U.S. DEPARTMENT OF
ENERGY

2020 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 20, 2020

Dear Colleague:

Welcome to the 2020 Genomic Science Annual PI Meeting! On behalf of the entire program staff of BER's Biological Systems Science Division (BSSD) I'd like to thank you for the success of the Genomic Science Program (GSP). Your scientific creativity, insight, and experimental results are what sustains this program.

From the program level, biological research continues to capture attention. We've seen much discussion in policy circles on how to accelerate a burgeoning bio-based economy, be more competitive in the global biotechnology arena, and lead development of next generation technologies. DOE has a significant role to play in all of these areas and this program will likely form a cornerstone of related efforts. The GSP and the BSSD portfolio with its combination of basic research, enabling capabilities and user facilities, is poised to address a wide range of research challenges in the bioeconomy, biotechnology, and next-gen technologies space, should we be afforded the opportunity to do so. We are optimistic about the future and how the Division is currently positioned to address national priorities and you should be too.

We have a full agenda again this year highlighting some exceptional research results. As always, we hope that the presentations will spark fruitful discussions among attendees. The PI meeting is an important mechanism for our staff to become aware of results from projects funded within the program. We also hope that it helps researchers connect with others working in complementary areas. Please use this opportunity to seek out and meet your colleagues.

We are pleased to have Dr. Tom Juenger (University of Texas, Austin) as our keynote speaker this year. Dr. Juenger's research touches on many facets of the GSP portfolio, and we know that his talk will have broad appeal. We are also pleased to welcome new early career awardees, new plant genomics and environmental microbiome researchers, and new computational biology and secure biosystems design pilot project researchers to this year's meeting.

The meeting agenda includes talks from each element of the program, beginning with the Bioenergy Research Centers and continues across the portfolio in several breakout sessions throughout the meeting. There are also some interesting plenary sessions from the DOE User Facilities and community resources highlighting new capabilities available to researchers in the portfolio. Of particular interest is a new plenary on making better use of biological data.

Finally, a reminder that we hope you treat this meeting as an opportunity to view your research in the broad context of the larger program. All principal investigator projects from our University and DOE National Laboratory portfolios, as well as User Facilities and enabling capabilities funded by the program are present at this meeting. Please take full advantage of this opportunity to meet with your colleagues and with your DOE program staff, representatives from elsewhere within the Department of Energy, and colleagues from other Federal Agencies.

Thank you again for making the GSP the success that it is. We look forward to an excellent meeting!

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.)

A - B - C - D - E - F - G - H - I - J - K - L - M - N - O - P - Q - R - S - T - U - V - W - X - Y - Z
(Click on letter above to advance to the alpha section of book.)

A

Paul Adams: Integrating Genomics, Physiology and Genetics of the Oak Ridge Field Research Site Microbiome
PRESENTER: Adam Deutschbauer

Paul Adams: Selective Carbon Sources Influence the End-Products of Microbial Nitrate Respiration
PRESENTER: Hans Carlson

Paul Adams: Spatiotemporal Dynamics of Groundwater and Sediment: Geochemistry, Microbial Communities and Activities in a Contaminated Aquifer
PRESENTER: Katherine Walker

Paul Adams: The ENIGMA Data Clearinghouse: A platform for rigorous self-validated data modeling and integrative, reproducible data analysis
PRESENTER: John-Marc Chandonia

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

Andrew Allen: Evolution and Metabolic Configuration of Nitrogen Flux in a Model Marine Diatom
PRESENTER: Sarah Smith

Andrew Allen: Identification of transcription factor binding sites and characterization of promoter architecture in the model diatom Phaeodactylum tricornutum
PRESENTER: Andrew Allen

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

Andrew Allen: Temporal profiles unravel resource allocation mechanisms under nitrogen starvation in the diatom Phaeodactylum tricornutum
PRESENTER: Cristal Zuniga

Steven Allison: Prolonged drought alters plant-litter decomposition via changes in bacterial communities and substrate availability
PRESENTER: Sarai Finks

Ana Alonso: Functional Analysis of Candidate Genes Involved in Oil Storage and Stability in Pennycress
PRESENTER: Emmanuel Ortiz

Ana Alonso: Natural Variation of Pennycress Metabolome and Transcriptome, an Emerging Crop for Aviation Biofuel
PRESENTER: Cintia Arias

Daniel Amador-Noguez: In vivo thermodynamic analysis of glycolysis in C. thermocellum and T. saccharolyticum using 13C and 2H tracers
PRESENTER: Tyler Jacobson

James A Anderson: Investigating Seed Size and Oil Content in Pennycress, Thlaspi arvense
PRESENTER: Katherine Frels

Adam Arkin: KBase: Omics driven discovery of novel functional capabilities in biological systems
PRESENTER: Janaka Edirisinghe

Adam Arkin: KBase: Leveraging Amplicon Analysis Tools to Generate Testable Hypotheses From Complex Natural Communities
PRESENTER: Pamela Weisenhorn

Adam Arkin: KBase: Microbiome and Phylogenomics Capabilities
PRESENTER: Dylan Chivian

Adam Arkin: KBase: The Systems Biology Knowledgebase for Predictive Biological and Environmental Research in an Integrated Data Platform
PRESENTER: Benjamin Allen

Adam Arkin: Large Scale Model-Driven comparison of Metagenome Assembled Genomes from Diverse Environments
PRESENTER: José Faria

Adam Arkin: NWChem Computational Modeling of Metabolites in KBase
PRESENTER: Neeraj Kumar

Adam Arkin: The 2019 KBase Fungal Biochemistry Curation Jamboree: Insights and Lessons Learned
PRESENTER: Jeremy Zucker

Jose Avalos: Genetically Encoded Biosensors for Mitochondrial and Cytosolic Biosynthesis of Branched-Chain Higher Alcohols in Saccharomyces cerevisiae
PRESENTER: Jose Avalos

B

Jill Banfield: Rapid automated curation of genomes from metagenomes and integration into KBase
PRESENTER: Rohan Sachdeva

Rebecca Bart: Optimizing Tradeoffs Implicit During Bioenergy Crop Improvement: Understanding the Effect of Altered Cell Wall and Sugar Content on Sorghum-associated Pathogenic Bacteria
PRESENTER: Rebecca Bart

Ivan Baxter: New Computational Pipelines to Prioritize Candidate Genes for Optimal Biomass Production under Drought in C4 Plants Setaria viridis and Sorghum bicolor
PRESENTER: Seung Rhee

Ivan Baxter: PEPC kinetics and the efficiency of C4 photosynthesis in Sorghum bicolor
PRESENTER: Asaph Cousins

Ivan Baxter: Phenomics of stomata and water use efficiency in C4 species
PRESENTER: Andrew Leakey

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

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

Ivan Baxter: Using synthetic genetic circuits to tightly control root architecture.
PRESENTER: Jose Dinneny

Michael Betenbaugh: Metabolic Flux Analysis of Sucrose-Secreting Cyanobacterium Synechococcus elongatus
PRESENTER: Cristal Zuniga

Michael Betenbaugh: Optimizing Carbon Metabolism in Co-Culture for Applications to Sustainable Biosynthesis
PRESENTER: Jackson Jenkins

Jennifer M Bhatnagar: Ectomycorrhizal fungi: mediators of plant-microbial interactions and terrestrial Biogeochemistry
PRESENTER: Jennifer M Bhatnagar

Crysten Blaby: Developing a molecular-level model of cofactor-trafficking in chloroplasts
PRESENTER: Crysten Blaby

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

Eduardo Blumwald: Molecular Regulation of Cell-type Specific Responses to Abiotic Stresses in Poplar
PRESENTER: Eduardo Blumwald

Eduardo Blumwald: Rational design and testing of osmotic-stress inducible synthetic promoters from poplar cis-regulatory elements
PRESENTER: Neal Stewart

Nicholas Bouskill: Microbial environmental feedbacks and the evolution of soil organic matter
PRESENTER: Nick Bouskill

Nanette Boyle: Enabling Predictive Metabolic Modeling of Diurnal Growth Using a Multi-Scale Multi-Paradigm Approach
PRESENTER: Nanette Boyle

Daniel Buckley: Harnessing Metagenomic Stable Isotope Probing to Uncover the Carbon Cycling Capacity of Soil Microbes
PRESENTER: Samuel Barnett

Daniel Buckley: Phenolic acid-degrading populations of Paraburkholderia prime decomposition in forest soils
PRESENTER: Roland Wilhelm

Stephen Burley: RCSB Protein Data Bank: Making connections from genes to ecosystems
PRESENTER: John D Westbrook

Kristin Burnum-Johnson: Spatiotemporal Mapping of Lignocellulose Decomposition by a Naturally Evolved Fungal Garden Microbial Consortium
PRESENTER: Kristin Burnum-Johnson

Posy Busby: Identifying Plant Genes Associated With Pathogen Antagonism in Populus trichocarpa
PRESENTER: Posy Busby

C

William Cannon: Elucidating Principles of Bacterial-Fungal Interactions
PRESENTER: William Cannon

William Cannon: LEARNING REGULATION AND OPTIMAL CONTROL OF ENZYME ACTIVITIES AND APPLICATION TO SYSTEMS BIOLOGY DATA OF NEUROSPORA CRASSA
PRESENTER: William Cannon

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

Patrick Chain: Examinations of the Fungal Genus Monosporascus Reveal its Potential as an Experimental Model for Studying Bacterial-Fungal Interactions
PRESENTER: Aaron Robinson

Patrick Chain: Integrating read-based microbiome taxonomy classification tools into KBase
PRESENTER: Mark Flynn

Patrick Chain: Investigating Chloroplast Signal within the Hyphae of Diverse Fungi
PRESENTER: Julia Kelliher

Patrick Chain: Resolving Intracellular Chloroplast in Fungi from Sequence to Slide.
PRESENTER: Demosthenes Morales

Clint Chapple: Coupling Metabolic Source Isotopic Pair Labeling And Genome Wide Association For Metabolite And Gene Annotation In Plants
PRESENTER: Clint Chapple

George Church: Characterizing the portability of RecT-mediated oligonucleotide recombination
PRESENTER: Gabriel Filsinger

George Church: Deep Mutational Learning of Protein Function for New Intracellular Biosensors
PRESENTER: Alexander Garruss

George Church: Genetic Code Expansion in Bacillus subtilis
PRESENTER: Devon Stork

George Church: High Throughput Functional Variant Screens via In-vivo Production of Single-stranded DNA
PRESENTER: Max Schubert

George Church: Parallelized in vivo Construction of a Synthetic 57-Codon E. coli Genome
PRESENTER: Akos Nyerges

George Church: Uncovering spatial taxonomic structures of synthetic microbial communities using subcellular RNA sequencing
PRESENTER: Andrew Pawlowski

John Coates: Revealing the Prevalence and Diversity of a Rare Phosphorus Metabolism through Selective Enrichments and Genome Resolved Metagenomics
PRESENTER: Sophia Ewens

John Coates: The Lineages of Dissimilatory Phosphite Oxidation Genes Indicate an Ancient, Vertically Transferred Metabolism
PRESENTER: Alexa Gomberg

Luca Comai: Discovery and characterization of disease resistance loci using a unique gene copy number variant population
PRESENTER: Luca Comai

Gloria Coruzzi: EvoNet: A phylogenomic and systems biology approach to identify genes underlying plant survival in marginal, low-Nitrogen soils
PRESENTER: Gil Eshel

Hugo Cuevas: Genome-wide association analysis of anthracnose resistance response in the NPGS sweet sorghum collection
PRESENTER: Hugo Cuevas

D

Gautam Dantas: Characterizing growth and metabolism of Rhodococcus opacus PD630 on real lignin breakdown products
PRESENTER: Garrett Roell

Gautam Dantas: Elucidating aromatic tolerance and utilization in adaptively evolved Rhodococcus opacus strains for lignin valorization
PRESENTER: Winston Anthony

Brian Davison: Pectin – lignin interactions in plant cell walls and model composites
PRESENTER: Hugh O'Neill

Brian Davison: Polymer and structural science behind valorizing lignin using solvents
PRESENTER: Loukas Petridis

Brian Davison: Precision Labeling of Membrane Fatty Acids in Bacillus subtilis and the Impacts on the Cellular Proteome and Lipidome
PRESENTER: James Elkins

Kristen DeAngelis: Drivers and Mechanisms of Long-Term Soil Response to Chronic Warming
PRESENTER: Kristen DeAngelis

Evan DeLucia: A role for differential Rca isoform expression in C4 bioenergy grass thermotolerance?
PRESENTER: Sang Yeol Kim

Evan DeLucia: An Ecosystem-Scale Comparison of Sorghum, Maize, and Miscanthus as Three Bioenergy Crop Candidates
PRESENTER: Caitlin Moore

Evan DeLucia: Application of Spatially Adjusted Machine Learning Approaches to Improve Sorghum Biomass Prediction Using Unmanned Aerial Vehicles
PRESENTER: Sebastian Varela

Evan DeLucia: Assessing Biological Nitrification Inhibition in the Rhizosphere of Field-Grown Bioenergy Sorghum
PRESENTER: Mark Burnham

Evan DeLucia: Building a Comprehensive Set of Genetic Tools for Metabolic Engineering of Issatchenkia orientalis
PRESENTER: Mingfeng Cao

Evan DeLucia: Coarse-grained Modeling of Saccharomyces cerevisiae Physiology
PRESENTER: Viviana Nguyen

Evan DeLucia: Development of High Throughput MALDI-TOF-MS Workflow for Profiling Medium-Chain Fatty Acids from Microbial Colonies
PRESENTER: Kisurb Choe

Evan DeLucia: Development of High Throughput Primer Design and Quantification for Nitrogen Cycle Genes in Bioenergy Crop Soils
PRESENTER: Jaejin Lee

Evan DeLucia: Effectiveness of Payments for Greenhouse Gas Mitigation to Induce Low Carbon Bioenergy Production
PRESENTER: Fahd Majeed

Evan DeLucia: Effects of Natural Variation in Self-Shading on Photosynthetic Traits in Sorghum
PRESENTER: Nikhil Jaikumar

Evan DeLucia: Energizing the Machinery of Storage Lipid Synthesis in Plant Vegetative Tissues
PRESENTER: John Shanklin

Evan DeLucia: Establishment of Efficient Multiplex Genome Editing in Sorghum Using Green Calli
PRESENTER: Praveena Kanchupati

Evan DeLucia: Gene Targeting and Targeted Mutagenesis for Genetic Improvement of Oilcane
PRESENTER: Baskaran Kannan

Evan DeLucia: Leveraging Comparative Population Genomics to Dissect the Mechanisms of Issatchenkia orientalis Fluconazole Resistance
PRESENTER: Yasuo Yoshikuni

Evan DeLucia: Linking Microbial Community Structure and Function for Sustainable Production of Bioenergy Crops
PRESENTER: Angela Kent

Evan DeLucia: Metabolic Engineering of Triacylglycerols with Specialized Fatty Acids in Sorghum
PRESENTER: Kiyoul Park

Evan DeLucia: Oleaginous Yeasts for the Production of Sugar Alcohols
PRESENTER: Sujit Jagtap

Evan DeLucia: Quantifying the Plant-Microbial Interactions Controlling Soil Organic Matter Formation in Bioenergy to Improve Model Representations of Sustainability
PRESENTER: Joanna Ridgeway

Evan DeLucia: Rewiring Metabolism to Construct a Yeast Strain Capable of Producing 2,3-butanediol Without Ethanol and Glycerol Production
PRESENTER: Jaewon Lee

Evan DeLucia: Towards a fully automated algorithm driven platform for biosystems design
PRESENTER: Shekhar Mishra

José Dinneny: Discovering innovations in stress tolerance through comparative gene regulatory network analysis and cell-type specific expression maps.
PRESENTER: Jose Dinneny

Mitch Doktycz: Pant-Microbe interfaces: Application of machine -learned protein-metabolite binding prediction models to plant-microbe interfaces
PRESENTER: Omar Demerdash

Mitch Doktycz: Plant-Micrboe Interfaces: Systems Biology Approaches to Understanding a Plant's Adaptation to and Regulation of the Phytobiome
PRESENTER: Daniel Jacobson

Mitch Doktycz: Plant-Microbe Interfaces: Experimental characterization of protein movement from plants to ectomycorrhizal fungus
PRESENTER: Xiaohan Yang

Mitch Doktycz: Plant-Microbe Interfaces: Identification of gene products involved in plant colonization by Pantoea sp. YR343 using a plant-responsive diguanylate cyclase
PRESENTER: Jennifer Morrell-Falvey

Mitch Doktycz: Plant-Microbe Interfaces: Identification and characterization of Proteolytic Cleavage Product (PCP) peptides that function as key signaling molecules for Plant-Microbe Interactions
PRESENTER: Robert Hettich

Mitch Doktycz: Plant-Microbe Interfaces: Metabolomics of non-host switchgrass plants expressing a poplar lectin receptor-like kinase in response to the mycorrhizal fungus Laccaria bicolor
PRESENTER: Timothy Tschaplinski

Mitch Doktycz: Plant-Microbe Interfaces: Pathway prediction and production through cell-free synthetic biology
PRESENTER: Mitchel Doktycz

Mitch Doktycz: Plant-Microbe Interfaces: Simplified community approach to investigate the dynamic host-microbiome relationship
PRESENTER: Dale Pelletier

Mitch Doktycz: Plant-Microbe Interfaces: Temporal Variation in Plant-Microbe Interactions
PRESENTER: Melissa Cregger

Tim Donohue: Accelerating Yield Improvement in Switchgrass through Genomic Prediction of Floral Anthesis
PRESENTER: Neal Tilhou

Tim Donohue: Associative Nitrogen Fixation (ANF) in High-Yielding Switchgrass Varieties
PRESENTER: Carolina Cordova

Tim Donohue: Crabtree-like aerobic xylose fermentation through increased metabolic flux and altered sugar signaling pathways in Saccharomyces cerevisiae
PRESENTER: Trey Sato

Tim Donohue: Data-driven design and engineering of biomolecules: mRNA and DNA
PRESENTER: Sanjan Gupta

Tim Donohue: Designing Mixed-Solvent Environments for Acid-Catalyzed Biomass Conversion Processes
PRESENTER: Alex Chew

Tim Donohue: Engineering of the Enzymes IspG and IspH from Zymononas mobilis to Increase Terpenoid Production

PRESENTER: Isabel Askenasy

Tim Donohue: Engineering Streptomyces to Capture Value from Lignocellulosic Biofuel Conversion Residue

PRESENTER: Caryn Wadler

Tim Donohue: Incorporation of protocatechuaic acid (3, 4-dihydroxybenzoate) conjugates into the lignin of transgenic poplar

PRESENTER: Faride Unda

Tim Donohue: Integrated spatially explicit optimization of biofuel supply chains and landscape design considering biomass yield uncertainty

PRESENTER: Eric O'Neill

Tim Donohue: Microbial conversion of chemically depolymerized lignin into valuable compounds

PRESENTER: Steven Karlen

Tim Donohue: Microbial Valorization of Lignin: Using Novosphingobium aromaticivorans to Break the Bonds in Lignin and Convert Lignin Deconstruction Products into Value-added Chemicals

PRESENTER: Wayne Kontur

Tim Donohue: Model-driven analysis of mutant fitness experiments improves genome-scale metabolic models of Zymomonas mobilis ZM4

PRESENTER: Dylan Courtney

Tim Donohue: MPK6-MYB46 Regulatory Module Suppresses Plant Biomass Formation During Salt Stress

PRESENTER: Kyung-Hwan Han

Tim Donohue: Pathway Engineering and Re-targeting Boosts Production of High-Value Bioproducts in Plants

PRESENTER: Jacob Bibik

Tim Donohue: Sorghum Dw2 controls stem growth by regulating PLDδ/endomembrane activity and cell proliferation

PRESENTER: Joel Oliver

Tim Donohue: Stability of Switchgrass Leaf Microbiome in the Face of Natural Aerial Colonizers

PRESENTER: Lukas Bell-Dereske

Tim Donohue: Systems Level Comparison of Medium Chain Fatty Acid Production

PRESENTER: Nathaniel Fortney

Tim Donohue: Using fungal diversity to improve biofuel conversion

PRESENTER: David Krause

Tim Donohue: Using the Zip-Lignin Strategy to Build the Optimal Sorghum Biofuel Crop
PRESENTER: Rebecca Smith

John Dunbar: Influence of microbial surface litter decomposer communities on CO2 emissions from natural soils
PRESENTER: Sanna Sevanto

John Dunbar: Merging fungal and bacterial community profiles via an internal control
PRESENTER: Miriam Hutchinson

John Dunbar: Microbial Community Composition Controls Carbon Flux Across Litter Types in Short-Term Litter Decomposition
PRESENTER: Marie Kroeger

John Dunbar: Organism interactions and substrate range are the primary mechanisms linked to divergent carbon flow during litter decomposition
PRESENTER: Michaeline Albright

John Dunbar: Role of geographic scale in likelihood of microbial-driven functional variation during litter decomposition
PRESENTER: Rae DeVan

Mary Dunlop: High-speed spectroscopic stimulated Raman scattering microscopy for measuring biofuel synthesis
PRESENTER: Haonan Lin

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

John Dyer: Genomics and Phenomics to Identify Yield and Drought Tolerance Alleles for Improvement of Camelina as a Biofuel Crop
PRESENTER: Hussein Abdel-Haleem

E

Robert Egbert: Genome remodeling to control the persistence of engineered functions in soil microbes
PRESENTER: Robert Egbert

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

Sarah Evans: Free-living Nitrogen Fixation in the Switchgrass Rhizosphere
PRESENTER: Darian Smercina

Sarah Evans: Relic DNA dynamics mask the resilience of switchgrass bacterial communities to extreme drying rewetting
PRESENTER: Heather Kittredge

Sarah Evans: Soil Microbes Affect Switchgrass Germination More than Seedling Growth Under Drought
PRESENTER: Tayler Ulbrich

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

F

Kjiersten Fagnan: Cultivation-independent expansion of the Nucleocytoplasmic Large DNA Viruses
PRESENTER: Frederik Schulz

Kjiersten Fagnan: The National Microbiome Data Collaborative: Empowering the Research Community to More Effectively Harness Microbiome Data
PRESENTER: Pajau Vangay

Mary Firestone: Arbuscular Mycorrhizal Fungi Transport Water to Host Plants
PRESENTER: Anne Kakouridis

Mary Firestone: Connecting switchgrass-microbe-soil interfaces for sustainable bioenergy crop production on marginal soils: stable-isotope labeling, genomics and exometabolomics
PRESENTER: Mary Firestone

Mary Firestone: Cross-Kingdom Interactions: the Foundation for Nutrient Cycling in Grassland Soils
PRESENTER: Joanne Emerson

