]Department of Molecular Biology & the Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ

https://www.igb.illinois.edu/research-areas/gsecenter

Project Goals: Understand metabolic regulation in I. orientalis

Omics technologies hold the potential to rapidly understand less studied organisms. Here we examine the potential to integrate proteomics, metabolomics, and fluxomics to dissect metabolic regulation in a non-model yeast of industrial metabolic engineering interest, *Issatchenkia orientalis*. Despite diverging from *Saccharomyces cerevisiae* 250 million years ago, *I. orientalis* responds similarly to nutrient limitation. Systematic identification of meaningful metabolic enzyme regulation from the multi-omics data revealed similar metabolic regulatory logic to the model yeast. This logic is, however, implemented through different allosteric regulatory events, suggesting convergent evolution. Among these is inhibition of the glycolytic enzyme glyceraldehyde dehydrogenase by ATP, which we also verified biochemically. By understanding the specific enzymatic regulatory events controlling metabolic flux in *I. orientalis,* we lay the groundwork for future rational and efficient engineering. More generally, we demonstrate the capacity for integrated omics to rapidly advance metabolic regulatory understanding in less studied microbes.

This work is supported by U. S. Department of Energy Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018260.

Model-driven Metabolic Engineering and 13C-Metabolic Flux Analysis for Non-model Yeast Organisms

Patrick F. Suthers[1*] (pfs13@psu.edu), Hoang V. Dinh[1*], John I. Hendry[1], Zia Fatma[2], Yihui Shen[3], Tianxia Xiao[3], Joshua D. Rabinowitz[3], 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; [3]Princeton University, Princeton, New Jersey

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 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 production of oleaginous compounds and *Issatchenkia orientalis* for 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 genome-scale. The final goal is to develop kinetic models to guide metabolic engineering accounting for reaction kinetics and allosteric regulations.

Unique metabolic capabilities and resilience to inhibitory stressors enable some non-model yeasts to be attractive microbial cell factories. *Rhodosporidium toruloides* is a basidiomycetes yeast that can accumulate large amount of lipids, and *Issatchenkia orientalis* is a promising host for industrial production of organic acids because of its low-pH tolerance. To better assess their metabolic capabilities and to draw comparisons with the model yeast *Saccharomyces cerevisiae*, we reconstructed separate genome-scale metabolic (GSM) models for each non-model yeast. These curated GSM models drew on in-house-measured macromolecular compositions and chemostat growth data which enabled estimating ATP maintenance requirements, and we performed model validation experiments. We applied these GSM models to make suggestions on genetic modifications to bolster targeted product formation for succinic and itaconic acids, as well as to examine the dependence of product formation on oxygen uptake levels. Upon the GSMs, we built carbon mapping models for ^{13}C-metabolic flux analysis (^{13}C-MFA) at genome-scale with labeling data using the tracers U-^{13}C-glucose and 1,2-^{13}C-glucose. Expanding to genome-scale accounts for cofactor balance reveals alternative flux distribution in central metabolism and supplies flux value for biosynthetic reactions. In the future, ^{13}C-MFA fluxes will be used in strain designs, systematic identification of allosteric regulations, and kinetic model development.

This work is supported by U. S. Department of Energy Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018260.