Mary Firestone: Effects of Switchgrass Cultivation on Deep Soil Carbon Stock and Long-term Carbon Dynamics in Marginal Lands
PRESENTER: Jialiang Kuang

Mary Firestone: Microbial and Viral Niche-Differentiation in Time-Resolved Metatranscriptomes from Rhizosphere and Detritusphere Soil
PRESENTER: Ella Sieradzki

Mary Firestone: Multitrophic and Metabolite Responses to Drought in Grassland Soils
PRESENTER: Javier Ceja-Navarro

Mary Firestone: Spectroscopic Diagnosis of Plant Phosphorus Availability and Relationship to Tissue Chemistry and Productivity of a Bioenergy Feedstock
PRESENTER: Zhao Hao

Mary Firestone: Succession of Rhizosphere Biotic Communities During Switchgrass Establishment in Marginal Soils
PRESENTER: Yuan Wang

Mary Firestone: Unravelling Rhizosphere-Microbial Interactions in the Rhizosphere of Alamo Switchgrass (Panicum virgatum) under Abiotic Stresses
PRESENTER: Nameer Baker

Brian Fox: Creation of an Acyltransferase Toolbox for Plant Biomass Engineering
PRESENTER: Brian Fox

G

Jeffrey Gardner: Using systems biology to untangle the complex physiology of bacterial xylan utilization
PRESENTER: Jeffrey Gardner

Jean Gibert: PROTIST PREDATION MEDIATES THE TEMPERATURE RESPONSE OF MICROBIAL COMMUNITIES
PRESENTER: Jean Gibert

Ryan Gill: A Marionette S. cerevisiae Strain to Control Metabolic Pathways
PRESENTER: Marcelo Bassalo

Ryan Gill: Design and engineering of native regulatory networks in non-model microbes
PRESENTER: Margaret Habib

Ryan Gill: Engineering of Regulatory Networks for Improved C3-C4 Alcohol Tolerance and Production in E. coli and S. cerevisiae
PRESENTER: Emily Freed

Ryan Gill: Towards Integration of Cello, the Computer-Aided Design Platform for Genetic Circuits, into KBase.
PRESENTER: Omree Gal-Oz

H

Kirsten Hofmockel: Controls on the Composition of Microbial Derived Necromass in Soil
PRESENTER: Kirsten Hofmockel

Kirsten Hofmockel: Deconstructing the Soil Microbiome into Reduced-Complexity Functional Modules
PRESENTER: Janet Jansson

Kirsten Hofmockel: Fungal hyphal networks play a key role in soil microbiome micronutrient acquisition and transport during drought
PRESENTER: Christopher Anderton

Kirsten Hofmockel: Generation and Analysis of Reduced Complexity Model Soil Consortia
PRESENTER: Ryan McClure

Kirsten Hofmockel: Unraveling the Molecular Mechanisms Underlying the Microbiome Response to Soil Rewetting
PRESENTER: Mary Lipton

Kirsten Hofmockel: Viral diversity: decoding hidden potential for metabolic functions in soils
PRESENTER: Emily Graham

Bruce Hungate: Effects of Warming on Bacterial Growth and Element Fluxes in Soil
PRESENTER: Bruce Hungate

Bruce Hungate: Measurement of Isotope Assimilation Rates into Microbial DNA Through Quantitative Stable Isotope Probing with Internal Standards.
PRESENTER: Egbert Schwartz

I

No entries

J

Daniel Jacobson: Exascale Networks for Arabidopsis in Kbase
PRESENTER: Daniel Jacobson

Daniel Jacobson: Peta- and Exa-scale for Arabidopsis in KBase
PRESENTER: Michael Garvin

Tiffany Jamann: Conserved Genetic Mechanisms for Biotic Stress in Sorghum
PRESENTER: Tiffany Jamann

Kolby Jardine: Cell Wall O-Acetyl and Methyl Esterification Patterns of Leaves Reflected in Atmospheric Emission Signatures of Acetic Acid and Methanol
PRESENTER: Rebecca Dewhirst

Michael Jewett: Accelerating Pathway Engineering of Non-Model Organisms Through Novel Cell-Free to In Vivo Workflows
PRESENTER: Michael Koepke

Michael Jewett: Determining Protospacer Adjacent Motif Preferences of Industrially Relevant Clostridial Type I-B CRISPR-Cas Systems
PRESENTER: Grant Rybnicky

Michael Jewett: Establishing an Automated High-throughput Screening Platform
PRESENTER: Rasmus Jensen

Michael Jewett: Integrating Proteomic and Metabolomic Analyses to Optimize Cellular Extract Preparation for Enhanced Cell-Free Protein Synthesis
PRESENTER: Richard J. Giannone

Michael Jewett: Kinetic Modeling Tools Using Cell-Free Experiments to Predict Metabolic Network Behavior in Non-Model Systems
PRESENTER: Jacob Martin

Michael Jewett: Predicting Novel Biosynthetic Pathways with Generalized Enzymatic Reaction Rules
PRESENTER: Zhuofu Ni

Michael Jewett: Sequencing and Gene Mining the Largest Collection of Industrially used Acetone-Butanol-Ethanol (ABE) Fermentation Strains
PRESENTER: Steve Brown

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

Thomas Juenger: Ecosystem responses in switchgrass monoculture stands across a latitudinal gradient
PRESENTER: Michael Ricketts

Thomas Juenger: Genetics of Climate Adaptation Using Genome-Wide Association in Switchgrass
PRESENTER: Alice MacQueen

Thomas Juenger: Host Genetics Control the Composition of Root-associated Microbiota in Switchgrass (Panicum virgatum)
PRESENTER: Joseph Edwards

Thomas Juenger: Spatiotemporal dynamics of a microbiome on Panicum hallii under drought stress
PRESENTER: Esther Singer

Thomas Juenger: The genomic basis of ecotype evolution in Switchgrass
PRESENTER: John Lovell

Thomas Juenger: Using Machine Learning to Identify Cultivar x Site Interaction and Environmental Variable Affecting Aboveground Biomass
PRESENTER: Li Zhang

K

Jay Keasling: A Droplet Microfluidic Platform for Lab Automation
PRESENTER: Kosuke Iwai

Jay Keasling: Adaptive Laboratory Evolution as an Efficient Technology for Strain Construction
PRESENTER: Adam Feist

Jay Keasling: An Automated Sample Preparation Workflow For High-throughput, Quantitative Proteomic Studies of Microbes
PRESENTER: Christopher Petzold

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

Jay Keasling: Carbon footprint and economics of integrating biogas upgrading process and carbon capture technologies in cellulosic biorefineries
PRESENTER: Minliang Yang

Jay Keasling: Catabolism of Lignin Oligomers by Soil-Derived Microbiomes
PRESENTER: Steven Singer

Jay Keasling: Collaboration with the Experiment Data Depot
PRESENTER: Nathan Hillson

Jay Keasling: Conversion of Ionic Liquid Pretreated Poplar into Jet Fuel
PRESENTER: John Gladden

Jay Keasling: Distillable Ionic Liquids/Deep Eutectic Solvents for an Effective Recycling and Recovery Approach
PRESENTER: Ezinne Achinivu

Jay Keasling: Engineered Polyketide Synthases as Platform for Synthetic Chemistry
PRESENTER: Jay Keasling

Jay Keasling: Environmental Impacts of Biomass Sorghum Production in the Continental United States
PRESENTER: Umakant Mishra

Jay Keasling: Field Testing of Engineered Switchgrass with Improved Biomass Yield and Sustainability Traits
PRESENTER: Henrik Scheller

Jay Keasling: Genome-scale metabolic rewiring to achieve predictable titers, rates and yields of non-native products at scale
PRESENTER: Aindrila Mukhopadhyay

Jay Keasling: High-Throughput Screening of Lignocellulosic Biomass Degrading Enzymes Utilizing Mass Spectrometry
PRESENTER: Noel Ha

Jay Keasling: Lowering Lignin Recalcitrance and Producing Value Bioproducts in Poplar
PRESENTER: Chang-Jun Liu

Jay Keasling: Machine Learning to Predict Biomass Sorghum Yields under Future Climate Scenarios
PRESENTER: Corinne Scown

Jay Keasling: Production of Platform Chemicals in Bioenergy Crops: Stacking Low-Recalcitrance Traits with Co-Products
PRESENTER: Aymerick Eudes

Jay Keasling: Redirecting metabolic flux via combinatorial multiplex CRISPRi-mediated repression for isopentenol production in E. coli
PRESENTER: Taek Soon Lee

Jay Keasling: Sorghum Secondary Cell Wall Nanoarchitecture Can Be Revealed By Solid State NMR
PRESENTER: Jenny Mortimer

Jay Keasling: Towards Whole Biomass Utilization: Development of Ionic Liquid Technologies for Lignin
PRESENTER: Seema Singh

Matias Kirst: Deciphering N-fixing symbiosis signaling in Medicago with dynamic regulatory module networks (DRMNs)
PRESENTER: Sara Knaack

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

Matias Kirst: Global scale phylogenomics of the nitrogen fixing clade
PRESENTER: Heather Kates

L

Peggy Lemaux: Lessons from the Field: How Sorghum and Its Microbiome Respond to Drought
PRESENTER: Peggy Lemaux

Karen Lloyd: Arctic Microbial Permafrost Degradation
PRESENTER: Karen Lloyd

Stephen Long: Increasing photosynthetic efficiency of energycane under fluctuating lights
PRESENTER: Moonsub Lee

Stephen Long: Stem Parenchyma Cell-specific Gene Characterization in Energycane
PRESENTER: Jiang Wang

Stephen Long: Toward transgenic sustainable productivity increases in Miscanthus giganteus
PRESENTER: Karolina Sobanska

Stephen Long: Towards Oil Cane: Engineering Energycane for Hyperaccumulation of Lipids and Improved Agronomic Performance
PRESENTER: Guangbin Luo

David Lowry: Identification of Adaptive Fungal Pathogen Resistance Loci in Switchgrass
PRESENTER: David Lowry

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

Ting Lu: Dissecting the Social Interactions of Yeast-Lactic Acid Bacteria Consortia
PRESENTER: Yongping Xin

M

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

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

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: Andreas Vasdekis

Josh Michener: Systems Metabolic Engineering of Novosphingobium aromaticivorans for Lignin Valorization
PRESENTER: Joshua Michener

Julie Mitchell: Structure to Function: Bringing Protein Structure and Ligand Screening to KBase
PRESENTER: Ada Sedova

James Moran: Evaluating Biogeochemical Processes Facilitated by Plant and Microbial Interactions within the Rhizosphere
PRESENTER: James Moran

Paula Mouser: Metabolic and membrane adaptations of the hydraulically fractured shale isolate Halanaerobium in response to temperature and growth rate fluctuations under continuous culture
PRESENTER: Paula Mouser

Wellington Muchero: Genome sequencing reveal structural and nucleotide-level divergence among immunosuppressing G-type Lectin Receptor kinases across multiple Salix species.
PRESENTER: Wellington Muchero

N

Krishna Niyogi: Elucidating Nutrient Dependent Effects on Regulation of Photosynthesis and Metabolism
PRESENTER: Tim Jeffers

Krishna Niyogi: Genome-based Protein Function Discovery in the Eukaryotic Alga Chromochloris zofingiensis
PRESENTER: Crysten Blaby

Krishna Niyogi: Genome-scale Metabolic Model of Chromochloris zofingiensis, an Emerging Model Green Alga for Sustainable Fuel Production
PRESENTER: Alexander Metcalf

Philippe Noirot: Machine Learning Guided Design of Safeguards That Operate Under Various Bacterial Physiologies
PRESENTER: Philippe Noirot

Trent Northen: Designing Synthetic communities for dissecting plant-microbe interactions in fabricated ecosystems (EcoFABs)
PRESENTER: Kateryna Zhalnina

Trent Northen: Fabricated Ecosystems (EcoFABs) design for controlled and reproducible habitats to investigate plant-microbe-soil interactions
PRESENTER: Dawn Chiniquy

Trent Northen: m-CAFEs Applications of Targeted Editing in Microbial Networks
PRESENTER: Matthew Nethery

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

O

Victoria Orphan: Methane and nutrient cycling by sediment-hosted archaeal-bacterial syntrophic consortia and their viral predators
PRESENTER: Victoria Orphan

Victoria Orphan: Microbial Interactions at Micro-scale and Pore-scale Revealed by Process-based Reactive Transport Modeling
PRESENTER: Xiaojia He

P

Himadri Pakrasi: Systems analysis of a fast growing N2-fixing cyanobacterium for production of advanced biofuels and nitrogen-containing petrochemical replacement compounds
PRESENTER: Anindita Banerjee

Terry Papoutsakis: Modeling growth kinetics and metabolism of Clostridium acetobutylicum/Clostridium ljungdahlii co-culture with cell fusion
PRESENTER: Charles Foster

Terry Papoutsakis: Syntrophic Co-Cultures of Clostridium Organisms to Produce Higher Alcohols & Other C6-C8 Metabolites
PRESENTER: Kamil Charubin

Kabir Peay: Symbiotic niche mapping reveals nutrient specialization and functional complementarity among ectomycorrhizal fungi
PRESENTER: Kabir Peay

Jennifer Pett-Ridge: A Workflow for Generating and Polishing Nanopore Reads from Low Biomass Samples
PRESENTER: Olivier Zablocki

Jennifer Pett-Ridge: How Drought Modulates Formation and Persistence of Microbe-Derived Soil Carbon from Rhizosphere, Detritusphere, and Bulk Soil Microbial Communities
PRESENTER: Noah Sokol

Jennifer Pett-Ridge: Trait-based Modeling of Mineral-associated Soil Organic Matter Formation in Distinct Soil Habitats
PRESENTER: Gianna Marschmann

Jennifer Pett-Ridge: Unearthing the Active Microbes, Viruses and Metabolites in Dynamic-Redox Tropical Soils with Quantitative SIP and Metagenomics
PRESENTER: Jennifer Pett-Ridge

Jennifer Pett-Ridge: Using Quantitative Stable Isotope Probing to Link Precipitation Regimes of Mediterranean-Grassland Ecosystems to Soil Microbial Ecophysiology
PRESENTER: Megan Foley

Jennifer Pett-Ridge: Viral Diversity and Potential Carbon Cycling Impacts Across a Soil Climate Gradient
PRESENTER: Christine Sun

Q

No entries

R

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

S

Howard Salis: A Thousand Highly Non-Repetitive Promoters for Controlling Transcription Rates in Clostridia during Syngas Fermentation
PRESENTER: Howard Salis

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

Davinia Salvachua Rodriguez: White-Rot Fungi Utilize Lignin-Derived Compounds as a Carbon Source
PRESENTER: Davinia Salvachua

Daniel Schachtman: Approaches to the Development of Sustainable Energy Sorghum Biofuel Feedstocks in Drought Prone and Low Nitrogen Environments
PRESENTER: Daniel Schachtman

Daniel Schachtman: Data access, mining and visualization. Tools to accommodate an interdisciplinary project.
PRESENTER: Phlip Ozersky

Daniel Schachtman: Leaf Carbon Isotope Composition in Diverse Sorghum Lines
PRESENTER: Asaph Cousins

Daniel Schachtman: Sorghum root microbiome dynamics under nutrient-limited and drought conditions
PRESENTER: Susannah Tringe

Jonathan Schilling: Gene Regulatory Networks Enabling Fungi to Selectively Extract Sugars from Lignocellulose
PRESENTER: Claire Anderson

Danny Schnell: A Systems Approach to Enhancing Seedling Establishment for Increased Yields in the Oilseed Crop, Camelina sativa
PRESENTER: Danny Schnell

John Sedbrook: Advancing Field Pennycress as a New Oilseed Biofuels Feedstock that does not Require New Land Commitments
PRESENTER: John Sedbrook

Jeremy Semray: Competition Between Methanotrophs for Copper
PRESENTER: Christina Kang-Yun

Kevin Solomon: Genetic tools to optimize lignocellulose conversion in anaerobic fungi and interrogate their genomes
PRESENTER: Ethan Hillman

Rhona Stuart: Beneficial Partners: Mycorrhizal Resource Exchange in Bioenergy Cropping Systems
PRESENTER: Rachel Hestrin

Rhona Stuart: Categorizing metabolic exchange and signaling reveal distinct mechanisms of mutualistic algal-bacterial interactions
PRESENTER: Xavier Mayali

Rhona Stuart: Characterizing algal metabolites and their role in biotic interactions
PRESENTER: Vanessa Brisson

Rhona Stuart: Examining the role of physical proximity and diffusion of metabolites in algal-bacterial interactions
PRESENTER: Hyungseok Kim

Rhona Stuart: System-level analyses of beneficial interactions in an algal-bacterial co-culture
PRESENTER: Ali Navid

Rhona Stuart: Tools for Importing, Comparing and Merging Functional Annotations for Improved Metabolic Modeling in KBase
PRESENTER: Patrik D'haeseleer

Rhona Stuart: Understanding Variation in the Switchgrass Microbiome Across Scales: Evidence for Both Host Filtering and Environmental Control
PRESENTER: Marissa Lee

Matthew Sullivan: KBase Science GSP: Towards a Viral Ecogenomics Toolkit at KBase
PRESENTER: Matthew Sullivan

Matthew Sullivan: The role of viruses in the carbon cycle along a permafrost thaw gradient
PRESENTER: Matthew Sullivan

Matthew Sullivan: Viruses may manipulate the global carbon cycle through carbohydrate active enzymes
PRESENTER: Lindsey Solden

T

Bob Tabita: Novel Microbial Routes to Synthesize Industrially Significant Precursor Compounds
PRESENTER: William Cannon

Bob Tabita: Novel nitrogenase-like C-S lyases link bacterial anaerobic methionine salvage to ethylene and methane production
PRESENTER: Justin North

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

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

Gail Taylor: Understanding the genetic basis of drought tolerance of bioenergy poplar
PRESENTER: Gail Taylor

Cong Trinh: Understanding and Eliminating the Detrimental Effect of Thiamine Deficiency on the Oleaginous Yeast Yarrowia lipolytica
PRESENTER: Caleb Walker

Cong Trinh: Understanding and Harnessing the Exceptional Robustness of Yarrowia lipolytica for the Conversion of Biomass Hydrolysate into Designer Bioesters
PRESENTER: Cong Trinh

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

Gerald Tuskan: Building a suite of CRISPR/Cas9 tools for efficient switchgrass gene editing
PRESENTER: Eudald Illa-Berenger

Gerald Tuskan: Catalytic Upgrading of n-Butanol to Fully Synthetic Jet Fuel
PRESENTER: Zhenglong Li

Gerald Tuskan: Characterization of drought tolerance and water-use efficiency related traits in switchgras
PRESENTER: Yongqin Wang

Gerald Tuskan: Economic Impact of Yield and Composition Variation in Bioenergy Crops: Populus trichocarpa
PRESENTER: Erin Webb

Gerald Tuskan: Engineering CRISPR-Cas Systems for Genome Editing in Pseudomonas putida KT2440 and Clostridium thermocellum
PRESENTER: Carrie Eckert

Gerald Tuskan: High-throughput functional characterization of Populus trichocarpa UDP-glycosyltransferases
PRESENTER: Stephanie Galanie

Gerald Tuskan: AkiraProt: An Ensemble Workflow for Proteome-Wide Structural Analysis
PRESENTER: Erica Prates

Gerald Tuskan: SNPeffect: Identifying Functional Roles for SNPs using Metabolic Networks
PRESENTER: Debolina Sarkar

Gerald Tuskan: QTL mapping and candidate gene discovery of metabolomic signatures in switchgrass
PRESENTER: Tom Pendergast

Gerald Tuskan: High-throughput Reductive Catalytic Fractionation for Lignin Characterization in the Genome Wide Association Study of Poplar
PRESENTER: Michael Stone

Gerald Tuskan: Functional genomic and cross-species studies uncover hidden elements of phenylpropanoid biosynthesis
PRESENTER: Meng Xie

Gerald Tuskan: Genetic engineering to produce C-lignin deposition in plant stems
PRESENTER: Jaime Barros

Gerald Tuskan: Genome shuffling and bacterial quantitative trail locus (QTL) mapping in Pseudomonas putida
PRESENTER: Julie Chaves

Gerald Tuskan: Improving Clostridium thermocellum Product Titers by Increasing the Thermodynamic Driving Force of its Glycolytic Pathway
PRESENTER: Lee Lynd

Gerald Tuskan: Enzymatic Synthesis of Xylan Microstructures
PRESENTER: Peter Smith

Gerald Tuskan: Investigating the spatial organization of aromatic catabolism in P. putida KT2440
PRESENTER: Allison Werner

Gerald Tuskan: Leveraging Super High Optical Resolution Microscopy to Probe the Interaction Zone Between Clostridium thermocellum and Biomass
PRESENTER: John Yarbrough

Gerald Tuskan: Lignocellulose-Fermenting Microbiomes: A Compass for Biofuel Process Development
PRESENTER: Evert Holwerda

Gerald Tuskan: Rapidly domestication poplar using genomic selection and machine learning
PRESENTER: David Kainer

Gerald Tuskan: The role of beneficial microbes in stress management of Populus
PRESENTER: Jessy Labbe

Keith Tyo: Prospecting thiamine diphosphate-dependent carboligases and characterizing their promiscuity to create novel metabolic pathways from primary metabolites
PRESENTER: Bradley Biggs

Keith Tyo: Using machine learning to model promiscuous activity of thiamine diphosphate-dependent carboligases and side reactions in the E. coli metabolome
PRESENTER: Tracey Dinh

U

James Umen: Deep Green: Structural and Functional Genomic Characterization of Conserved Unannotated Green Lineage Proteins
PRESENTER: James Umen

V

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

W

Jim Wang: Dynamic Genome-Scale Metabolic Network Modeling for a Novel Methanotroph-Cyanobacteria Coculture
PRESENTER: Kiumars Badr

Jim Wang: Tuning C1-metabolism for efficient utilization of biogas in synthetic photoautotroph-methanotroph binary consortium.
PRESENTER: Marcus Bray

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

Thea Whitman: Dissection of Carbon and Nitrogen Cycling in Post-Fire Soil Environments using a Genome-Informed Experimental Community
PRESENTER: Thea Whitman

Mari Winkler: Integrating single-cell wetland microbiome structure, function, and activity to ecosystem-scale biogeochemical fluxes
PRESENTER: Mari-Karoliina Winkler

Kelly Wrighton : Coupling KBASE with PFLOTRAN
PRESENTER: Roelof Versteeg

X

No entries

Y

Todd Yeates: Innovations in Enzyme and Pathway Engineering for Cell-Free Production of Biofuels and High-Value Chemicals
PRESENTER: Saken Sherkhanov

Todd Yeates: Microbial Metabolism, Chemistry, and Communities under Study at the UCLA-DOE Institute for Genomics and Proteomics
PRESENTER: Rachel Loo

Todd Yeates: New Atomic Imaging Technology Development at the UCLA-DOE Institute
PRESENTER: Jose Rodriguez

Todd Yeates: Transcriptomic analyses of bulk and single cell Chalmydomonas RNA-seq data reveal new gene functions and cell state heterogeneity
PRESENTER: Matteo Pellegrini

Jamey Young: Metabolic engineering of cyanobacteria for enhanced production of ethylene and free fatty acids
PRESENTER: Bo Wang

Jamey Young: Rapid flux phenotyping to accelerate metabolic engineering of cyanobacteria
PRESENTER: Piyoosh Babele

Z

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

Huimin Zhao: Evidence for Metabolic Channeling of Glucose into the Oxidative Pentose Phosphate Pathway to Drive NADPH Production in Rhodosporidium toruloides
PRESENTER: Tianxia Xiao

Huimin Zhao: Exploring Oleaginous Yeast Rhodosporidium toruloides as a Platform Organism for Production of Chemicals and Fuels
PRESENTER: Carl Schultz

Huimin Zhao: Genome-scale Model Reconstruction and 13C-Metabolic Flux Analysis for Non-model Yeast Organisms Rhodosporidium toruloides IFO0880 and Issatchenkia orientalis SD108
PRESENTER: Hoang Dinh

Improving the safety and outcome of research using next-generation genome engineering

Paul E. Abraham[1*]*(abrahampe@ornl.gov) Michael M. Vergara[1], Guoliang Yuan[1], Mahmudul Hassan[1], Amber McBride[1], Xiaohan Yang[1], Daniel Jacobson[1], and Jessy Labbe[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831.

Project Goals: The goals of this pilot program as an activity in the Secure Biosystems Design initiative is to contribute awareness and inform decision-making across biosystems design research. This is a part of BERs Genomic Science program, integrating ongoing efforts in microbiome, environmental genomics, and sustainability research in mission-relevant ecosystems.

The simplicity and flexibility of CRISPR/Cas offers unprecedented opportunities to rewrite genomes. Unfortunately, researchers developing new techniques for advanced genome editing rarely have the resources to properly assess all the risks they are introducing. Analysis is typically performed using a single, well-characterized genotype of interest under laboratory conditions. As such, current research lacks sufficient knowledge to detect, assess, and mitigate unintended consequences of these techniques. Until these knowledge gaps are addressed, scientists emphasize that precautions are necessary because this biotechnology is moving faster than regulation considerations and actions. Given the risk CRISPR/Cas-enabled gene drive systems pose (e.g., gene drives systems may escape confinement through accidents) and potential for far-reaching, even global spread from small releases, safeguarding genomes with a countermeasure against unwanted gene editing is a high priority. To mitigate these risks, this project leverages a nucleic acid-based approach to have an invading CRISPR/Cas system self-identify and self-destruct. Initial work has focused on implementing and evaluating this approach in microbial and plant systems. To this end, preliminary breakthroughs demonstrate that locking mechanisms can be incorporated into genomes to provide a useful containment or countermeasure measure to regulate or avert CRISPR/Cas gene-editing.

Funding statement. This work was supported by the U.S. Department of Energy (DOE), Office of Biological and the Environmental Research (BER) Secure Biosystems Design program.

Genome-scale metabolic and regulatory network reconstruction of *Caldicellulosiruptor bescii*

Ying Zhang[1]*(yingzhang@uri.edu), Ke Zhang, [1] Irina A. Rodionova,[2] Dmitry A. Rodionov,[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]University of Rhode Island, Kingston, RI; [2]Sanford-Burnham-Prebys Med. Discovery Institute, La Jolla, CA; [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 non-model, microbial metabolic engineering platform based on the most thermophilic lignocellulose-degrading organism known, *Caldicellulosiruptor bescii,* which grows optimally 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 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.

Caldicellulosiruptor bescii is an extremely thermophilic, strictly anaerobic, gram-positive bacterium. It is the most thermophilic cellulolytic bacterium known to date (T_{opt}=78~80°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. Engineered *C. bescii* strains has 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.

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 complete genomes of thirteen species of *Caldicellulosiruptor* were used for ortholog mapping and comparative analysis. Functional gene assignments, genome context analysis, comparative analysis of orthologous genes and DNA upstream regions, gene co-occurrence analysis and protein similarity searches were performed in the SEED environment (1). We also used genome annotations from Swiss-Prot, KEGG, TCDB, and RegPrecise databases and published experimental data. The previously generated RNASeq datasets for whole-genome gene expression and transcriptional start sites obtained for *C. bescii* grown on five different carbon sources (glucose, xylose, cellobiose, xylan, cellulose) were used for validation of reconstructed transcriptional regulons and for refinement of transporter specificities. The curation of a genome-scale model (GEM) was done with the support of PSAMM software (2) to incorporate the

known and predicted metabolic functions of enzymes and transporters.

The global reconstruction of carbohydrate utilization includes almost 200 *C. bescii* genes, encoding enzymes, transporters and transcription factors (TFs), involved in more than 20 distinct pathways for the utilization of various carbohydrates. Using comparative genomics, we identified *ab initio* novel DNA-binding motifs and reconstructed regulatory networks for 24 TFs controlling individual sugar catabolic pathways. The global reconstruction of carbohydrate utilization is being integrated with a *C. bescii* GEM, which contains 721 metabolites and 772 metabolic reactions associated with 520 metabolic genes, covering 17% or the entire genome and 73% of the metabolic and carbohydrate transport genes in *C. bescii*. Besides the carbohydrate utilization pathways, the *C. bescii* GEM contains a diverse range of catabolic and anabolic pathways, including central carbon metabolism and the biosynthesis of proteins, nucleotides, lipids, vitamins, and cofactors. Biomass production of *C. bescii* is represented in the GEM using a biomass objective function that is carefully calibrated using experimental measurements of major cell components. Growth predictions made by the *C. bescii* GEM is validated through matching metabolic simulations to with growth measurements in batch and chemostat culture using defined media. This model serves as a stepping stone for the engineering of *C. bescii* strains to enable and enhance the yields of bio-based fuels and chemicals.

IMPORTANCE In this study, we built a predictive model for simulating the metabolism of the non-model organism, *C. bescii*. The simulation predictions made by simulations can provide potential directions for the more efficient metabolic engineering of *C. bescii*.

References

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

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

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

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

Gabriel M. Rubinstein[1*] (grubinst@uga.edu), James Crosby,[2] Ryan G. Bing,[2] Tania Tanwee,[1] Ke Zhang,[4] Ying Zhang,[4] Dmitry Rodionov,[3] Robert M. Kelly,[2] and **Michael W. W. Adams[1]**

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

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.

Bio-processing above 70°C can have important advantages over near-ambient operations. Highly genetically modified microorganisms typically have a fitness disadvantage and can be easily overtaken in culture by contaminating microbes. The high growth temperature of extreme thermophiles precludes growth or survival of virtually any contaminating organism or phage. This reduces operating costs associated with reactor sterilization and maintaining a sterile facility. In addition, at industrial scales, heat production from microbial metabolic activity vastly outweighs heat loss through bioreactor walls such that cooling can be required. Extreme thermophiles have the advantage that non-refrigerated cooling water can be used, and heating requirements can be met with low-grade steam, typically in excess capacity on plant sites. This project is leveraging recent developments in the PIs' labs for *C. bescii* that enable the proposed effort (1-11). We are developing approaches that provide a comprehensive description of this bacterium's physiology and metabolism to inform metabolic engineering strategies, validate the models with experimental data, and demonstrate that unpretreated lignocellulose can be converted into value-added industrial chemicals at bioreactor scale. The specific aims of this research are: 1) to construct and test a robust metabolic model based on a metabolic reconstruction of *C. bescii* growing on the simple sugars, glucose and xylose, 2) to construct and test a robust metabolic model of *C. bescii* growing on complex biomass-related sugars, cellulose and xylan, 3) to optimize the production of acetone and other industrial chemicals from simple sugars guided by metabolic modeling, and 4) to demonstrate conversion of cellulose, xylan, cellulose/xylan, and the model biomass switchgrass to valuable fermentation products.

At present, high temperature chemostat cultures are being used in conjunction with transcriptomic analysis to determine bioenergetic parameters and gene regulation patterns for *C. bescii* growth on lignocellulose-relevant sugars, including glucose, xylose, cellulose, and xylan. Notably, an alternative glycolytic pathway was recently identified in *C. bescii*, and the role of this pathway in regulating redox pools and electron flux is currently under investigation (12). Additionally, a strain of *C. bescii* has now been engineered to produce acetone during growth on cellobiose. Furthermore, vectors for engineering strains to produce other industrial chemicals have been constructed. These efforts are informing comprehensive metabolic reconstruction and modeling analyses with the ultimate goal of optimizing the production of useful products from renewable feedstocks by recombinant strains of *C. bescii*.

References

1. Lee, L. L., et al. (2018) Comparative biochemical and structural analysis of novel cellulose binding proteins (Tāpirins) from extremely thermophilic *Caldicellulosiruptor* species. *Appl. Environ. Microbiol.* (doi: 10.1128/AEM.01983-18)
2. Williams-Rhaesa, A. M., et al. (2018) Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii. Metab. Engineering Commun.* **7,** e00073 (doi: 10.1016/j.mec.2018.e00073)
3. Zeldes, B.M. et al. (2018) A synthetic enzymatic pathway for extremely thermophilic acetone production based on the unexpectedly thermostable acetoacetate decarboxylase from *Clostridium acetobutylicum*. *Biotechnol. Bioeng.* 115:2951-2961 (doi: 10.1002/bit.26829)
4. Williams-Rhaesa, A.M. et al. (2018) A xylose inducible promoter expands the genetic tools for the biomass-degrading, extremely thermophilic bacterium *Caldicellulosiruptor besci*. *Extremophiles* (doi: 10.1007/s00792-018-1023)
5. Conway, J.M. et al. (2018) Novel multi-domain, multi-functional glycoside hydrolases from highly lignocellulolytic *Caldicellulosiruptor* species. *AIChE J.* 64:4218-4228 (doi.org/10.1002/aic.16354)
6. Conway, J.M. et al. (2018) Parsing *in vivo* and *in vitro* contributions to microcrystalline cellulose hydrolysis by multi-domain glycoside hydrolases in the *Caldicellulosiruptor bescii* secretome. *Biotechnol. Bioeng.* 115:1-15.
7. Williams-Rhaesa, A., et al. (2017) Genome stability in engineered strains of the extremely thermophilic, lignocellulose-degrading bacterium *Caldicellulosiruptor bescii Appl Environ Microbiol* 83 (doi: 10.1128/AEM.00444-17)
8. Zurawski, J.V., et al. (2017) Bioavailability of carbohydrate content in natural and transgenic switchgrasses for the extreme thermophile *Caldicellulosiruptor bescii*. *Appl. Environ. Microbiol*. (doi: 10.1128/AEM.00969-17)
9. Conway, J.M., et al. (2017) Functional analysis of the Glucan Degradation Locus (GDL) in *Caldicellulosiruptor bescii* reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. *Appl. Environ. Microbiol.* 83:e01828-17 (doi: 10.1128/AEM.01828-17)
10. Lee, L.L., et al. (2018) Genus-wide assessment of lignocellulose utilization in the extremely thermophilic *Caldicellulosiruptor* by genomic, pan-genomic and metagenomic analysis. *Appl. Environ. Microbiol.* e02694-17 (doi: 10.1128/AEM.02694-17)
11. Straub, C.T. et al. (2019) Quantitative fermentation of unpretreated transgenic poplar by *Caldicellulosiruptor bescii*. *Nat. Commun.* 10, 3548 (doi:10.1038/s41467-019-11376-6)
12. Scott, I. et al. (2019) The thermophilic biomass-degrading bacterium *Caldicellulosiruptor bescii* utilizes two enzymes to oxidize glyceraldehyde-3-phosphate during glycolysis. *J Biol. Chem.* 294: 9995-10005. doi:10.1074/jbc.RA118.007120

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

A Method for Circularizing Microbial Genomes from Metagenomics Data

L.M. Lui (lmlui@lbl.gov)[1]*, T. Nielsen[1], H.J. Smith[2], F. von Netzer[3], E.L-W. Majumder[4], J.V. Kuehl[1], F. Song[1], A. Sczesnak[1,5], M.P. Thorgesen[6], X. Ge[6], F.L. Poole[6], C.J. Paradis[7], K.F. Walker[8], K.A. Lowe[9], D.C. Joyner[9], D. Ning[9], M. Rodriquez, Jr.[8], A.B. Aaring[1], B.A. Adams[8], D. Williams[8], J.D. Van Nostrand[10], G.M. Zane[11], M.W.W. Adams[6], J. Zhou[10], R. Chakraborty[1], J.D. Wall[11], D.A. Stahl[3], T.C. Hazen[8,9], M.W. Fields[2], AP Arkin[1,5], **PD Adams[1]**

[1]Lawrence Berkeley National Lab, Berkeley CA; [2]Montana State University, Bozeman MT; [3]University of Washington, Seattle WA; [4]State University of New York, Environmental Science and Forestry; [5]University of California, Berkeley CA; [6]University of Georgia, Athens GA; [7]University of Wisconsin, Milwaukee WI; [8]University of Tennessee, Knoxville TN; [9]Oak Ridge National Lab, Oak Ridge TN; [10]University of Oklahoma, Norman OK; [11]University of Missouri, Columbia MO

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.

Metagenomics facilitates the study of the genetic information from unculturable microbes and complex microbial communities, but achieving complete microbial genomes (*i.e.*, circular) from metagenomics data is difficult because of inherent qualities of short read sequencing data and assembly and binning methods currently available. To our knowledge, only 62 circularized genomes have been assembled from metagenomics data despite the thousands of datasets that are available. We believe that circularized genomes are important for (1) building a reference collection as scaffolds for future assemblies, (2) providing complete gene content of a genome, (3) confirming little or no contamination of a genome, (4) studying the genomic context of genes, and (5) linking protein coding genes to ribosomal RNA genes to aid metabolic inference in 16S rRNA gene sequencing studies. We developed a method to achieve circularized genomes using iterative assembly, binning, and read mapping. In addition, this method exposes potential misassemblies from k-mer based assemblies. We chose species of the Candidate Phyla Radiation (CPR) to focus our initial efforts because they have small genomes and are only known to have one copy of the ribosomal RNA genes. From our work, we present 42 CPR genomes and one Margulisbacteria genome from 19 published datasets and from ENIGMA sequencing of sediment and groundwater samples from Oak Ridge National Lab Field Research Center. We demonstrate findings that would likely be difficult without circularized genomes, including that ribosomal genes are likely not operonic in the majority of CPR, diverged forms of RNase P in CPR, and presence of megaplasmids in the datasets.

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

A Multi-Laboratory Effort to Use Synthetic Communities to Discover, Characterize, and Dissect Key Microbial Processes Relevant to Field Observations

J.J. Valenzuela[1]*(jvalenzu@isbscience.org), A.V. Carr[1], J.A. Wilson[1], K. Hunt[2], H.J. Smith[3], F.L. Poole[4], X. Ge[4], C.M. Gionfriddo[5], R.L. Wilpiszeski[5], V. Li[6], A. M. Deutschbauer[6], T. R. Northen[6], M. W.W. Adams[4], R. Chakraborty[6], D. A. Elias[5], D. A. Stahl[2], M. W. Fields[3], G. E. Siuzdak[7], **N. S. Baliga[1,2], A. P. Arkin[6,8]** and **P. D. Adams[6,8]**

[1]Institute for Systems Biology, Seattle, WA; [2]University of Washington, Seattle, WA; [3]Montana State University, Bozeman, MT; [4]University of Georgia, Athens, GA; [5]Oak Ridge National Lab, Oak Ridge, TN; [6]Lawrence Berkeley National Lab, Berkeley CA; [7]Scripps Research Institute, La Jolla, CA; [8]University of California at Berkeley, CA

http://enigma.lbl.gov

Project Goals: ENIGMA -Ecosystems and Networks Integrated with Genes and Molecular Assemblies use a systems biology approach to understand the interaction between microbial communities and the ecosystems they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods. Thus, ENIGMA has been organized into several campaigns involving multiple institutes with varying expertise. *Here we describe the core efforts of the Environmental Simulation and Modeling campaign (EnvSim) to establish synthetic communities to discover, characterize, and dissect key microbial processes relevant to field observations, in particular as they relate to denitrification and the emission of the greenhouse gas nitrous oxide (N_2O).*

The Environmental Simulation and Modeling campaign (EnvSim) has strategically aligned efforts across ENIGMA to address two major sets of hypotheses relevant to ecologically important phenomena observed at the Oak Ridge Field Research Center (FRC). The first hypothesis, which originated from field observations, was that sulfate respiration and nitrate respiration processes operate in mutual exclusivity down the transect of a sediment core. Using the field isolate (*Intrasporangium calvum* C5) this phenomenon was simulated and the mechanistic underpinnings delineated through laboratory investigations. Insights from this study are currently being tested back at the FRC using groundwater chemostats to demonstrate the "field-to-lab-to-field" iteration model employed across ENIGMA for investigating a phenomenon that emerged through the investigations of complex interactions within a microbial community. The second set of hypotheses being tested by the EnvSim campaign involve the interplay of biotic and abiotic factors that lead to N_2O emissions at the FRC. It has been shown that in FRC wells which have a lower pH (~6.5 - 3) tend to have higher concentrations of N_2O. The EnvSim campaign has deduced four potential mechanisms that may account for the N_2O emissions at the FRC and are currently being investigated by labs across ENIGMA. Denitrification at the FRC[1] may be driven by complete denitrifiers, however, their NosZ enzymes, which catalyze the final step in denitrification by converting N_2O to N_2, may be sensitive at lower pHs. Additionally, the denitrification process could be partitioned among organisms and some may have pH-sensitive NosZ genes. Another factor contributing to variable N_2O emissions relates to the metal co-factors involved in the denitrification pathway[2,3]. For instance, excess Cu, Al, Mn, U, Ni, Co, Cu, and/or Cd may have inhibitory effects on multiple

enzymatic steps during denitrification, while the enzymatic production of nitrite, the precursor of N_2O, can be limited by the essential metal Mo. Lastly, abiotic production of N_2O may occur as a result of chemodenitrification, in which metals like Fe, Mn, and some organic compounds can drive redox reactions that convert nitrogen cycle intermediates to N_2O under the right conditions[4].

Here, we describe an important subset of ongoing collaborative projects that are dissecting the above hypotheses. We highlight the complementary nature of each project and how the use of different reactor systems answer distinct questions, which provides important information and data that can be funneled into our predictive models. For instance, we are using column reactors and highly controlled, constantly stirred, planktonic reactors to study how microbial community structure and function changes in response to pH shifts, oxygen concentrations, and sulfurous compounds. Batch culture experiments are being used to study the abiotic and biotic impacts of metals on denitrification processes. In addition, transcriptomic analysis of these experiments are aiding the construction of a metabolic and gene regulatory network model. Importantly, the work being showcased illustrates the field-to-lab and lab-to-field framework of the ENIGMA campaign strategy.

References

1. Alvarez, L., Bricio, C., Blesa, A., Hidalgo, A. and Berenguer, J., 2014. Transferable denitrification capability of Thermus thermophilus. Appl. Environ. Microbiol., 80(1), pp.19-28.
2. Glass, J. and Orphan, V.J., 2012. Trace metal requirements for microbial enzymes involved in the production and consumption of methane and nitrous oxide. Frontiers in microbiology, 3, p.61.
3. Tavares, P., Pereira, A.S., Moura, J.J.G. and Moura, I., 2006. Metalloenzymes of the denitrification pathway. Journal of inorganic biochemistry, 100(12), pp.2087-2100.
4. Zhu-Barker, X., Cavazos, A.R., Ostrom, N.E., Horwath, W.R. and Glass, J.B., 2015. The importance of abiotic reactions for nitrous oxide production. Biogeochemistry, 126(3), pp.251-267.

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

Applying Stable Isotopes for Source Fingerprinting of Dissolved Organic Nitrogen in Groundwater

Romy Chakraborty[1]* (rchakraborty@lbl.gov), Xiaoqin Wu,[1] Dominique C. Joyner[2], Terry C Hazen[2], Ria Gracielle Malana,[1] Adam P. Arkin[1,3] and Paul D. Adams[1,3]

[1]Lawrence Berkeley National Lab, Berkeley; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]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.

In this project, we aim to: 1) differentiate the multiple sources of dissolved organic nitrogen (DON) in FRC contaminated groundwater via isotopic and elemental analysis; 2) quantify the contributions of DON in FRC contaminated groundwater during four successive seasons using stable isotope analysis in R (SIAR) mixing model; 3) verify the impact of source on labile fractions of DON in groundwater.

Abstract

DON constitutes a major pool of dissolved N in many aquatic ecosystems, playing an important role in biogeochemical cycling of both carbon (C) and N. Our recent work on natural organic matter at FRC indicated that dissolved organic nitrogen (DON) contributes to more than 50% of dissolved N pool in uncontaminated sediments, serving as an important source of C and N for microbes when labile C is limiting. Although considerable research has been carried out on dissolved inorganic N (DIN) transformation at FRC, the role of DON is largely unknown and has been overlooked thus far. Variations in source determine the quantity and quality of DON, therefore greatly affecting the microbial turnover of DON, and resultant microbial community structure. Stable isotope signature at natural abundance is a powerful tool in source fingerprinting of bulk C and N in soil and sediment, as well as nitrate in aquatic environments. However, only a handful of studies have been reported on isotopic survey of DON in marine, lake, and soil environments, none available for groundwater.

To obtain baseline information of DON in Oak Ridge FRC groundwater, we investigated the quantity and quality of dissolve organic matter (DOM) in groundwater during an ENIGMA 2-month sampling campaign, carried out in spring 2019 (mid March to mid May) at both FRC uncontaminated background area and nitrate contaminated area. Both DON and DOC content in nitrate-contaminated wells during late campaign period (mid April to mid May) was significantly lower ($p < 0.05$) than those during early campaign period. This suggests the existence of a

temporal fluctuation of DON quantity in FRC groundwater especially at nitrate-contaminated area, and the fluctuation can likely be explained as dilution by water from melting existing snowpack on ground. DOM quality in groundwater changed during this period as well. We used the ratio of ON to OC (ON/OC) as an indicator of DOM quality. During late campaign period (mid April to mid May), the ON/OC decreased in nitrate-contaminated groundwater, indicating that water input from melting snowpack potentially introduces different types of DOM (N-poor molecules) to groundwater.

Following up on these observations we are developing a reliable analytical method to measure N isotope signature of trace level of DON in groundwater. Currently, there is no well-established method for directly measuring δ^{15}N-DON in environmental water sample. We are testing solid-phase extraction based methods to concentrate trace level of DON from groundwater and to eliminate inorganic N such as nitrate and ammonium. We are testing different sorbents, and will apply Elemental Analyzer-Isotope Ratio Mass Spectrometry (EA-IRMS) for measuring δ^{15}N-DON in the solid sorbents.

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

Core Values: Spatial Variation in Microbial Function, Activity, and Community Assembly in Groundwater and Sediment from a Contaminated Subsurface Aquifer

L.M. Lui*[1] (lmlui@lbl.gov), H.J. Smith*[2] (heidi.smith@montana.edu), F. von Netzer[3], K.B. De León[4], E.L-W. Majumder[5], J.V. Kuehl[1], F. Song[1], A. Sczesnak[1,6], T. Nielsen[1], M.P. Thorgesen[7], X. Ge[7], F.L. Poole[7], B.P. Bowen[1], S.M. Kosina[1], C.J. Paradis[8], K.F. Walker[9], K.A. Lowe[10], D.C. Joyner[9], M. Rodriquez, Jr.[10], B.A. Adams[9], D. Williams[9], J-W. Moon[11], J.D. Van Nostrand[12], D. Ning[12], Y. Fu[12], W. Shi[12], Y. Li[12], D.J. Curtis[12], Y. Fan[12], L. Wu[12], R. Tian[12], G.M. Zane[4], A.B. Aaring[1], X. Wu[1], A.E. Kazakov[1], J-M. Chandonia[1], P.S. Novichkov[1], P.J. Walian[1], R. Chakraborty[1], M.W.W. Adams[7], J. Zhou[12], T.R. Northen[1], J.D. Wall[4], D.A. Stahl[3], D.A. Elias[10], T.C. Hazen[9,10], M.W. Fields[2], A.P. Arkin[1,6], **P.D. Adams[1]**

[1]Lawrence Berkeley National Lab, Berkeley CA; [2]Montana State University, Bozeman MT; [3]University of Washington, Seattle Washington; [4]University of Missouri, Columbia MS; [5]Scripps Research Institute, San Diego CA; [6]University of California, Berkeley CA; [7]University of Georgia, Athens GA; [8]Los Alamos National Laboratory, Los Alamos NM; [9]University of Tennessee, Knoxville TN; [10]Oak Ridge National Lab, Oak Ridge TN; [11]National Minerals Information Center, US Geological Survey Reston, VA; [12]University of Oklahoma, Norman OK

http://enigma.lbl.gov

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

Primary motivations for studying the subsurface are to expand what is known about Earth's microbial diversity and the subsurface microorganisms under low nutrient conditions that significantly impact C, S, N, P and mineral cycles. However, on an ecosystem scale, there is limited information regarding the exact relationship between microbial diversity, environmental parameters, and biogeochemical processes between subsurface groundwater and porous media (*i.e.*, sediment). The Department of Energy ENIGMA Scientific Focus Area seeks to map the causal interactions that constrain microbial community assembly, functionality, and dispersal in chemically and physically complex environments. We recently initiated an in depth study of microbial distribution and activity throughout sediment and water compartments in the shallow subsurface at the Oak Ridge Field Research Center, a site of nuclear weapon development during the Manhattan Project.

We hypothesize that strong gradients of pH, heavy metals, nitrate, and other contaminants at the site influence the distribution, structure, and activity of microbial communities. We performed large-scale analysis of two sediment cores and associated groundwater for which we produced depth-index data sets of physical, chemical, bulk biological and sequencing measurements. One core (466 cm) was from a region outside the area of heavy chemical contamination and the other core (815 cm) from within the contaminated area. We divided the cores into ~23 cm segments for processing, resulting in 56 segments which allowed a finer-grained analysis of the vertical transect as compared to other subsurface studies.

We observed differences in diversity and distribution of dominant metabolisms between the heavily contaminated and less contaminated cores. The heavily contaminated core was less diverse as measured by 16S amplicon sequencing. Approximately 250 exact sequence variants (ESVs) accounted for half the observed reads in the heavily contaminated core as compared to 660 from the uncontaminated core, suggesting strong selective pressure from contamination. In general, there is little overlap in ESVs between the two cores (~300-350 meters apart). Analysis of ESV distribution and inferred functional potential using FAPROTAX suggests that the less contaminated core has strongly ordered communities with well-defined functional zones for nitrification, denitrification, methanogenesis, and sulfate reduction, and that these functions are carried out by specific diverse subcommunities differentially distributed with depth. There was a clear interplay among communities mediating denitrification, methanogenesis, and sulfate reduction within the variably saturated zone. These metabolisms are also correlated with the presence of key chemical constraints, such as uranium, nitrate, and pH, not just location in the core. Conversely, the heavily contaminated core has a far more heterogenous population structure with little evident intersegmental transfer (*i.e.*, little mixing between adjacent layers). Within the heavily contaminated core we observed two distinct cosmopolitan communities: (1) a large and diverse community enriched in denitrifying organisms and (2) a less diverse highly abundant community not clearly enriched in any metabolism. We found that the ENIGMA isolate collection is not representative of all of the organisms in the cores, but in some cases we have greater than 80% coverage of the strains present in a subcommunity. Based on the isolates we have and the strong chemical and physical correlates to specific community compositions, we are currently in the process of generating enrichments and synthetic communities representative of each of the identified subcommunities. From activity assays, we identified a competitive exclusion of sulfate reduction and denitrification; species representative of these metabolisms are found and enriched in different locations in the core. This exclusion was further supported by shotgun metagenomics. From 22 groundwater and 91 sediment shotgun metagenomes, we have successfully circularized 8 genomes and have over 50 genomes with >98% completeness and <2% contamination. We are tracking these genomes with depth in the core and analyzing the potential functional roles of these organisms in denitrification and sulfate reduction.

To distinguish intact and potentially viable cells from "relic" DNA, we used complementary culture-independent methods to determine respiratory rates and identify the active fractions of microbial assemblages from groundwater and sediment. Groundwater from the less contaminated area had higher diversity and three- to four-fold higher activity than heavily contaminated samples. Additionally, in the less contaminated area, the activity on a per cell basis was two to three-fold greater for planktonic cells compared to particle associated organisms, with small cells (<0.1μm) contributing up to 19% of total activity. Conversely, in heavily contaminated aquifers, activity was greater for sediment-associated cells. To understand the distribution of the active ESVs across the ORNL landscape we are analyzing the top active ESVs and checking their abundance in historic 16S rRNA sequence data from nearly 100 wells from ORFRC.

This study integrates over 12 measures of microbial community composition, activity, and environmental controls to provide new insights into how contamination impacts the distribution of activity between attached and planktonic populations in subsurface microbial communities. The application and refinement of *in situ* measurements spanning different scales will aid in the development predictive frameworks for understanding large scale biogeochemical cycling from groundwater environments.

ENIGMA (http://enigma.lbl.gov) at LBNL supported by Office of Biological and Environmental Research US Dept of Energy Contract No: DE-AC02-05CH11231

High nitrous oxide emissions from a nitrate contaminated subsurface indicate significant metabolic activity.

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

[1]University of Washington, Seattle, Washington; [2]Institute for Systems Biology, Seattle, Washington; [3]Oak Ridge National Lab, Oak Ridge, Tennessee; [4]Woods Hole Oceanographic Institution, Woods Hole, MA; [5]Lawrence Berkeley National Laboratory, Berkeley, California

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.

Linking field observations with laboratory studies and vice versa is essential for advancing predictive understanding of environmental systems and for stewardship of those systems. However, some environments that provide critical ecosystem services, such as the subsurface and groundwater systems, are extremely difficult to sample and monitor in real-time, and doing so is both expensive and invasive. Thus, nondestructive approaches to process analyses are essential tools for connecting lab and environment. To this end, we are developing protocols that use the flux and isotopic composition of the greenhouse gas nitrous oxide to resolve alternative biotic (e.g., nitrification and denitrification) and abiotic processes, such as iron catalyzed reduction of nitrite, that contribute to its emission from the subsurface. ENIGMA has developed a program to connect measurements of nitrous oxide to biologically mediated processes through field observations (For more information on ENIGMA field observations see "Spatiotemporal Dynamics of Groundwater and Sediment: Geochemistry, Microbial Communities and Activities in a Contaminated Aquifer" by Walker et. al.) and laboratory simulations (For more information on ENIGMA lab to field plan see "A Multi-Laboratory Effort to Use Synthetic Communities to Discover, Characterize, and Dissect Key Microbial Processes Relevant to Field Observations" by Valenzuela et. al.), thereby establishing a noninvasive metric for quantifying activity without destructive sampling.

Contamination by nitrogen species is a concern in many terrestrial and aquatic environments impacted by past and current human activities, including release associated with intensive agriculture and industrial activity, and from wastewater treatment. This contamination has been shown to lead to altered plant, animal, and microbial communities and to increased production of the greenhouse gas nitrous oxide, primarily through either nitrification or denitrification. The subsurface of the Field Research Center (FRC), near Oak Ridge National Lab in Tennessee, has been contaminated with low pH (3-7), heavy metal laden nitrate (~10 g/l) for decades. To understand how this contamination has influenced subsurface processes we are investigating environmental variables influencing nitrous oxide emissions, considering both biotic and abiotic contributions to this important greenhouse gas. By using flux analysis and isotopic

characterization of nitrous oxide, combined with complementary molecular and chemical characterization of multiple observation wells, we anticipate developing a more predictive understanding of the controlling variables. Current analyses are focused on wells positioned at different depths and spanning a range of pH, nitrate contamination, and metals contamination to resolve biotic and abiotic sources of production and to identify controlling environmental variables. Initial data sets have revealed that subsurface fluxes are orders of magnitude higher than those observed in other systems, including agricultural soils and the marine oxygen minimal zone. In contrast, the observed surface fluxes are in the range observed for other sites, indicating additional consumptive processes within the vadose zone that mitigate surface emissions.

Funding statement: ENIGMA is a Scientific Focus Area Program at Lawrence Berkeley National Laboratory and is supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

High throughput approaches for investigation of microbial interactions within synthetic microbial communities

Trent R. Northen[1*](TRNorthen@lbl.gov) Markus de Raad[1], Nicholas R. Saichek[1], Suzanne M. Kosina[1], Benjamin P. Bowen[1], Lauren M. Lui[1], Hans K. Carlson[1], Adam M. Deutschbauer[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, Berkeley CA

http://enigma.lbl.gov

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

There is an urgent need to improve our understanding of the connections between microbial community composition and interactions to their *in situ* activities. Recently, we reported the implementation of BONCAT (Biorthogonal Non-Canonical Amino Acid Tagging) to capture the translationally active cells in soils[1] from Oak Ridge, TN. When comparing microbial populations from two soil depths incubated under the same conditions for seven days, we found that active populations ranged from 25 – 75% of total cells, and accounted for 3-4 million active cells per gram of soil. The BONCAT positive cell fraction for each depth was recovered by fluorescence activated cell sorting (FACS) and identified by16S amplicon sequencing. On average, 86% of sequence reads recovered from the active community shared >97% sequence similarity with cultured isolates from the same location suggesting that we can use our existing isolate collection to construct synthetic communities (SynComs) designed to mimic active populations and investigate relevant microbial interactions, especially resource partitioning.

Currently, tracking growth rates and resource use in mixed communities is technically challenging. Significant progress has been made using sequencing-based approaches. However, there are limitations in sensitivity and throughput. Use of mass spectrometry for analysis of strain specific protein biomarkers represents an alternative and promising complementary approach. This approach is routinely used for rapid, highly sensitive identification of bacteria in clinical settings. We are adapted this protein-profiling approach for use with stable isotope probing to both determine community structure and track resource partitioning in our BONCAT positive SynCom experiments. This is accomplished by using strain-specific biomarker profiling with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Initial results show that this approach accurately quantified individual strain abundance in a mixed community of six different bacterial strains. Using exometabolomics profiling, we have predicted co-culture resource partitioning and are now feeding the SynComs with stable isotope labeled forms of these metabolites. With this approach, we will track label incorporation into newly synthesized protein biomarkers to compare metabolite use in pure cultures vs. SynComs. These integrated technologies will provide important new insights into resource competition and cross-feeding within sediment communities to help address our science goals.

References

[1] E. Couradeau, J. Sasse, D. Goudeau, N. Nath, T.C. Hazen, B.P. Bowen, R. Chakraborty, R.R. Malmstrom, T.R. Northen, Probing the active fraction of soil microbiomes using BONCAT-FACS, Nat. Commun. 10 (2019). doi:10.1038/s41467-019-10542-0.

ENIGMA is a Scientific Focus Area Program at Lawrence Berkeley National Laboratory and is 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.

Integrating data and algorithms from the ENIGMA project into KBase

D.M. Needham[1], A. Zhang[1], J-M. Chandonia[2], D. Chivian[2], L.M. Lui[2], W. Zheng[3], S. Zhao[1], Y. Yin[1], D.A. Weitz[3], T.C. Hazen[4,5], P.S. Novichkov[2], J. Zhou[6], E.J. Alm[1], A.P. Arkin[2,7], P.D. Adams[2,7]

[1]Massachusetts Institute of Technology, Cambridge MA; [2]Lawrence Berkeley National Lab, Berkeley CA; [3]Harvard University, Cambridge MA; [4]University of Tennessee, Knoxville TN; [5]Oak Ridge National Lab, Oak Ridge TN; [6]University of Oklahoma, [7]University of California at Berkeley

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.

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.

As sequencing becomes less expensive, researchers are turning from 16S rRNA surveys to metagenomics in order to add a new level of functional and phylogenetic resolution to their sequence-based analyses. Metagenomic data harbors additional layers of data on population structure, strain dynamics, and genome evolution that cannot be inferred from 16S alone. Yet, powerful and user-friendly tools for the analysis of this data are not publicly available.

Microbial communities across nearly all biological systems, from sediment to groundwater, are diverse, dynamic ecosystems comprised of genetically diverse populations. Such diversity can be broad, as between species, as well as, within populations, as in strains of species. Typical metagenomic approaches explore this diversity in a manner than loses information the genomes of individual cells. In contrast, characterizations at the individual cell level yield information about interactions between organisms, such as between bacteria and phage, as well as strain level differences within a population. However, Public available genomes are good resources as reference for functional and taxonomy annotation, while most of them are culturable species from host-associated environments. Genomes of good quality and novelty are lacking to serve as references for not-yet cultured bacteria in natural environments. Thus, population genetic and evolutionary data analysis tools within the KBase environment and single cell sequencing of ENIGMA samples could have an outsize impact on environmental microbiology research.

Here we report new functions that we add to the KBase environment to catalyze metagenomic data analysis. First, we have built a standard and comprehensive set of reference genomes to which metagenomic reads can be compared. We have designed the pipeline to compare metagenomes to references and tested it in our samples. We built the estimators of strain level

nucleotide diversity, and even inference of strain genomes. We designed the tools to study within-population genome rearrangements and mutations. We published one compact tool, the meta_decoder (https://github.com/caozhichongchong/meta_decoder), that automatically identifies and compares the bacterial strains, mobile genetic elements, and phase variation across samples. We have tested meta_decoder using simulated datasets and we are now applying it to ENIGMA genomes and metagenomes.

We profiled 15,343 single genomes by droplet microfluidics (Microbe-seq) of an enigma groundwater sample GW-FW-305. We assembled many genomes, including 16 high quality bacterial genomes, > 75% completion and < 5% genomic redundancy. Several genomes were >97% complete and < 1% redundant, characteristics that are unusual for traditional metagenomes from a single sample. We uploaded these novel genomes with high quality (of completeness and contamination) into KBase as good references for not-yet cultured bacteria in natural environments.

We collaborate with ENIGMA data management team (John-Marc Chandonia) and KBase team (Dylan Chivian). The ENIGMA data we use is the Pseudomonas genomes from Lauren M. Lui (Arkin Lab).

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

Integrating Genomics, Physiology and Genetics of the Oak Ridge Field Research Site Microbiome

A.M. Deutschbauer[1,2](AMDeutschbauer@lbl.gov) H.K. Carlson[1], R. Chakraborty[1], V. Trotter[1], J.V. Kuehl[1], A. Kothari[1], A.E. Kazakov[1], M.N. Price[1], X. Wu,[1] T.R. Northen[1], J.-M. Chandonia[1], P.S. Novichkov[1], A. Mukhopadhyay[1], **A.P. Arkin**[1,2] and **P.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.

Over many years of isolation efforts, ENIGMA has leveraged a network of groundwater wells at the Oak Ridge Field Research Site in which multi-dimensional gradients of inorganic ion contaminants are a complex combinatorial constraint on the microbiome to amass a large culture collection. The microbial isolates in this collection represent a significant percentage of the culturable microbial populations from this contaminated aquifer. Genome sequencing of hundreds of these isolates has revealed both macro and microdiversity, and the isolates display a range of phenotypes with respect to metal resistance, energy metabolism and nutrient utilization. We are using this phenotypic variation to infer the probable niche and range of the various sub-populations. Comparison of laboratory phenotypic characterization with field survey data enables the inference of probable selective pressures acting on the microbiome in the field. Through streamlining genetics and scaling gene function discovery in our isolates, we are linking specific genetic mechanisms of survival or sensitivity to field relevant selective pressures. The observations and tools developed through these efforts are being propagated through publicly available web interfaces and KBase to facilitate new discoveries from across the scientific community.

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

Selective Carbon Sources Influence the End-Products of Microbial Nitrate Respiration

H. K. Carlson[1] (**hkcarlson@lbl.gov**), L. M. Lui[1], M. N. Price[1], A. E. Kazakov[1], A. V. Carr[3], J. V. Kuehl[1], T. K. Owens[1], T. Nielsen[1], A. M. Deutschbauer[1], **A.P. Arkin**[1,2], **P.D. Adams**[1,2]

[1]**Lawrence Berkeley National Lab, Berkeley CA**; [2]University of California at Berkeley, Berkeley CA; [3]University of Washington, Seattle WA;

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: Respiratory and catabolic pathways are differentially distributed in microbial genomes. Thus, specific carbon sources may favor different respiratory processes. We profiled the influence of 94 carbon sources on the end-products of nitrate respiration in microbial enrichment cultures from diverse terrestrial environments. We found that some carbon sources consistently favor dissimilatory nitrate reduction to ammonium (DNRA/nitrate ammonification) while other carbon sources favor nitrite accumulation or denitrification. For an enrichment culture from aquatic sediment, we sequenced the genomes of the most abundant strains, matched these genomes to 16S rDNA exact sequence variants (ESVs), and used 16S rDNA amplicon sequencing to track the differential enrichment of functionally distinct ESVs on different carbon sources. We found that changes in the abundances of strains with different genetic potentials for nitrite accumulation, DNRA or denitrification were correlated with the nitrite or ammonium concentrations in the enrichment cultures recovered on different carbon sources. Specifically, we found that either L-sorbose or D-cellobiose enriched for a *Klebsiella* nitrite accumulator, other sugars enriched for an *Escherichia* nitrate ammonifier, and citrate or formate enriched for a *Pseudomonas* denitrifier and a *Sulfurospirillum* nitrate ammonifier. Our results add important nuance to the current paradigm that higher concentrations of carbon will always favor DNRA over denitrification or nitrite accumulation, and we propose that, in some cases, carbon composition can be as important as carbon concentration in determining nitrate respiratory end-products. Furthermore, our approach can be extended to other environments and metabolisms to characterize how selective parameters influence microbial community composition, gene content and function.

References

1. Selective carbon sources influence the end-products of microbial nitrate respiration; Hans K. Carlson, Lauren M. Lui, Morgan N. Price, Alexey E. Kazakov, Alex V. Carr, Jennifer V. Kuehl, Trenton K. Owens, Torben Nielsen, Adam P. Arkin, Adam M. Deutschbauer
bioRxiv 829143; doi: https://doi.org/10.1101/829143

Spatiotemporal Dynamics of Groundwater and Sediment: Geochemistry, Microbial Communities and Activities in a Contaminated Aquifer

K. F. Walker[1,2](kfitzge4@utk.edu), E. R. Dixon[1,3], D. C. Joyner[1,3], K. A. Lowe[1], F. L. Poole[4], X. Ge[4], M. P. Thorgersen[4], D. Ning[5], Y. Fan[5], J.P. Michael[5], Y. Fu[5], R. Tian[5], Y. Wang[5], J.D. Van Nostrand[5], L.M. Lui[6], X. Wu[6], K.J. Davis[7], M.W.W. Adams[4], R. Chakraborty[6], D. A. Elias[1], R. L. Wilpiszeski[1], J. Zhou[5,6], M.W. Fields[7], **T. C. Hazen**[1,2,3,6], A.P. Arkin[6] and **P.D. Adams**[6]

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

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.

Spatiotemporal variability of groundwater levels and sources could greatly impact the geochemistry and the drift, dispersal, and selection of associated microbial communities. As a pilot study, one uncontaminated site and two contaminated sites (Area 2 and Area 3) at the DOE Oak Ridge Reservation Y-12 Complex in Oak Ridge, Tennessee, were selected for bi-weekly sampling. Three groundwater wells at each site were sampled for geochemistry and microbial activity measurements. Groundwater was also filtered in succession through 10 µm, 0.2 µm and 0.1µm filters for assessment of microbial communities (16S and metagenomic). After completing five sampling time points from March to May 2019, the preliminary results of bacterial communities revealed distinct succession patterns affected by contamination and cell size. For the uncontaminated site and Area 2, total microbial cell counts ranged between 5 to 18 x 10^6 cells /ml and the counts for Area 3 averaged between 2 to 8 x 10^6 cells/ml. Cell counts for most contaminated site (Area 3) were more consistent over time when compared to the other two sites. Total microbial activity was assessed via the uptake of ^{3}H-leucine. Activity measurements changed temporally and were an order of magnitude higher for uncontaminated wells compared to the contaminated wells (1-3 x 10^{-5} ng C/cell/d versus 0.6-8 x 10^{-6} ng C/cell/d, respectively). The changes in total microbial activity did not always correlate to changes in microbial cell numbers. These results suggest that not all microbes were active over the same times and places (*i.e.*, mechanisms of dispersal and selection were likely impacting different populations spatiotemporally). The bacterial communities from contaminated wells (Area 2 and 3) were similar in diversity and structure compared to those in uncontaminated wells. The unique ESVs

respective to each well (contaminated or uncontaminated) were typically lower in abundance compared to ESVs detected across areas and wells. The results supported environmental filtering, particularly in contaminated wells, for unique, low-abundance populations. In addition, detected ESVs from the 0.1μm filters showed decreased relative abundance of area-specific species over time, but ESVs from larger fractions (0.2μm to 10μm) did not. While larger-size bacteria from different wells always showed different community structure, small-size bacteria from different areas became convergent after late April. This implicates the influence of migration, corresponding to precipitation changes in April and May that coincided with observed changes in the δO^{18} values of groundwater. While α- and/or β-Proteobacteria generally dominated in larger-size bacteria, the phylum Bacteroidetes significantly increased or even predominated in the small-size fraction after late April, mainly attributed to the genus *Hydrotalea*. Sulfate-reducing bacteria, a relevant functional group at this site, belonged to the orders Syntrophobacterales and Desulfobacterales, which were nearly undetectable in small-size bacteria. In contrast, *Rhodanobacter*, a dominant genus in contaminated wells of this site, was detectable in different size fractions and showed decreased relative abundance after late April. These preliminary data demonstrated the value of more frequent sampling for an in-depth time series analysis.

Starting in July 2019 a comprehensive, high-resolution time series survey of 27 wells was carried out to obtain diurnal and seasonal fluctuations within three levels (mild, moderate, and high) of nitrate and heavy metal contamination. With this data, we aim to model these areas and study changes within the attached and unattached microbial communities in relation to groundwater geochemistry. Measurements were gathered from 27 previously established groundwater wells four days/week over the span of 17 weeks (70 days total, July to December) to build both diurnal and seasonal time series. In-field geochemical measurements were obtained for dissolved oxygen (DO), pH, conductivity, oxidation-reduction potential (ORP), and nitrate concentration. Samples were also taken for metals, anions, organic acids, and total organic and inorganic C/N. Preliminary results show diurnal and seasonal changes in geochemistry with wide variations between each well and levels of contamination. Additionally, one well in each level of contamination (3 wells total) was selected to complete a “deep-dive” analysis by sampling for microbial communities in groundwater (unattached) and sediment (attached). Groundwater was filtered through 8μm and 0.2μm filters for 16S rRNA and metagenomic analysis for a total of 420 filters. In each of the three “deep-dive” wells, 18 unamended sediment traps were deployed throughout the sampling period in order to complete a time series soil analysis. The attached microbial communities and soil geochemistry will be compared to the unattached communities and groundwater geochemistry. Results for each stage of analysis will be linked to groundwater flow vectors and on-site weather data. With this data, we aim to establish a predictive systems model to understand potential distribution of microbial communities and associated activities in the shallow subsurface.

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

The ENIGMA Data Clearinghouse: A platform for rigorous self-validated data modeling and integrative, reproducible data analysis

John-Marc Chandonia[*,1] (JMChandonia@lbl.gov), Pavel S. Novichov[*,1], Adam P. Arkin, and **Paul D. Adams**[1,2]

[1]Lawrence Berkeley National Lab, Berkeley; [2]University of California at Berkeley; [*]co-first authors

http://enigma.lbl.gov

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

One of the Grand Challenges of data science is to facilitate knowledge discovery by enabling datasets to be readily analyzable both by humans and by machine learning algorithms. In 2016, a diverse group of stakeholders formalized a concise and measurable set of principles, called FAIR, to increase the utility of datasets for the purpose of knowledge discovery. The four principles of FAIR are Findability, Accessibility, Interoperability, and Reusability. *Findability* means that data are assigned stable identifiers, and properly indexed. *Accessibility* means the data are easily retrievable by people authorized to have access. *Interoperability* means the data are clearly documented using a formal language, in order to facilitate integrated analyses that span multiple datasets. *Reusability* means the data are documented sufficiently well that it may be used by other people than those who originally generated it, and that the provenance of all data is clear.

The latter two principles are particularly challenging, yet critical to achieve, for organizations such as ENIGMA that draw conclusions based on highly integrative analyses of many types of data generated by multiple labs. *Reusability* can be difficult because non-specialized data formats often do not allow or require specification of key details, even basic ones such as units of measurement. As a result, it can be challenging to reproduce or reuse data, because of undocumented assumptions and conventions. Ensuring *Interoperability* between datasets is hard for many of the same reasons: when different teams within an organization produce data, impedance matching must be done in order to perform an integrative analysis. Some sources of impedance are differing units, incompatible scaling or normalization of different datasets, and different identifiers used by different teams to refer to the same objects.

We surveyed hundreds of data types throughout ENIGMA, and discovered that the vast majority of data (from raw assays to processed results) can be represented by a limited number of mathematical data models, such as multi-dimensional arrays of scalars. We believe that this result is generalizable across many fields of research, and indeed, storage formats such as HDF5 and NetCDF-4, along with libraries such as Xarray, are well supported and widely used technologies. However, a common file format alone is not sufficient to ensure adherence to the FAIR principles of *Interoperability* and *Reusability*: in

addition, all contents, dimensions, and units in these multidimensional arrays must be formally and rigorously documented.

We developed the ENIGMA Data Clearinghouse, the first general-purpose platform that solves this problem. This relies on three key technologies: 1) to rigorously document context for all data, we introduce the concept of a "contexton," or unit of context. Contextons are built using "microtypes," which we define as atomic data types representing a simple concept relevant to a domain of interest. Both rely on ontologies, which define a controlled vocabulary for describing a domain of interest. Together, these microtypes and ontologies represent a language that allows users to formally describe all data in that instance in a way that is both *Interoperable* and *Reusable*. 2) Dynamic data types, which make up the vast majority of data, are defined by the users of the system as they are needed, by combining commonly used mathematical data structures with contextons. This "building blocks" approach enables new data types to be defined as needed, with low costs, but also ensures that they are documented in the formal and rigorous manner that is necessary for *Interoperability* and *Reusability* of the data. A limited number of static core types, which are fully specified traditional data structures, are also built using contextons in order to ensure *Interoperability* with the dynamic data. These static core types include the system type *Process*, which is a special core type needed to document the provenance of each data object. 3) All static and dynamic data are referenced in an object graph, where nodes are static or dynamic datasets, and edges are processes. This graph formally annotates the provenance of all data.

In addition to storing ENIGMA data, the Data Clearinghouse includes rich functionality to make the system useful for data analysis, visualization, and managerial oversight. This functionality includes graphing tools, advanced search, an upload wizard, and an API for merging data. ENIGMA data scientists access the Clearinghouse through Jupyter notebooks, running in a shared directory of a server running JupyterHub.

We are also collaborating with the KBase project to harden and deploy these technologies for use by ENIGMA team members as well as all other KBase users. Until the time when these technologies are deployed for general use in KBase, we plan to continue to develop our API and store current ENIGMA datasets using the Data Clearinghouse server. This will ensure that all current ENIGMA data types are compatible with our tools, and that all data can be seamlessly transferred to KBase when our technology is deployed there.

ENIGMA is a Scientific Focus Area Program at Lawrence Berkeley National Laboratory and is 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.

Optimal Experimental Design (OED) of Biological Systems

Francis J. Alexander,[1,*] (falexander@bnl.gov), Ian Blaby[2], Edward Dougherty[3], 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.

To accelerate strain improvement strategies, significant paradigm shifts in life science research are needed, particularly a move toward tool development for optimizing and controlling highly uncertain systems. 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. One valid and successful approach is to understand each component (regulation, transport, biochemical pathway, etc.) of the system via experiments that capture specific data points. However, data derived from biological experiments can be as complex as the organisms from which said information is collected. Data sets and experimental results 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. For example,

a system designed for elevated levels of a given metabolite may not differ significantly from the parent strain because of unknown pathway branch points or regulation (i.e., "metabolic buffering capacity").

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. To achieve this goal, the team is using a multidisciplinary approach involving two interlinked aims. Aim 1 identifies, adapt, and implement the necessary algorithms to make OED applicable to biological problems by reducing the cost of uncertainty in cellular metabolism. This is achieved by inferring the transcriptional regulatory network (TRN) from *E. coli* gene expression compendia using a Bayesian network (BN) and employing the gene expression values predicted by the BN under various control actions/conditions to infer their impacts on the metabolite yield. The metabolic outcomes will then be predicted through flux balance analysis (FBA) with proper constraints on the metabolic pathways regulated by the TRN modeled by the BN. Aim 2 will define, execute, and iterate genome-scale engineering approaches guided by MOCU and MOCU-based OED using the aforementioned models in Aim 1. Realization of Aim 1 will yield robust predictions, even in the face of uncertainty (i.e., incomplete information), informing the model. In Aim 2, the necessary genome manipulations guided by these predictions will be performed to quantitatively assess their success and iterate.

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

Evolution and Metabolic Configuration of Nitrogen Flux in a Model Marine Diatom

Sarah Smith[1*], Christopher Dupont[1], McCarthy James[1], Zoltán Füssy[1], Kim Hixson[3], Bernhard O. Palsson[3], **Andrew E. Allen**[1,4]

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037, USA; USA; [2]Pacific Northwest National Laboratory, Environmental Molecular Sciences Laboratory, Richland, Washington, 99352, USA; [3]University of California, San Diego, Division of Biological Sciences, La Jolla, California, 92093, USA; [2]Scripps Institution of Oceanography, Integrative Oceanography Division, University of California, San Diego, La Jolla, CA 92093[4]

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.

Abstract:

Diatoms dominate phytoplankton communities by outcompeting other groups for nitrate, yet little is known about the mechanisms underpinning this ability. Genome and genome-enabled studies have shown that diatoms possess unique metabolic features compared to other phototrophs, such as mitochondrial glycolysis and the presence of a urea cycle. In diatoms, the cycle is known to be important for recovery from nitrogen limitation, however there are open questions about how the cycle is integrated within the cell-wide metabolic network. To develop a whole-cell level understanding of the impact of nitrogen source and status on *Phaeodactylum tricornutum*, we investigated gene expression and metabolic flux in in experiments aimed at eliciting shifts in nitrogen status over the short term. Using a combination of transcriptomics, proteomics, metabolomics, fluxomics, and flux balance analysis, we have arrived at a systems-level understanding of how nitrogen is assimilated and distributed within *P. tricornutum*. We found a high degree of metabolic network connectivity between the chloroplast and mitochondria of pathways at the critical intersection of carbon and nitrogen metabolism. We characterize the differentiated function of organellar GS-GOGAT cycles and describe aspartate and alanine systems used to exchange amino moieties between organelles. We also describe an arginine biosynthesis pathway that is split across organelles in diatoms, clarifying the role of the urea cycle. We propose that the unique configuration and high degree of metabolic integration between the major energy organelles allows diatoms to efficiently respond to changing nitrogen status, conferring an ecological advantage over other phytoplankton taxa.

Supported by DE-SC0008593*, Biological and Environmental Research*

Identification of Transcription Factor Binding Sites and Characterization of Promoter Architecture in the model Diatom *Phaeodactylum tricornutum*

Andrew E. Allen[1,2*], Sarah R. Smith, Hong Zheng, Pratap Venepally

[1]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037, USA; [2]Scripps Institution of Oceanography, Integrative Oceanography Division, University of California, San Diego, La Jolla, CA 92093, USA.

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.

Abstract:

Transcription factors (TFs) regulate gene expression by binding DNA in gene promoters and have an important role in activating the cellular response to shifting environmental conditions. To date, several diatom transcriptome studies have shown that suites of genes are co-expressed in response to shifting conditions (i.e. nutrients or light), identifying putative gene regulons. However, little is known about the roles or binding sites of specific TFs that elicit these transcriptional responses as only a few TFs have been characterized in diatoms. We investigated transcription factor binding sites, using a combination of bioinformatics-based promoter analysis and high-throughput in vitro DNA affinity purification sequencing (DAP-seq). DAP-seq enables genome-wide characterization of transcription factor binding sites. We identify a nitrate-response element enriched in the promoters of nitrate assimilation genes that is similar to the binding site for the human transcription factor ETS. We also characterize the optimal binding sites for different genes in the bZIP transcription factor family from *P. tricornutum*, facilitating description of the genes they regulate. Specific knowledge of promoter architecture is valuable for the development of tools for molecular investigations and genetic engineering of diatoms and is essential in order to understand how reprogramming of gene expression is accomplished to achieve appropriate cellular responses to environmental signals.

Supported by DE-SC0008593*, Biological and Environmental Research*

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], Andrew Allen[4,5], **Jamey D. Young**[1,2]

[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

Pt 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 goal of the project is to promote efficient production of high-value and fuel-related compounds through optimization of metabolic fluxes in the diatom *Phaeodactylum tricornutum* (Pt), a model photosynthetic eukaryotic microbe. Our lab focuses on using ^{13}C metabolic flux analysis (MFA) to understand changes in metabolism by generating a quantitative flux map of the metabolic reaction network.[2,3] To create the flux map, we use experimental measurements of growth, product formation, and stable isotope labeling to constrain a model of central carbon metabolism. We developed a model of Pt based on genomic annotations and incorporated subcellular compartments to reflect the organization of central carbon metabolism within the cell. We combined our model with experimental isotope labeling studies to elucidate the metabolism of Pt under conditions relevant for biomanufacturing.

Our current focus is to investigate metabolic adjustments to variations in light and nitrogen availability, two variables which strongly impact Pt growth and lipid accumulation. Under low light growth conditions (60μE), we observed significantly higher chlorophyll content compared to the chlorophyll content at high light growth conditions (250μE). We also unearthed dramatic shifts in metabolic fluxes and pool sizes in Pt under nitrogen-limiting conditions. Particularly, most of the TCA cycle metabolite pool sizes were elevated while pool sizes of most nitrogen-containing metabolites, except urea, decreased significantly. Our near-term goal is to use ^{13}C MFA to understand how Pt metabolism adapts to various environmental conditions that are essential for maximizing TAG biosynthesis.

(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 Ma, F. et al. Isotopically nonstationary ^{13}C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation. *Proc Natl Acad Sci U S A.* **111**(47), 16967-16972, doi: 10.1073/pnas.131948511 (2014)

Temporal profiles unravel resource allocation mechanisms under nitrogen starvation in the diatom *Phaeodactylum tricornutum*

Juan D. Tibocha-Bonilla[1], Cristal Zuniga[1*] (crzuniga@eng.ucsd.edu), **Andrew E. Allen[2]**, and **Karsten Zengler[1]**

[1]University of California, San Diego; [2]J. Craig Venter Institute, San Diego

The goal of this project is to gain new insights into the photosynthetic eukaryote *Phaeodactylum tricornutum* for bioenergy production using a systems biology approach. We integrated time-course target metabolomics data into a metabolic modeling framework to systematically identify and quantify the partitioning of carbon and nitrogen between cellular metabolism, cross-talk between organelles, and channeling of photosynthetic electron flows.

Diverse conditions, i.e. growth during day and night, and compartmental cellular organization require phototrophs to shift their proteome demands and therefore adjust their metabolism and biomass composition during the course of growth. This complex interplay between energy and carbon metabolism and its dynamics in phototrophs is still not fully understood. Constraint-based modeling is a systems biology tool that contextualizes experimental data, such as uptake rates and biomass composition, for successful prediction of growth phenotypes. Currently, the lack of time-course biomass composition data has restricted prediction accuracy. Instead of recapitulating dynamic changes in cell composition, current models are forced to assume that the biomass remains constant. Here, we used experimentally determined metabolomics data to determine biomass composition constrains. Dynamic constraints were applied to our previously published genome-scale metabolic model of the diatom *P. tricornutum*[1]. We identified temporal profiles of metabolic fluxes that indicate long-term trends in pathways and organelle-specific activities in response to nitrogen depletion. Additionally, a growth rate sensitivity analysis of time-course flux distributions enabled identifying the main metabolite affecting growth (e.g. amino acids and lipids). Surprisingly, our dynamic simulations hinted at free energy instead of the molecular weight as the main drivers of biosynthetic cost. Our *P. tricornutum* temporal-flux-profiles will be used to enlarge the number of gene associations in the model by scanning dynamic transcriptomics data sets and by evaluating hidden metabolic and transport activities using our in-house developed tools[2].

References

1. Levering, J. *et al.* Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS One* **11,** 1–22 (2016).

2. Zuñiga, C. *et al.* Predicting dynamic metabolic demands in the photosynthetic eukaryote *Chlorella vulgaris*. *Plant Physiol.* **176,** 450–462 (2018).

Supported by DE-SC0018344, Biological and Environmental Research

Prolonged drought alters plant-litter decomposition via changes in bacterial communities and substrate availability

Sarai S. Finks[1*], Claudia Weihe[2*], Eoin L. Brodie[3], Michael L. Goulden[4], Adam C. Martiny[5], Kathleen K. Treseder[6], Jennifer B.H. Martiny[7*], and **Steven D. Allison**[8]
*sfinks@uci.edu, cweihe@uci.edu
[1-2, 4-8] University of California Irvine
[3] Lawrence Berkeley National Laboratory

Microbial decomposers of plant litter play an important role in carbon and nitrogen cycling in terrestrial systems. Past research has focused on the direct effects of global changes, such as increased temperature, added nitrogen, and reduced water, on microbial abundance and composition. However, there is limited understanding of the indirect effects that plant communities may have, via plant-microbe interactions, on the microbial response to climate change. Here we investigated how drought impacts bacterial communities: directly through abiotic changes in water availability to the microbes and/or indirectly through changes in plant communities. To do this, we first surveyed the bacterial communities on plant litter within a decade long global change experiment applied to two adjacent ecosystems (a grassland and coastal sage scrub, CSS). The Loma-Ridge Global Change Experiment (LRGCE) employs a split-plot design where a nitrogen addition treatment is nested within plots receiving either ambient rainfall or a ~50% reduction in rainfall. Bacterial composition was characterized for each treatment over 2.5 years (n = 214) by 16S rRNA sequencing. We then conducted a reciprocal transplant experiment (n = 192) over 1.5 years to disentangle the abiotic effects of drought versus plant community effects on microbial composition. The cages contained sterile plant litter from grassland or CSS ambient plots and were inoculated with grass or CCS microbial communities exposed to ambient or drought conditions. To assay the response of the bacterial communities to the separate factors, we characterized composition and their functioning (mass loss) over time . Within the LRGCE, bacterial community composition differed greatly between the grassland and CSS ecosystems (explaining 12% of compositional variation), indicating a strong effect of the plant-litter substrate on the decomposer community. Drought also significantly altered bacterial community composition across both ecosystems (4% of the variation). However, drought had a much stronger effect on bacterial communities in the grassland than the CSS, suggesting that the influence of drought might be mediated by the plant community. The main taxonomic drivers to microbial drought response were among the most abundant genera overall in grassland or CSS, namely, *Curtobacterium, Janthinobacterium*, *Pedobacter,* and *Methylobacterium*. The reciprocal transplant experiment confirmed an indirect role of the plant community in the bacterial response to prolonged drought. Lastly, decomposition of the litter substrate was strongly dependent upon water availability and the ecosystem the plant-litter was taken from. This work highlights the importance of considering plant-microbial interactions on the microbial response to climate change.

Functional Analysis of Candidate Genes Involved in Oil Storage and Stability in Pennycress

Emmanuel Ortiz[1]* (Emmanuel.Ortiz@unt.edu), Kent Chapman,[1] Erich Grotewold,[2] and **Ana Paula Alonso**[1]

[1]University of North Texas, Denton; and [2]Michigan State University, East Lansing

Project Goals: 1) To investigate the function of candidate genes involved in lipid storage and stability using transient expression in *Nicotiana benthamiana*. 2) To use stable transformation to evaluate the impact of these candidate genes on lipid droplets, biomass content, lipid spatial distribution and composition in pennycress plants. 3) To characterize the effect that co-expression of these genes has on central metabolism and the temporal organization of storage lipids in developing pennycress seeds.

The US military and commercial aviation industry together consume 20 billion gallons of jet fuel per year, the cost of which as more than tripled since 2000. In order to reduce the reliance on fossil fuels, and our environmental footprint as well as costs, it is imperative to develop renewable sources of aviation fuel. In recent years, pennycress (*Thlaspi arvense* L.) has been identified as a promising alternative oilseed crop suitable for aviation fuel production due to its high oil content and fatty acid composition. Biodiesel produced from pennycress exhibits excellent characteristics such as low could point temperature, pour point, and cold filter plugging point, with 50% less greenhouse gas emissions compared to petroleum-based fuel. Pennycress requires low agricultural inputs and can serve as a cover crop to utilize excess nitrogen, slow soil erosion, suppress weeds, and attract pollinators when grown in a summer/winter rotation cycle with other conventional commodity crops such as corn and soybean. Its yields reach 2,000 Lbs/acre, producing up to 100 gallons of oil per acre, which is twice more than soybean or camelina (another bioenergy crop under consideration). Given that this crop could be grown in the 90 million-acre US Midwest Corn Belt without displacing commodity crops, pennycress has the potential to produce 9 billion gallons of aviation fuel per year, nearly meeting half of the national demand.

Our group has been developing tools and resources to improve oil content and quality in pennycress (DE-SC0019233). Technoeconomic analyses of the costs associated with production and logistic operations revealed that pennycress would be more competitive than other crops considered for renewable jet fuel production, such as camelina, carinata, and canola. Also, the preliminary findings from our analyses of pennycress natural variation have identified candidate genes and metabolites associated with increased oil accumulation. These biomarkers are involved in lipid synthesis/degradation, and in primary metabolism. Of particular interest is the correlation between transcript levels of candidate genes involved in fatty acid storage and stability with total oilseed content at maturity. This finding underlines the central role of the packaging and storage of triacylglycerols in our quest to enhance oil production in plant organs.

While pennycress, as a member of family Brassicaceae, 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. A detailed functional genomics study of these genes will be the focus of this project. To achieve this goal, we will first evaluate the function and subcellular localization of 43 priority candidates in *N. benthamiana*. In parallel, we will screen for changes in lipid droplet morphology and organization as well as variations in overall biomass and

lipid composition. We will then test up to 15 single/combined candidates on pennycress plants using stable transformation. Seeds from homozygous mutant plants and wild type will be examined to evaluate the impact of the mutation(s) on lipid droplets, biomass content, lipid spatial distribution and composition. Finally, we will analyze the effect that co-expression of these genes has on the pathways of central metabolism. We will also assess the impact on the temporal organization of storage lipids in developing seeds using ^{13}C-labeling and MALDI-MS imaging.

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.

Natural Variation of Pennycress Metabolome and Transcriptome, an Emerging Crop for Aviation Biofuel

Cintia Arias[1*](cintia.arias@unt.edu), Emmanuel Ortiz,[1] Christopher Johnston,[1] Tatiana Garcia,[2] Mukundi Eric,[2] Fabio Gomez-Cano,[2] Tyler Swanson,[1] Fan Yang,[3] Erich Grotewold,[2] and **Ana Paula Alonso**[1]

[1]University of North Texas, Denton, TX; [2]Michigan State University, East Lansing, MI; [3]The Ohio State University, Columbus, OH

Project Goals: Investigating pennycress natural variation to identify candidate genes and biomarkers associated with oil accumulation.

Pennycress (*Thlaspi arvens*e) is a member of family Brassicaceae which produces seed oil ideally suited for use as a renewable source of biodiesel and aviation fuel[1]. However, for this plant to become an economically viable and sustainable bioenergy crop, molecular and genetic resources must be developed to guide strategies for increasing oil production through breeding and/or genetic manipulation. In this work, we investigated the natural variation of central carbon metabolism in pennycress among twenty-one accessions sourced from distinct geographic regions to identify candidate genes and biomarkers associated with oil accumulation. The biomass components, targeted metabolome, and transcriptome of each line were analyzed at two different stages of embryo development, and positive and negative correlations between the levels of intracellular metabolites and transcripts with oil accumulation were identified. Additionally, two accessions with diverse oil content (30% *vs* 40%), were analyzed in greater detail under more controlled conditions at five developmental time points. One of the pathways that illustrated the largest metabolic variation between the two lines was the tricarboxylic acid cycle, in which intermediaries displayed significant differences. Also, the content of several compounds involved in sucrose biosynthesis and/or degradation suggest contrasting regulation of sugar metabolism. Altogether, this work has identified oil biomarkers that may prove useful in the breeding and rational bioengineering of pennycress.

References

1. Fan, J.Q., Shonnard, D.R., Kalnes, T.N., Johnsen, P.B. & Rao, S. A life cycle assessment of pennycress (Thlaspi arvense L.) -derived jet fuel and diesel. *Biomass & Bioenergy* **55**, 87-100 (2013).

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0019233.

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

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

[a]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[b]Department of Bacteriology, University of Wisconsin – Madison, Madison, Wisconsin, USA
[c]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, USA
[d]Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire, USA

Project Goals: To develop experimental-computational approaches for *in vivo* genome-scale determination of Gibbs free energies (ΔG) in metabolic networks.

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*. The use of a near equilibrium glycolytic pathway, with potentially increased ATP yield, by *C. thermocellum* may represent an evolutionary adaptation to growth on cellulose but it has the drawback of being highly susceptible to product feedback inhibition. The results of this study will facilitate future

engineering of high-performance strains capable of transforming cellulosic biomass to biofuels at high yields and titers.

Investigating Seed Size and Oil Content in Pennycress, *Thlaspi arvense*

Katherine Frels[1*] (kfrels@umn.edu), Ratan Chopra[1], Zenith Tandukar[1] John Sedbrook[2], M. David Marks[1], and **James A. Anderson**[1]

[1]University of Minnesota, St. Paul, MN; [2]Illinois State University, Normal, IL

Project Goals: Our goal is to characterize and improve seed size and oil content in pennycress to improve its efficiency and utility as a biofuel feedstock and make the seed easier for producers to handle. To advance towards this goal, we aim to: 1) Investigate the genetic control of these traits in a wild-germplasm collection using association mapping and biparental linkage mapping to identify quantitative trait loci (QTL); 2) Identify and characterize these traits in EMS induced mutation lines using high-throughput screening tools in combination with next generation sequencing techniques; and 3) Generate pennycress CRISPR-Cas9 knockouts in the genes known to regulate these traits in *Arabidopsis*.

Pennycress is a winter hardy cover crop that provides ecosystem services such as reduced soil erosion and nutrient loss in between fall corn harvest and spring soybean planting. Unlike traditional cover crops, field pennycress produces a mature oilseed in late spring, allowing farmers to harvest two cash crops in one year. Wild-derived pennycress lines have been shown to yield >1,000 kg ha^{-1} [1]. Pennycress seeds contain on average 33% oil by weight, and the oil is an excellent biofuel feedstock. However, despite these environmental and economic benefits, pennycress is currently limited by its small seed size (1 mg/seed), which can complicate planting, harvesting, and handling of the seed. Increasing seed size would also increase the efficiency of oil extraction. In conjunction with improved seed size, increased oil content in the seed would also improve the economics of growing and processing of pennycress as a biofuels feedstock. We have collected wild pennycress accessions representing genetic diversity from North America, Europe, and western Asia[2]. By characterizing these accessions for seed size and oil content, we can identify useful variants for improvement. With USDA NIFA funding, we previously developed several EMS-induced pennycress mutant lines exhibiting key domestication traits such as reduced seed pod shatter, earlier flowering, and improved fatty acid profiles[3]. We have also developed and demonstrated the utility of pennycress *Agrobacterium*-mediated plant transformation and CRISPR-Cas9 genome editing by generating pennycress lines with undetectable levels of erucic acid in seed oil[4]. Using these recently developed techniques and germplasm, our goal is to identify and characterize traits that will improve pennycress efficiency and utility as a biofuel feedstock species and make the seed easier for producers to handle. Finally, we will introgress these traits into our elite breeding lines to develop new pennycress varieties with increased oil and seed yield.

To complete these objectives, we have compiled a pennycress association mapping panel with 319 genotypes (267 winter-type, and 52 spring-type individuals). The panel was planted in St.

Paul, MN for the 2018-2019 growing season. Harvested seed was screened for size using a Marvin Seed Analyzer and for oil content using NIRS. Phenotypic variation for both seed size and oil content was observed and will be combined with genotyping-by-sequencing derived genetic markers to identify marker-trait associations. We expect that the genotypic analysis will help us identify QTL associated with seed size and oil content and that these QTL can be used for marker assisted selection to develop improved breeding lines with larger seeds and higher oil content. In addition to using natural variation, we have also employed mutagenesis and gene editing techniques to rapidly improve seed size and oil content. We screened a pennycress ethyl methanesulphonate (EMS) mutagenesis population containing approximately 15,000 M_2 plants for larger seed size. One thousand mutant lines were identified for further screening, and 15 lines were identified as large seed mutants without lethal embryo phenotypes. Thousand seed weight for these lines ranged from 1.3 to 1.7 g compared to the wild type average of 1.1 g. These lines will be grown in the field to test the trait heritability and if the increase in seed size causes a reduction in seeds per pod under normal growing conditions. If the large seed size is heritable, we will work to identify the causative gene and develop a genetic marker for use in marker assisted trait introgression. We have generated high oil pennycress mutants using CRISPR gene editing and confirmed the phenotype in field trials. The *tt8* mutant has 10% higher oil content and has shown growth and yield characteristics agronomically indistinguishable from wild-type pennycress.

References

1. Cubins, J. A. *et al.* Management of pennycress as a winter annual cash cover crop. A review. *Agronomy for Sustainable Development* **39**, (2019).
2. Frels, K. *et al.* Genetic diversity of field pennycress (Thlaspi arvense) reveals untapped variability and paths toward selection for domestication. *Agronomy* **9**, (2019).
3. Chopra, R. *et al.* Identification and stacking of crucial traits required for the domestication of pennycress. *Nat. Food* **1**, 84–91 (2020).
4. McGinn, M. *et al.* Molecular tools enabling pennycress (Thlaspi arvense) as a model plant and oilseed cash cover crop. *Plant Biotechnol. J.* **17**, 776–778 (2019).

This research was supported by the Plant Feedstock Genomics for Bioenergy: A Joint Research Funding Opportunity Announcement USDA, DOE, award no. 2019-67009-29004.

KBase : Omics driven discovery of novel functional capabilities in biological systems

Janaka N. Edirisinghe*[2] (janakae@anl.gov), Benjamin Allen[3], Jason Baumohl[1], Jay Bolton[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Paramvir Dehal[1], Meghan Drake[3], José P. Faria[2], Annette Greiner[1], Tianhao Gu[2], James Jeffryes[2], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Akiyo Marukawa[1], Sean McCorkle[4], Cheyenne Nelson[1], Dan Murphy-Olson[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyenth Terry[1], James Thomason[5], Doreen Ware[5], Pamela Weisenhorn[2], Elisha Wood-Charlson[1], Shinjae Yoo[4], Qizhi Zhang[2], **Robert Cottingham[3], Chris Henry[2] and Adam P. Arkin[1]**

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

Project Goals: Discovery and characterization of novel biochemistry is key to understanding the behavior of complex biological systems. To explain a microbe's role in biogeochemical cycles, or synthesis of various biofuel products from plant biomass, or to better understand community interactions within complex biological systems, it is essential to identify poorly explained or novel biochemical pathways. Though recent advances in system biology and the exponential growth of reference data suggest that there is an enormous amount of untapped enzyme potential, the traditional approaches for discovery of new functions and pathways are still mostly trial and error experimental processes. Using KBase, we demonstrate a mechanistic-modeling-based approach coupled with multi-omics data to support high throughput discovery of new metabolic pathways.

While there are system-biology approaches that have been developed in isolation for novel pathway identification, it has been a great challenge to translate raw data into an improved understanding of microbe-mediated chemical transformations in degradation and biosynthesis pathways. In KBase, users will be able to combine genomic, metabolic modeling and cheminformatic predictions reconciled with omics-data to predict novel pathway reactions as depicted in Figure 1. Currently, users are able to upload sequencing data then assemble, annotate and build metabolic models from isolate genomes or from metagenomes. Multi-omics data such as chemical abundance data (e.g.; FTICR data, MS2 metabolomics data) (Fig 1D), RNA-seq data (Fig 1E), and proteomics data (Fig IF) can then be mapped onto models and applied to predict active known pathways, reactions, and genes. This is done by optimizing the extent to which active reactions in the models are associated with genes with high expression, quantitatively measured enzyme levels and positively identified metabolites. However, it is often the case that much of the chemical abundance data does not get mapped onto the existing biochemistry databases, which suggests that many functioning pathways remain unidentified.

To predict these new pathways and compounds, we use a cheminformatics pipeline in order to expand the existing biochemistry based on enzymatic and spontaneous chemical rules, then reconcile against unmapped metabolomics data (Fig. 1C-D). With a continuously interconnected

metabolic network of known and predicted reactions, it is feasible to activate all pathways that are implicated based on omics data. While this strategy can generate multiple probable pathway routes, we would be able to filter out high confidence pathways based on thermodynamic feasibility of the predicted reactions (Fig. 1F) and also by eliminating the pathways that lack metabolomic, transcriptomic, or proteomic evidence (Fig 1. D, F, E). By having a set of high confidence novel reactions, we would be able to map potential gene candidates via chemical and or structural based gene finding approaches (reaction finger-printing) (Fig 1C, D2, D3) coupled with genes that are expressed according to transcriptomic and/or proteomic data. Finally, a subset of these predicted genes can be validated experimentally (Fig 1J).

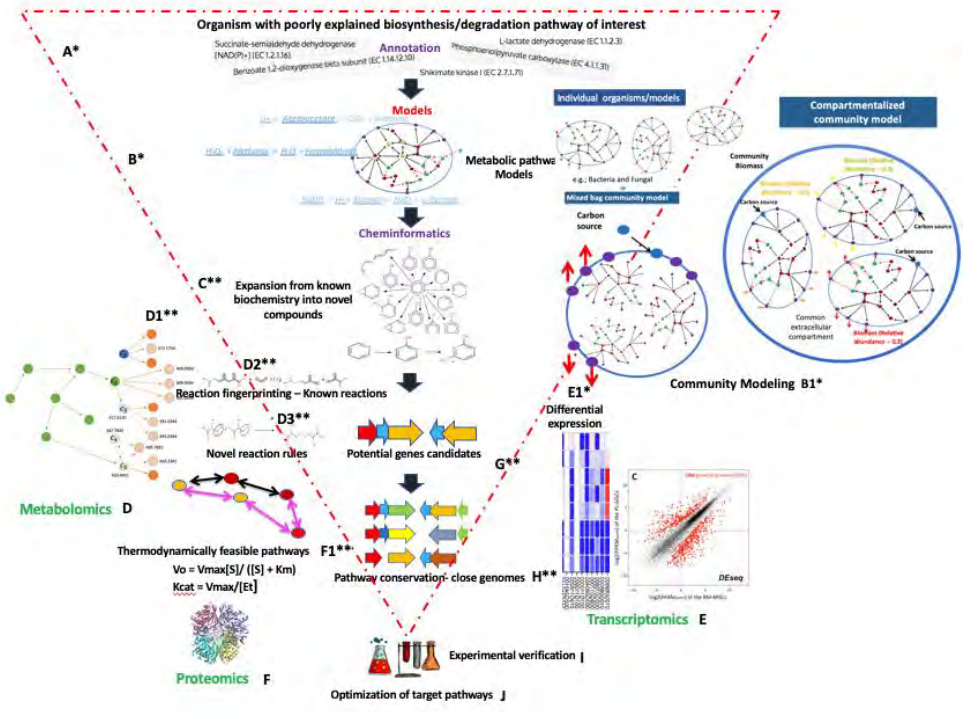


Figure 1. The figure shows the pathway discovery pipeline that has being implemented in DOE KBase and the integration of omics data at certain levels. The red triangle indicates the process of narrowing down into few target genes from thousands of genes in a genome for experimental validation. Steps labeled with a single asterisk (*) indicate that the specific component of the pipeline has already been implemented in KBase while the labels with a double asterisk (**) indicate that the functionality is still being developed. **B1**: **Community Modeling** - Construction of community models from individual models. There are two types of models that can be built in understanding the microbial communities. (i) Mixed-bag models (on left): Able to assess the overall metabolic capability of a community (ii) Compartmentalized community models (on right) able to assess the contribution of each member in community

KBase is funded by the Genomic Science program within 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: Leveraging Amplicon Analysis Tools to Generate Testable Hypotheses From Complex Natural Communities

Pamela Weisenhorn*[2](pweisenhorn@anl.gov), Benjamin Allen[3], Jason Baumohl[1], Jay Bolton[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Paramvir Dehal[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Annette Greiner[1], Tianhao Gu[2], James Jeffryes[2], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Akiyo Marukawa[1], Sean McCorkle[4], Cheyenne Nelson[1], Dan Murphy-Olson[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyenth Terry[1], James Thomason[5], Doreen Ware[5], Elisha Wood-Charlson[1], Shinjae Yoo[4], Qizhi Zhang[2], **Robert Cottingham[3], Chris Henry[2], Adam P. Arkin[1]**

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

http://kbase.us

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

The U.S. Department of Energy (DOE) supports biological and environmental research to investigate the complex interactions within biological systems and the processes that shape soil, water, and ecological dynamics of our biosphere, and to harness these processes for sustainable production of energy and materials. KBase enables researchers to advance our fundamental knowledge of complex biological and environmental systems by providing the computing infrastructure necessary to integrate and analyze heterogenous data types and share their findings with the broader community. By simplifying data integration and exploration, KBase enables researchers to identify patterns in their data and provides workflows to move beyond patterns to predictive understanding of processes.

For many researchers, amplicon datasets (e.g. 16S targeted metagenomes) often provide the first insights into the dynamics and functioning of complex natural or synthetic microbial communities. Recently implemented tools in KBase will allow researchers to use these amplicon datasets to examine: patterns in taxon abundance at different taxonomic levels; differences in the overall composition of communities in response to experimental treatments or environmental conditions; and microbe-microbe interactions in relation to their environment. While many of these approaches can be implemented independently, the computational infrastructure of KBase allows researchers to use publicly available data and a breadth of bioinformatic tools to quickly

move beyond mere identification of patterns and begin to explore the potential underlying mechanisms.

Using data collected as part of the Argonne Wetland Hydrobiogeochemistry project, we demonstrate how the amplicon tools in KBase can be used to identify patterns in microbial community composition and dynamics in response to environmental heterogeneity. We then demonstrate how KBase's diverse and integrated capabilities allow researchers to maximize the impact of their data and accelerate scientific discovery through deeper exploration of these patterns. Specifically, we demonstrate how connection between amplicon data in KBase and application of KBase's metabolic modeling, genome comparison, and auxotrophy prediction tools can generate testable hypotheses regarding the mechanisms underlying predicted microbe-microbe interactions. These predictions can then be used for the efficient design of focused experiments or field campaigns.

The ability to easily explore patterns and predictions both within and across projects will continue to be advanced through implementation of metadata standards for environmental samples and taxonomic abundance matrices (currently under development by DOE Environmental System Science's Data Infrastructure for a Virtual Ecosystem project). We discuss how these standards will be used to facilitate import of data from the ESS-DIVE archive and the role of such cross-platform standards in advancing our fundamental understanding of complex biological and environmental systems.

KBase is funded by the Genomic Science program within 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. Data were collected by the Argonne Wetland Hydrobiogeochemistry SFA which is funded by the Subsurface Biogeochemical Research Program, under Contract DE-AC02-06CH11357.

KBase: Microbiome and Phylogenomics Capabilities

Dylan Chivian*[1] (DCChivian@lbl.gov), **Adam P. Arkin[1]**, **Robert Cottingham[3]**, **Chris Henry[2]**, Benjamin Allen[3], Jason Baumohl[1], Jay Bolton[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Zachary Crockett[3], Paramvir Dehal[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Annette Greiner[1], Tianhao Gu[2], James Jeffryes[2], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Akiyo Marukawa[1], Sean McCorkle[4], Cheyenne Nelson[1], Dan Murphy-Olson[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyenth Terry[1], James Thomason[5], Doreen Ware[5], Pamela Weisenhorn[2], Elisha Wood-Charlson[1], Shinjae Yoo[4], Qizhi Zhang[2]

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

http://kbase.us

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

KBase was designed to enable systems biology analysis of communities of microbes and/or plants. KBase is extensible and currently includes powerful tools for metabolic modeling, comparative and phylogenomics of microbial genomes that can be used for developing mechanistic understanding of functional interactions between species in microbial ecosystems. Essential to gaining new insight is obtaining high-quality genomes to annotate, either via cultivation or genome extraction, from metagenome assembly. KBase has incorporated and added to a suite of microbiome analysis apps meant to be used in concert, including sequence QA/QC tools such as Trimmomatic and FastQC, taxonomic structure profiling of shotgun metagenome sequence with Kaiju, custom KBase apps for generating sample-specific *in silico* reads for downstream benchmarking, several metagenome assembly algorithms including MEGAHIT, IDBA-UD, and metaSPAdes, custom KBase apps for comparing metagenome assemblies, grouping assembled genome fragments (contigs) into putative genomes (bins) with MaxBin2 and other binners, and genome completeness and contamination assessment and filtering with CheckM. Tools for fractionation of unassembled reads and unbinned contigs to permit taxonomic and functional assessment of unbinned portions of samples are offered, and can also be used to reassemble individual bins of interest. Additionally, we've recently released tools and services that allow users to search rapidly (seconds to minutes) all reference genome databases, metagenomes and published metagenome-assembled genomes (MAGs) using their

reads, assemblies or MAGs. This is implemented using a MinHash like sketching process that works well for identifying matches above ~90% identity.

We have greatly expanded microbiome analysis in KBase. In addition to support for amplicon-based analyses (please see poster KBase: Leveraging Amplicon Analysis Tools to Generate Testable Hypotheses From Complex Natural Communities), it is now possible to incorporate and use tools that enable users to get from shotgun reads through to MAGs to phylogenomics and metabolic modeling. As an example from our initial set of tools, a user can upload or find data from collaborators or the public and apply one of the metagenome assembly apps and bin the assembled contigs so that individual genomes can be extracted from the bins. Once individual MAGs are extracted, the highest quality fraction can be piped into a wide range of downstream analysis apps in KBase, including genome annotation, phylogenetic placement and genome content comparison with respect to one another, KBase reference genomes, and other public genome and MAG collections. Unbinned metagenome assemblies can also have gene annotation for analysis. For high-quality MAGs, metabolic modeling and RNA-seq alignment can be performed (please see poster KBase: Omics driven discovery of novel functional capabilities in biological systems). Pangenome calculations among related genomes can be combined with phylogenetic and functional analysis to capture evolutionary histories of gene families and allow researchers to investigate functional repertoires and niche roles of microbial lineages.

In addition to efforts by KBase developers to expand the functionality of our Microbiome tool suite, community developers have been adding tools that they use and have developed, including members of the DOE Joint Genome Institute (BBMap, MetaBAT2, RQCFileter, JGI Metagenome Assembly Pipeline), the ENIGMA SFA, the LLNL Soils SFA (vConTACT2, VirSorter), and the LANL Bacterial:Fungal Interactions SFA (GOTTCHA2). All Apps in KBase are openly available for users to apply with their own data.

KBase is funded by the Genomic Science program within 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: The Systems Biology Knowledgebase for Predictive Biological and Environmental Research in an Integrated Data Platform

Adam P. Arkin[1], **Robert Cottingham**[3], **Chris Henry**[2], Benjamin Allen*[3] (allenbh@ornl.gov), Jason Baumohl[1], Jay Bolton[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Zachary Crockett[3], Paramvir Dehal[1], Meghan Drake[3], Janaka N. Edirisinghe[2], José P. Faria[2], Annette Greiner[1], Tianhao Gu[2], James Jeffryes[2], Marcin P. Joachimiak[1], Sean Jungbluth[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sebastian Le Bras[1], Zhenyuan Lu[5], Akiyo Marukawa[1], Sean McCorkle[4], Cheyenne Nelson[1], Dan Murphy-Olson[2], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Boris Sadkhin[2], Samuel Seaver[2], Alan Seleman[2], Gwyenth Terry[1], James Thomason[5], Doreen Ware[5], Pamela Weisenhorn[2], Elisha Wood-Charlson[1], Shinjae Yoo[4], Qizhi Zhang[2]

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

http://kbase.us

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

The U.S. Department of Energy (DOE) supports biological and environmental research to investigate the complex interactions within biological systems and the processes that shape soil, water, and ecological dynamics of our biosphere, and to harness these processes for sustainable production of energy and materials. KBase is an open-source data-science platform funded by DOE to enable sharing, integration, and analysis of many types of data associated with microbes, plants, and their communities using scalable computing infrastructure. This extensive public resource is designed to facilitate large-scale analyses of biological and environmental systems while accelerating scientific discovery, improve reproducibility, and foster open collaboration.

KBase offers a suite of scientific applications to enable users to build sophisticated analytical workflows, share their findings, and organize their projects. Over 200 apps in KBase offer diverse scientific functionality across the realms of comparative genomics, microbial communities, metabolic modeling, and transcriptomics. Several tools and services in KBase have been co-developed with the DOE Joint Genome Institute. Users can build and share sophisticated workflows through a combination of chaining together multiple analysis tools, writing scripts for automation, and using batch processing, all within notebook-style *Narrative*s that contain the employed data and tools. Projects, laboratories, and even whole institutions can organize their users and associated Narratives into a shared *Organization* with multiple permission levels and

management features. Developers can build, test, register, and deploy new or existing software as KBase apps using the Software Development Kit, thereby extending the platform's scientific capabilities.

Newly added services include several tools collaboratively developed by DOE SFA programs and KBase staff, including NWchem, GOTTCHA2, VIRSorter, and vConTACT2. KBase is unique in offering these diverse and integrated capabilities to a growing community of several thousand users. A central premise of KBase is to maximize the impact of data shared and developed between all users on system. This is the premise of KBase's emerging Knowledge Engine technology, which draws inferences between user data and public repositories, so scientists can better understand their work in the context of public knowledge. As data propagates across the system, it can be continually updated with new information like predictions of gene function, allowing the user to see how these changes scale from genes to ecologies and better predict outcomes.

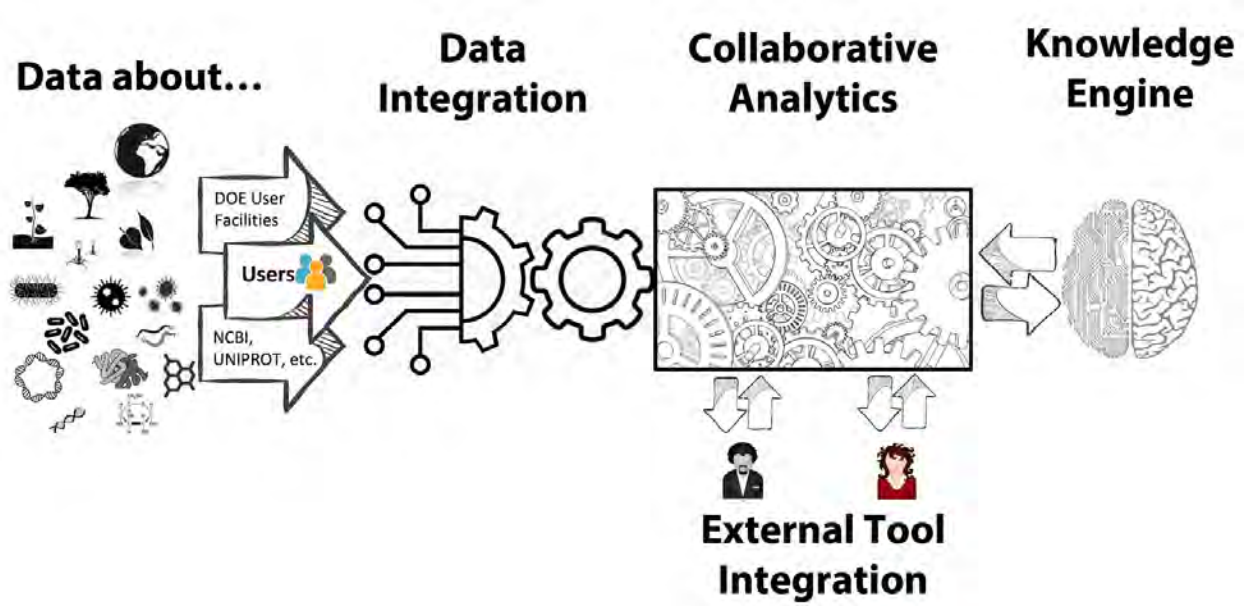


Figure 1. Integration of data and tools into KBase.

KBase is funded by the Genomic Science program within 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.

Large Scale Model-Driven comparison of Metagenome Assembled Genomes from Diverse Environments

Stephen Nayfach[1], José P. Faria[2*](jplfaria@anl.gov), Simon Roux[1], Rekha Seshadri[1], Daniel Udwary[1], Neha Varghese[1], Frederik Schulz1, Dongying Wu[1], David Paez-Espino[1], I-Min Chen[1], Marcel Huntemann[1], Krishna Palaniappan[1], Joshua Ladau[1], Supratim Mukherjee[1], T.B.K. Reddy[1], Torben Nielsen[1], Edward Kirton[1], Janaka N. Edirisinghe[2], Christopher S. Henry[2], Sean P. Jungbluth[3], Dylan Chivian[3], Paramvir Dehal[3], Elisha M. Wood-Charlson[3], Adam P. Arkin[3], Susannah Tringe[1], Axel Visel[1], IMG/M Data Consortium, Tanja Woyke[1], Nigel J. Mouncey[1], Natalia N. Ivanova[1], Nikos C. Kyrpides[1], Emiley A. Eloe-Fadrosh[1]

[1]DOE Joint Genome Institute, Berkeley, California, USA

[2]Argonne National Laboratory, Argonne, Illinois, USA

[3]Lawrence Berkeley National Laboratory, Berkeley, California, USA

https://genome.jgi.doe.gov/portal/GEMs/GEMs.download.html

https://narrative.kbase.us/#org/jgimags

Project goals:

Over 10,000 metagenomes collected from diverse habitats covering all of Earth's continents and oceans, human- and animal-host associated microbiomes, engineered environments, and natural and agricultural soils were used to generate over 52,000 metagenome-assembled genomes (MAGs). Metabolic models were constructed using the high-quality non-redundant MAGs to explore the distribution of metabolic functions across ecosystems. To evaluate the quality of the constructed metabolic models, species-level genomic references were compared and resulted in a high correlation of predicted functions, indicating the MAGs encode representative pathways to their isolate counterparts.

The reconstruction of bacterial and archaeal genomes from shotgun metagenomes provides insight into the ecology and evolution of environmental and host-associated microbiomes. Genome-scale metabolic models were built and analyzed for the non-redundant, high quality GEMs with >40 representatives per environment (n = 3255) using the ModelSEED[1] pipeline in KBase[2] (See supplemental materials). In brief, GEMs from similar environments, such as human and mammal, were shown to cluster by pathway presence (containing at least one complete flux pathway) implying that pathways are differentially sorted by distinct environmental factors. For the 607 GEMs with close (>95% ANI) RefSeq genomes identified, a comparison of GEM to RefSeq models revealed a very high (>0.98) correlation suggesting that these high-quality

GEMS are very near complete and representative of their full metabolic potential. To demonstrate that the high correlation was not the result of all models being similar, correlations were also computed for random pairs of GEM models and RefSeq models, resulting in a much lower correlation of 0.83. All GEMs, associated RefSeq genomes, and metabolic models are freely available in the "JGI MAG Database" KBase organization (https://narrative.kbase.us/#org/jgimags). To validate the high-quality GEMs metabolic models, pathway presence profiles were computed for reference genomes associated with humans and the built environment, as these two environments have >100 GEMs with associated reference genomes. The resulting profiles were nearly identical for all pathways. Pearson correlation coefficients were calculated for each GEM and corresponding reference genome across 55 metabolic pathways, with an average value >0.98. When the GEM and reference genomes were randomly paired and a Pearson correlation was calculated, the average correlation dropped to ~0.82, indicating that the high correlation previously reflects the similarity of the GEM and reference genome.

References

1. Arkin, Adam P., Robert W. Cottingham, Christopher S. Henry, Nomi L. Harris, Rick L. Stevens, Sergei Maslov, Paramvir Dehal, et al. 2018. "KBase: The United States Department of Energy Systems Biology Knowledgebase." *Nature Biotechnology* 36 (7): 566–69.
2. Henry, Christopher S., Matthew DeJongh, Aaron A. Best, Paul M. Frybarger, Ben Linsay, and Rick L. Stevens. 2010. "High-Throughput Generation, Optimization and Analysis of Genome-Scale Metabolic Models." *Nature Biotechnology* 28 (9): 977–82.

KBase is funded by the Genomic Science program within 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.

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.

NWChem Computational Modeling of Metabolites in KBase

Neeraj Kumar[1]* (Neeraj.Kumar@pnnl.gov), Andrew McNaughton[1], Rajendra Joshi[1], Eric Bylaska[1], Edoardo Apra[1], Lee Ann McCue[1], Gu Tianhao[2], Shane Canon[3], **Christopher Henry,[2] Adam P. Arkin[3], and Bob Cottingham[4]**

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Argonne National Laboratory, Argonne, IL; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Oak Ridge National Laboratory, Oak Ridge, TN

https://github.com/nkkchem/CompMolNWChem and https://github.com/nwchemgit/nwchem-dockerfiles

Project Goals: The DOE Systems Biology Knowledgebase (KBase) is a free, open-source software and data platform that enables researchers to collaboratively generate, test, compare, and share hypotheses about biological functions; analyze their own data along with public and collaborator data; and combine experimental evidence and conclusions to model plant and microbial physiology and community dynamics. KBase's analytical capabilities currently include (meta)genome assembly, annotation, comparative genomics, transcriptomics, and metabolic modeling. Its web-based user interface supports building, sharing, and publishing reproducible, annotated analysis workflows with integrated data. Additionally, KBase has a software development kit that enables the community to add functionality to the system.

Accurate prediction of quantum chemical properties of metabolites and their structure is needed to understand metabolism and to design new metabolic pathways for engineered systems. Such values can be precomputed for biochemistry databases like the ModelSEED resource that underlies KBase, but support is also needed for metabolites that do not yet appear in these databases (e.g. compounds proposed by cheminformatics applications or compounds involved in newly discovered synthetic pathways). Thus, it is not sufficient just to run property prediction apps in existing databases, but also to make these apps available for users to run themselves on new compounds of interest. To achieve this task, we recently deployed NWChem[1] quantum mechanical code in KBase to automate computational chemistry calculations and obtain optimized chemical structure with partial charges. This app accepts an arbitrary compound-set as input and enables users to select individual compounds from the input set to compute a structure for. The app adds mol2 structures to the input compound set for subsequent use in downstream applications like AutoDock Vina. Of course, it is also useful to precompute structures for all current compounds in the ModelSEED database. However, physics-based computational chemistry calculations are quite challenging to apply to a large database as they grow exponentially with the system size and require high-performance computing resources.[2] Thus, we are in the process of applying NWChem to compute high-quality predicted structures for as many compounds in the ModelSEED as possible, prioritizing the compounds that appear most prevalently in models in KBase. Thus far, structures have been precomputed for all 800

compounds in a *Yarrowia lipolytica* Yeast model. We are also currently exploring how to use HPC resources to power the NWChem app in KBase to greatly speed the pace of structure prediction in this app. In our talk, we will demonstrate this app on some example compounds and discuss the results of our analysis of the *Yarrowia* metabolites. Overall, this new integration in KBase empowers the synthetic biology community to test new compounds with known enzyme targets via docking calculations and paves the way for computation of a broad set of thermodynamic properties (e.g. free energy of formation of metabolites and free energy of metabolic reaction) that are useful for numerous systems biology applications.

References

1. M. Valiev, E.J. Bylaska, et al. "NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations" *Comput. Phys. Commun*. 181, 1477 (2010)
2. N. Kumar, et al. “Mechanistic Implications of Reductive Co–C Bond Cleavage in B_{12}-Dependent Methylmalonyl CoA Mutase” *J. Phys. Chem. B* 123, 10, 2210 (2019).

Funding statement. KBase is funded by the Genomic Science program within 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. N.K. was supported by the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. 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.

The 2019 KBase Fungal Biochemistry Curation Jamboree: Insights and Lessons Learned

Jeremy Zucker,[1]* (jeremy.zucker@pnnl.gov), Neeraj Kumar,[1] Janaka Eirisinghe,[3] Louise Glass[2], Kyle Pomraning, Scott Baker, Erin Bredeweg, and **Chris Henry[3]**

[1]Pacific Northwest National Laboratory, Richland, WA; [2]University of California, Berkeley, Berkeley, CA; [3]Argonne National Laboratory, Lemont, IL

https://kbase.us

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

Fungal genome-scale metabolic models are an efficient way of predicting phenotypes across various environmental conditions. These metabolic models are a key tool in understanding fungal-bacterial and plant-fungal community behavior.

However, automating the construction of high-quality fungal models has been a challenge. Recently, Kbase has introduced a fully automated fungal model construction pipeline. However, the underlying biochemistry data that is derived from published fungal models still needs to be significantly improved to build high-quality models.

To achieve this task, KBase organized a fungal biochemistry curation jamboree at PNNL. They reached out to the fungal community in the DOE space who were willing to contribute to curation of the fungal biochemistry. KBase has also developed an Escher map-based curation environment where curators can visualize biochemistry in the form of biochemical pathways and have the ability to curate reactions and assign gene families to those reactions. Based on feedback from the curation jamboree, KBase, has now set up standards in their curation process such as linking of literature citations and evidence-codes justifying the curation events.

Abstract

Recently, KBase has introduced a methodology to construct genome-scale fungal models in an automated fashion based on a set of reactions that are derived from 14 published fungal metabolic models. As the basis for the method, they produced a fungal model template that encompasses the biochemistry data from the published fungal models and the structural annotations from the associated fungal genomes.

KBase's approach uses structural annotations of any user-submitted fungal genome and computes a set of orthologous proteins against the curated fungal template in order to assert the presence or absence of specific biochemical reactions and pathways. These orthologous families are then curated and mapped to biochemistry by expert curators. The related biochemistry data is then propagated to construct a new draft metabolic model. Once the draft metabolic models are derived, additional reactions are added based on available functional annotations. This method is deployed in the Department of Energy Systems Biology Knowledgebase (KBase) (https://narrative.kbase.us/) as an app called "Build Fungal Model". This method is able to produce a draft fungal metabolic model in an around one hour.

KBase recognized that the core component of a pipeline for producing high-quality draft fungal models is a well curated biochemistry. However, the underlying biochemistry data that is derived from the published models needs to be significantly improved and reconciled onto a controlled

vocabulary in order to avoid redundancy and the incorrect representation of fungal biochemistry. To accomplish this, they reached out to fungal experts in DOE space and hosted a “Fungal Curation Jamboree” at Pacific Northwest National Laboratory”. Parallel to this effort they built a fungal curation environment based on Escher maps (https://escher.github.io/#/) where the biochemistry data and the gene associations can be visualized in the form of biochemical pathways. This allows the curators to quickly make the curation decisions on reactions and associate gene families to those reactions. This talk will focus on the outcomes of the Fungal Biochemistry Curation Jamboree and the lessons learned.

Genetically Encoded Biosensors for Mitochondrial and Cytosolic Biosynthesis of Branched-Chain Higher Alcohols in *Saccharomyces cerevisiae*

Yanfei Zhang,[1] Sarah Hammer,[1] Cesar Carrasco-Lopez,[1] Sergio Garcia, and **José L. Avalos** [1,2]* (javalos@princeton.edu)

[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 systems 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), and their blends belong to a select group of advanced biofuels with the highest potential to increase spark ignition engine efficiency [1]. Furthermore, BCHAs can be upgraded to jet fuel, making them excellent renewable fuels for ground as well as air transportation.

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. Isobutanol and isopentanol are derived from valine and leucine, respectively, which are biosynthesized in a common pathway that has been engineered in either the mitochondria or the cytosol. Efforts to commercialize these biofuels are challenged by limited productivities, as well as their high toxicity. While significant progress has been made in boosting yields and titers, particularly of isobutanol, there are currently no biosensors to enable high throughput screens to accelerate further stain improvement, or study their biosynthesis. In addition, very little is known about the interplay between different cellular networks and BCHA production and tolerance. Our recent study on the relationship between yeast tolerance to isobutanol and its adaptive response to nitrogen starvation begins to shed light into these complex phenotypes [2].

In this presentation I will provide a detailed description of the development, characterization, and application of the first genetically encoded biosensor for BCHA production in *Saccharomyces cerevisiae* [3]. This biosensor is based on the transcriptional regulator of branched chain amino acid biosynthesis, Leu3p. The activity of this transcription factor depends on the intracellular concentrations of α-iosopropylmalate (α-IPM), which is a byproduct and precursor of isobutanol and isopentanol biosynthesis respectively. Therefore, expressing green fluorescent protein (GFP) from a Leu3p-controlled promoter results in a robust biosensor for BCHA biosynthesis. Small modifications make the biosensor specific for either isobutanol or isopentanol production with correlation coefficients of $R^2 = 0.97$ and $R^2 = 0.99$, respectively. With these biosensors, we have been able to develop high throughput screens to isolate high BCHA producing strains, identify hyperactive variants of three enzymes in BCHA biosynthesis, and support the construction and optimization of whole metabolic pathways for isobutanol production in both mitochondria and cytosol.

These biosensors are essential to achieve the goals of the project. We are currently in the process of introducing the isobutanol biosensor into the yeast gene deletion library; this has already enabled the discovery of new genes that, when deleted, enhance isobutanol production. In addition, we are conducting the initial characterization of strains containing both the isobutanol biosensor and optogenetic controls of isobutanol biosynthesis [4], with the objective of developing a closed-loop control system for BCHA biosynthesis. These advances are significant milestones towards our ultimate goal of increasing our basic understanding of BCHA biosynthesis and tolerance, as well as expanding our capabilities to control and improve strains for their production.

References

1. Farrell, John, John Holladay, and Robert Wagner. "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. Kuroda K[#], Hammer SK[#], Watanabe Y, Montaño-López JJ, Fink GR, Stephanopoulos G, Ueda M, Avalos JL. Critical Roles of the Pentose Phosphate Pathway and GLN3 in Isobutanol-Specific Tolerance in Yeast. *Cell Systems* 9; 534-547 (2019).
3. Zhang Y, Hammer SK, Carrasco-López C, García-Echauri S, Avalos JL. A genetically encoded biosensor for mitochondrial and cytosolic biosynthesis of branched-chain higher alcohols in *Saccharomyces cerevisiae*. (In preparation).
4. Zhao EM, Zhang Y, Mehl J, Park H, Toettcher JE, Avalos JL. Optogenetic regulation of engineered cellular metabolism for microbial chemical production. *Nature* 29; 555 (7698):683-87 (2018).

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

Rapid automated curation of genomes from metagenomes and integration into KBase

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; [2]University of São Paulo, São Paulo, [3]Lawrence Berkeley National Laboratory, Berkeley; [4]Lawrence National Livermore National Laboratory, Livermore

http://genomicscience.energy.gov/research/sfas/llnlsoil.shtml
http://kbase.us

Project Goals:

The large scale recovery of metagenome assembled genomes (MAGs) has often resulted in the placement of the genome as the operational unit through which microbial biology and systems are understood. The underlying assembled sequences of these mass-produced genomes have characteristic errors from the assembly process that can result in MAGs that do not represent a true biological organism. Ultimately, erroneous genomes limit their utility for making robust ecological, evolutionary, and energetic inferences about microbial systems. Assembly errors can be fixed, but repair necessitates time consuming manual human-guided curation that ultimately restricts the number of curation targets. To address this, we are developing FixAME, an automated curation system that is being integrated into KBase. FixAME rapidly identifies and repairs many common assembly errors across entire metagenomic assemblies and can substantially increase the accuracy and completion of MAGs. FixAME also improves the overall contiguity of metagenomic assemblies while simultaneously sidestepping the bottleneck of costly human-guided manual curation.

Funding statement.
Funding for this project was provided by the Department of Energy - Microbes Persist: Microbes Persist: Building a KBase Foundation for Microbial and Viral Ecogenomics in Soil" under the Lawrence Livermore National Laboratory Soil Microbiome Science Focus Area.

Optimizing Tradeoffs Implicit During Bioenergy Crop Improvement: Understanding the Effect of Altered Cell Wall and Sugar Content on Sorghum-associated Pathogenic Bacteria

Qi Wang,[1] David Braun,[2] Anna Casto,[1] Singha Dhungana,[2] Noah Fahlgren,[1] Malia Gehan,[1] Jose Tovar,[1] Wilfred Vermerris,[3] and **Rebecca Bart**[1,*] (rbart@danforthcenter.org)

[1]Donald Danforth Plant Science Center, St. Louis, MO; [2] University of Missouri, Columbia; [3] University of Florida, Gainesville

Project Goals: The first project goal is to quantitatively model the disease triangle that describes sorghum, pathogenic bacteria, and the environment. Field and laboratory experiments are being combined to determine bacterial susceptibility of genetically diverse sorghum genotypes that differ in cell wall and sugar composition. Standard plant pathology techniques combined with powerful phenomics approaches are providing a holistic view of this pathosystem within variable environments. Further, transcriptomics is being employed to elucidate mechanisms used by bacterial pathogens to induce sorghum susceptibility. Microbial pathogens are known to manipulate the sugar and cell wall characteristics of their hosts. Consequently, these characteristics are being analyzed during pathogen invasion. This research will reveal the mechanisms underlying resistance to pathogens that must be maintained during biofuel trait optimization.

Plant-derived production of renewable fuels and chemicals has the potential to enhance US farming and agricultural economic opportunities, increase domestic energy security, and reduce fossil fuel dependency and greenhouse gas emissions. Realizing the potential of these alternative energy sources necessitates the development of high-biomass-yielding crops. These specialized crop varieties may harbor modifications to cell walls, which are a major barrier to pathogen entry, and to the tissue distribution of sugars, which are the food source for the pathogen; hence they are likely to present previously unseen challenges for disease resistance. Over the last several years, disease from the bacterial pathogen *Xanthomonas*, has caused significant yield losses in many crops where bacterial diseases had historically been rare, including corn and cotton. It is currently unclear why these diseases are emerging. *Xanthomonas* is a known pathogen of sorghum (*Sorghum bicolor* (L.) Moench), though similar to corn and cotton, the incidence and impact of the disease has historically been low. Taken together, these observations highlight a vulnerability in sorghum resilience to pathogens that is likely to be magnified by alterations in cell wall and sugar content. In this project, we aim to establish the sorghum – *Xanthomonas* pathosystem as a model for deducing how latent microbial pathogens might exploit key biofuel crop traits.

We will report on a screen of over one hundred and fifty sorghum varieties for resistance/susceptibility to *Xanthomonas* and an accompanying GWAS analysis. We will also summarize the results from a dual-RNA sequence experiment to simultaneously reveal gene expression patterns in both host and pathogen during compatible and incompatible interactions and characterization of cell wall and sugar profiles during pathogen attack.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018072.

New Computational Pipelines to Prioritize Candidate Genes for Optimal Biomass Production under Drought in C4 Plants *Setaria viridis* and *Sorghum bicolor*

Cheng Zhao[1], Fan Lin[1], Elena Lazarus[1], Pascal Schläpfer[1], Hye-In Nam[1], Scott Lee[3], Phil Ozersky[3], Edward J. Wolfrum[2], Jennifer Barrett[3], Allen Hubbard[3], Hui Jiang[3], Xiaoping Li[3], Erica Agnew[3], Todd Mockler[3], **Ivan Baxter[3]**, Seung Y. Rhee[1,*] (srhee@carnegiescience.edu)

[1]Carnegie Institution for Science, Stanford, CA; [2]National Renewable Energy Lab, Golden, CO; [3]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

C4 plants, such as *Sorghum bicolor* and *Setaria viridis*, have CO_2 concentrating mechanisms in specialized cell types (bundle sheath and mesophyll cells) to enhance water use and photosynthetic efficiencies. Current mathematical modeling of C4 photosynthesis does not sufficiently capture leaf biochemical and anatomical phenotypes under dynamic environments. Linkage mapping has been widely used to identify quantitative trait loci (QTL) in many plant species but usually requires a time-consuming and labor-intensive fine-mapping process. Here, we developed two computational pipelines to identify candidate genes to improve important agricultural traits, such as height and biomass production. First, we developed QTG-Finder2, a machine learning-based algorithm to prioritize the causal genes in QTLs, and used orthologs of known causal genes as a training set. The model trained with orthologs could recall about 64% of Arabidopsis and 83% of rice causal genes when the top 20% ranked genes were considered, which is similar to the performance of models trained with known causal genes. Using QTG-Finder2, we trained and cross-validated *Sorghum bicolor* and *Setaria viridis* models. The sorghum model was validated by causal genes curated from the literature and could recall 70% of causal genes when the top 20% ranked genes were considered. We also applied the *Setaria* model and public transcriptome data to prioritize a plant height QTL and identified thirteen candidate genes. Second, we will present a computational framework of multiscale 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 *S. bicolor* 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 data to inform the metabolic network model, such as photosynthesis data, biomass composition data, and RNA-seq data for Sorghum under well-watered and water-limiting conditions at multiple time points. Using the computational framework, we simulated plant growth and predicted that knocking out 23 genes and overexpressing 28 genes can improve biomass production. Finally, we have developed a pipeline to mine pan-genome information to identify potential causative polymorphisms in linkage with GWAS identified loci. We will cross-validate these independent pipelines computationally and test the prioritized candidate genes experimentally by generating and phenotyping knock-out lines.

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 nos. DE- SC0018277 and DE-SC0008769.*

PEPC kinetics and the efficiency of C_4 photosynthesis in *Sorghum bicolor*

Asaph B. Cousins[1]* (acousins@wsu.edu), Ryan L. Wessendorf[1], Robert J. DiMario[1], Kuenzang Om[1], and **Ivan Baxter**[2]

[1]Washington State University, Pullman, WA 99163, [2]Donald Danforth Plant Science Center, St. Louis, MO 63132

url: http://foxmillet.org

Overall 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. Here we specifically focus on Objective #1*: Engineering photosynthesis to improve performance under water stress.*

Abstract: Due to the predicted increase in food demand, studying the biochemical components of C_4 photosynthesis may provide insight into enhancing photosynthesis in crop plants to increase yield. Currently, photosynthesis can be reduced in C_4 crops by drought conditions which reduce intercellular CO_2 concentrations (C_i) in the plant. The initial carboxylation reaction in C_4 plants is catalyzed by phospho*enol*pyruvate carboxylase (PEPC) and leads to elevated CO_2 around Rubisco. The C_4 isozyme of PEPC originated from a non-photosynthetic PEPC and it has been suggested that specific amino acid substitutions in PEPC confer differences in the affinity of the enzyme for PEP (K_{PEP}). These changes in K_{PEP} may be an unavoidable side effect of selecting for a higher affinity for HCO_3^- (K_{HCO3}) to maintain rates of PEPC when stomatal conductance (g_s) is low. However, experimental evidence for amino acid changes influencing *in planta* kinetic properties of PEPC and rates of C_4 photosynthesis is lacking. Therefore, the objective of this aim is to determine how specific amino acid differences between the C_3 and C_4 isozymes of PEPC influence the efficiency of C_4 photosynthesis when the availability of atmospheric CO_2 is low. To accomplish this objective, we are measuring the kinetic properties of 28 PEPC isozymes from both C_3 and C_4 plants from members of the Poaceae family. These enzymes were overexpressed and purified from the PEPC-less *PCR1 Escherichia coli* strain. The kinetic measurements have been compared to determine if there is a tradeoff between K_{PEP} and K_{HCO3}. These PEPC kinetic parameters were measured in a temperature-controlled cuvette linked to a mass spectrometer. The ultimate goal of this research is to introduce an enhanced PEPC enzyme into sorghum to increase photosynthesis under drought conditions. The outcome of this research will enhance C_4 photosynthetic efficiency and will lead to an increase in whole plant water use efficiency.

Funding statement: This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0008769).

Phenomics of stomata and water use efficiency in C_4 species

Andrew D.B. Leakey[1]* (leakey@illinois.edu), John Ferguson[1], Jiayang Xie[1], Charles Pignon[1], Niteen Kadam[1], Ashish Rajukar[1], Dustin-Mayfield Jones[1], Maximilian Feldman[2], Parthiban Prakash[1], Darshi Banan[1], Rachel Paul[1], Scott Lee[2], Greg Ziegler[2], **Ivan Baxter[2]**, Thomas E. Clemente[3]

[1]University of Illinois at Urbana-Champaign, Urbana, IL; [2]Donald Danforth Plant Science Center, St Louis, MO; [3]University of Nebraska, Lincoln, NE

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

Water use efficiency (WUE), which is physiologically distinct from drought tolerance, is a key target for improving crop productivity, resilience and sustainability. This is because water availability is the primary limitation to crop yield globally and irrigation uses the largest fraction of our limited freshwater supply. The exchange of water and CO2 between a leaf and the atmosphere is regulated by the aperture and pattern of stomata. Mechanistic modeling indicates that stomatal conductance could be reduced or stomatal movements accelerated to improve water use efficiency in important C4 crops. While molecular genetics has revealed much about the genes regulating stomatal patterning and kinetics in Arabidopsis, knowledge of the genetic and physiological control of WUE by stomatal traits in C4 crops is still poor. Understanding of natural diversity in stomatal traits is limited by the lack of high-throughput phenotyping methods. In response, a rapid method to assess stomatal patterning in model C4 species grown in the field has been implemented. The leaf surface is scanned in less than two minutes with an optical tomographer, generating a quantitative measurement of a patch of the leaf surface. An algorithm was designed to automatically detect stomata in 10,000s of these images via training of a neural network approach. We identified trait correlations as well as genotype to phenotype associations for stomatal patterning, leaf gas exchange and canopy water use through quantitative trait loci and genome wide association studies. Transgenically modified expression of stomatal patterning genes has produced sorghum with greater WUE. Plants were grown in a new field facility for comprehensive evaluation of leaf, root and canopy WUE traits under Midwest growing conditions.

Funding statement: This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0008769).

Targeted Mutagenesis and Programmed Transcriptional Regulation in Setaria and Sorghum

Yang Liu,[1] Arjun Khakhar,[1] Matt Zinselmeier, [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 project is funded by grant DE-SC0018277 from The DOE Department of Biological and Environmental Research

The Centrality of the Development of Transgenic Lines for the Analysis of Photosynthetic and Water Use Efficiencies in Sorghum

Albert Kausch[1], Kimberly Nelson-Vasilchik[1], Joel Hague[1], Jiayin Liu[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].**

[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, 63130

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: 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. A required but often underappreciated technology for this goal involves the capability to create, test and cultivate transgenic and genome edited plants. The need and centrality of transgenic lines in sorghum is technically difficult yet essential to the analysis of traits genes. Standard plant transformation protocols are often limited by various factors including: (1) tissue culture expertise and facility intensive resources; (2) genotype and explant dependence, (3) low efficiencies; and, (4) time and labor intensive efforts. We have established reliable protocols for the standard *Agrobacterium*-mediated introduction for stable genetic constructs into sorghum cv BTx430. Using these protocols, we have established lines of transgenic sorghum to phenotypically study various genes of interest. Stable lines of sorghum 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 have explored current applications using transient expression of *Baby boom* and *Wuschel2* delivered by microprojectile bombardment in sorghum. Our program supports 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-SC0008769).

Using synthetic genetic circuits to tightly control root architecture.

Jennifer Brophy[1], Katie Magallon[1], **José R. Dinneny**[1*] (dinneny@stanford.edu), Ivan Baxter[2]

[1]Stanford University, Stanford, CA; [2]Donald Danforth Plant Science Center

[2]Donald Danforth Plant Science Center, St. Louis, MO[1]

Project Goals:

Establish a functionally characterized parts list of gene regulatory parts to construct orthogonal circuits in plants.

Build a new robotics system for the reliable automated imaging of root systems in plants

Previous work has demonstrated that the development of crown roots increases the flux of water through the plant, while the density of stomata and their aperature determine the rate water evaporates from the leaf surface. By modulating the growth of crown roots and stomatal patterning, we will change the rate at which water is taken up and lost to the atmosphere by the plant. This work will test the hypothesis that plants tend to exhibit responses that are more conservative than necessary due to their origins as wild species. Fine-scale control of developmental pathways affecting root and shoot development will take advantage of synthtetic biology approaches that enable reconstruction of regulatory circuits in plants with specific design specifications.

Funding statement: This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0008769).

Metabolic Flux Analysis of Sucrose-Secreting Cyanobacterium *Synechococcus elongatus*

Bo Wang (bo.wang.2@vanderbilt.edu) [1*], Cristal Zuniga (crzuniga@ucsd.edu) [2*], **Karsten Zengler** [2-4], **Jamey D. Young** [1, 5], **Michael J. Betenbaugh** [6]

[1] Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN. [2] Department of Pediatrics, University of California San Diego, La Jolla, CA. [3] Department of Bioengineering, University of California San Diego, La Jolla, CA. [4] Center for Microbiome Innovation, University of California San Diego, La Jolla, CA. [5] Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN. [6] Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD.

The goal of this project is to combine autotrophs and heterotrophs as a novel synergistic and symbiotic platform for the production of sustainable biofuel precursors. Photosynthetic microorganisms fix sunlight and CO_2 and provide organic carbon source and oxygen to the heterotrophs that are prolific producers of complex metabolites. Synthetic microbial communities of cyanobacterium-fungus will be studied through genome-scale metabolic modeling and ^{13}C metabolic flux analysis. Our current focus is to evaluate the metabolic response of the cyanobacterium *Synechococcus elongatus* PCC 7942 to osmotic stress and sucrose secretion.

Cyanobacterial strains that are capable of secreting sucrose to support growth of heterotrophs in a co-culture system have gained significant interest from the biotechnology community. Efforts on strain development and process optimization have taken place since a decade ago, and the technology has been advanced significantly. However, most efforts have been focused on investigating local pathways that are closely related to sucrose biosynthesis and secretion, and the global intracellular metabolism, which is crucial for detecting bottlenecks in the network, has not yet been investigated.

In this study, we investigate the global metabolic fluxes in the sucrose-secreting ($cscB^+$) strain versus the wild type. *Synechococcus elongatus* PCC 7942 synthesizes sucrose under NaCl stress as a means of coping with osmotic pressure, while overexpression of CscB, a sucrose transporter, confers the secretion of sucrose out of cells. We use two complementary approaches, *i.e.*, ^{13}C metabolic flux analysis (MFA) and Genome-Scale Modeling (GSM), to elucidate the difference of intracellular resource allocation by quantifying metabolic fluxes between these strains. ^{13}C-MFA uses ^{13}C-labelled compound as tracer to displace the ^{12}C atom of intracellular metabolites, and the metabolic fluxes can be quantitatively mapped through tracking the labelling status of these metabolites along a time course and computational modeling. Constrain-based GSM combine genome-scale network reconstructions with detailed biochemical and physiological information to provide *in silico* estimation of intracellular metabolic flux under a condition-specific set of constraints. ^{13}C-MFA and GSM can be applied to decipher network regulations and hence guide metabolic engineering.

We performed ^{13}C-MFA and GSM for three cyanobacterial cultures – [wild type], [wild type + NaCl], and [$cscB^+$ + NaCl] – under photoautotrophic conditions. For ^{13}C-MFA, we use ^{13}C-

labelled bicarbonate as tracer, and use GC-MS and LC-MS/MS to quantify labeling trajectories of more than a dozen of intracellular metabolites along a time course. We use INCA software to simulate the flux and generate the quantitative flux map. For GSM, we evaluated the solutions space using random sampling, using experimentally observed sucrose secretion rates as constrains in the model of the cyanobacteria to validate growth phenotypes and resource allocation under various growth conditions. This unbiased assessment provides a distribution of all possible flux distributions of the network at given conditions. Additionally, we evaluated osmotic stress conditions by increasing the salts concentration five times from 28 mM to 130 mM. Predicted flux distributions across the solution space were compared for all three conditions, showing significant changes in transport and carbohydrate metabolism.

This project is supported by Biological and Environmental Research, Department of Energy, through grant DE-SC0019388: Creating Multifunctional Synthetic Lichen Platforms for Sustainable Biosynthesis of Biofuel Precursors.

Optimizing Carbon Metabolism in Co-Culture for Applications to Sustainable Biosynthesis

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

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

Project Goals: The goal of this project is to develop synthetic lichen communities of autotrophic and heterotrophic microbes as a novel sustainable symbiotic platform for the production of biofuel and its precursors. Carbon-fixing autotrophs provide oxygen and organic substrates to their heterotrophic neighbors, which in turn produce carbon dioxide. By optimizing and enhancing these interactions, we can create a robust, sustainable synthetic lichen community. Multi-omics driven genetic engineering will improve metabolite exchange and product generation capabilities with the microbial co-culture.

Lichens are communities of auto- and heterotrophic microbes that collect sunlight and carbon dioxide and apply it to power the group's activities. They also 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. Key metabolite excretion bottlenecks will be identified in each partner, then the organisms will be modified as appropriate in order to share particular metabolic intermediates with their heterotrophic partners for channeling into key metabolic pathways, thus generating energy-related precursors of biochemicals or biofuels with high commercial value. To achieve these goals, we have screened several photoautotrophs and heterotrophs to find mutualistic coculture partners. One such photoautotroph, *Picochlorum renovo*, exhibits a rapid growth rate and is tolerant to a wider range of temperature and salinity than many other microalga. Genetic engineering toolboxes have been created to improve coculture applications with this organism. Additionally, we have engineered overexpression of the invertase enzyme in *Yarrowia lipolytica*, allowing for the consumption of sucrose and other less-utilized organic carbon sources. Other coculture partnerships with heterotrophic fungi and yeast (including *R. glutinis* and filamentous *Aspergillus* fungi) have been successfully conducted with cyanobacteria including *Nostoc* and engineered sucrose-secreting *S. elongatus*.

We have further explored the growth of *S. elongatus/R. glutinis* cocultures in continuous conditions using a custom-built, feedback-controlled photobioreactor equipped with a radial LED-based irradiation system. Since the organisms are cultured together, it is challenging to assign the origin of most essential metabolites to a particular specie in both the extracellular and

intracellular domains. To address this issue, we tested a new co-culture incubation strategy through incubation of cells using membrane segregation, hydrogel matrix immobilization, and other segregation techniques.. We also developed an analytical pipeline allowing investigators to identify extracellular and intracellular metabolites, estimate their fluxes and to assign them to the originating species.

References

1. Dahlin, L.R., Gerritsen, A.T., Henard, C.A. *et al.* Development of a high-productivity, halophilic, thermotolerant microalga *Picochlorum renovo*. *Commun Biol* 2, 388 (2019). https://doi.org/10.1038/s42003-019-0620-2

Supported under award DE-SC0019388, Biological and Environmental Research

Title: Ectomycorrhizal fungi: mediators of plant-microbial interactions and terrestrial biogeochemistry

Jennifer M. Bhatnagar[1]* (jmbhat@bu.edu), Colin Averill,[2] Hui-Ling Liao,[3] Ko-Hsuan Chen,[3] Edward Brzostek,[4] and Rytas Vilgalys[2]

[1]Boston University, Boston, MA; [2]ETH Zurich, Switzerland; [3]University of Florida, Gainesville, FL; [3]University of West Virginia, Morgantown, WV; and [5]Duke University, Durham, NC

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

Interactions between plants and soil microorganisms can reshape projections of terrestrial biogeochemical cycles, yet the mechanisms by which these interactions operate are poorly understood. **The goal of this project is to determine the role of plant and soil resources in shaping interactions between coniferous plants, their major root fungal symbionts (ectomycorrhizal fungi, EMF), and free-living saprotrophic decomposers (SAPs) in soil that control forest biogeochemistry.** Our project leverages a model plant-EMF system that dominates coniferous forests in U.S. and abroad; plants in the Pinaceae and their EMF symbionts in the genus *Suillus*. We have discovered that plant-EMF-SAP interactions in this system are driven by soil carbon (C) and nitrogen availability to SAPs. Ongoing experiments now focus on how increases in plant C availability to EMF (caused by elevated CO_2) affect plant-microbial interactions and biogeochemistry. Future work will integrate experimental data into a microbial-explicit biogeochemistry model, FUN-CORPSE and validate the model with data on soil microbial community and biogeochemistry from the Duke FACE experiment.

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 contract numbers JGI/EMSL FICUS 48480, 49514, and 49989; JGI CSP #503285; and Award Numbers DE-SC0012704 and DE-SC0020403.

Developing a molecular-level model of cofactor-trafficking in chloroplasts

Crysten E. Blaby-Haas*[1] (cblaby@bnl.gov), Desigan Kumaran[1], Miriam Pasquini[1], Qizhi Zhang[2], Sam Seaver[2]

[1]Brookhaven National Laboratory, Upton, NY; [2]Argonne National Laboratory, Argonne, IL.

Project Goals: Bioenergy crops that thrive in marginal soils and maintain performance in diverse and fluctuating environments are an essential component of a sustainable energy and carbon portfolio. However, understanding and predicting productivity in these environments is challenging, in part, because of the general lack of sequence-to-function information in the plant lineage. The Quantitative Plant Science Initiative (QPSI) is a versatile and scalable 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 gene-, protein-, and molecular-level experimentation. In this way, we leverage the scalability of 'omics data and bioinformatic approaches to capture system-level information for DOE-relevant crops, while generating experimentally determined sequence-specific understanding of gene and protein function. By incorporating molecular-level experimentation into our workflow, we can address the question of how a protein functions and establish mechanistic insight into how sequence variation impacts phenotype. This knowledge also 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.

To ensure that sustainable, low-impact bioenergy can be developed, a fundamental understanding of how plants can acclimate to poor nutrient quality is needed. Micronutrients are of growing importance in maximizing bioenergy/bioproduction crop yield, and the trace metal nutrients present unique challenges. Bioavailability of these nutrients in the soil is dynamic and variable, and yield-impacting deficiency can suddenly appear as more intensive cropping, higher yielding crops, insufficient nutrient management, and nutrient imbalances are becoming more widespread. Additionally, because these elements are essential for the proper assimilation and metabolism of macronutrients, such as nitrogen, poor macronutrient availability is exacerbated by metal deficiency. To support the development of bioenergy crops with improved micronutrient stress resilience, our goal is to develop a genome-based, molecular-level and system-level understanding for the two most abundant trace metal nutrients in plants: zinc and iron.

In addition to the thylakoid membrane, many metal-dependent proteins localize to the chloroplast. They do not necessarily participate in photosynthetic electron transfer directly but serve a support role ensuring the proper functioning of photosynthesis. The chloroplast is also the site of multiple metabolic pathways that are dependent on a transition metal at one or more steps. To ensure metalloprotein biogenesis in the chloroplast, plants have unique challenges to overcome. Cofactors must be transported across multiple membranes, while avoiding potentially cytotoxic interactions.

While permeases involved in chloroplast metal transport are relatively well characterized, how metal ions are loaded into most metalloproteins is unknown. Since unchelated metal ions can be toxic, mechanisms that mediate trafficking from the site of transport to each metalloprotein should exist. Here, we present the discovery of (1) a chloroplast metal transferase, which we propose is a zinc-chaperone essential for proper translation and protein modification during zinc scarcity, and (2) an unprecedented trinuclear zinc-heme-zinc protein, which we propose may function as a heme chaperone in cyanobacteria, but with the addition of an oxidase domain, has evolved a function in heme degradation in plants. We also present our work toward developing a plant subsystem editor. This public platform will enable the *in silico* integration of functional annotations, comparative plant genomics and metabolic reconstruction, which we are using to build process-level models of how plants mediate cofactor trafficking and metabolism-centric adjustments to cofactor use in the chloroplast.

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

Understanding poplar and sorghum micronutrient stress by integrating functional genomics with molecular-level experimentation

Meng Xie*[1] (xiem@ornl.gov), Sam Seaver[2], Daifeng Wang[3], Doreen Ware[4,5], Qun Liu[1], Timothy Paape[1], **Crysten E. Blaby-Haas[1]**

[1]Brookhaven National Laboratory, Upton, NY; [2]Argonne National Laboratory, Argonne, IL; [3]University of Wisconsin, Madison, WI; [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; [5]US Department of Agriculture, Agricultural Research Service, Ithaca, NY, USA

Project Goals: Bioenergy crops that thrive in marginal soils and maintain performance in diverse and fluctuating environments are an essential component of a sustainable energy and carbon portfolio. However, understanding and predicting productivity in these environments is challenging, in part, because of the general lack of sequence-to-function information in the plant lineage. The Quantitative Plant Science Initiative (QPSI) is a versatile and scalable 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 gene-, protein-, and molecular-level experimentation. In this way, we leverage the scalability of 'omics data and bioinformatic approaches to capture system-level information for DOE-relevant crops, while generating experimentally determined sequence-specific understanding of gene and protein function. By incorporating molecular-level experimentation in our workflow, we can address the question of how a protein functions and establish mechanistic insight into how sequence variation impacts phenotype. This knowledge also 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.

The development of 'omics technologies provides an unprecedented opportunity for the study of plants as complex biological systems. With these technologies and the associated genome-wide data, nearly any plant species, including DOE-relevant crops such as poplar and sorghum, can be fast-tracked to a level of understanding that was previously attainable for only a few "model organisms". The challenge is shifting from acquiring genomic data, such as whole-genome sequences and transcriptomics datasets, to using that knowledge to enable predictive biology. Central to this challenge is the paucity of sequence-to-function understanding in the plant lineage, generally, and in bioenergy crops, specifically. While our aspiration is a large-scale capability that generates gene and protein functional understanding in bioenergy crops, our initial focus addresses the need for micronutrient stress resilience for sustainable bioenergy/bioproduction. Of the essential micronutrients, insufficient zinc and iron bioavailability causes widespread decreases in crop yield. We will broadly use resources developed in the DOE complex together with a multi-omics experimental strategy to discover adaptive responses to suboptimal zinc and iron availability in two DOE flagship bioenergy crops, poplar and sorghum. This data will be used for a computational simulation of cofactor availability in the chloroplast, the major metal sink and site

of carbon fixation and energy generation. At the same time, we will deploy an interdisciplinary approach and use high-throughput methodologies to provide a layer of experimentally grounded sequence-specific understanding of molecular-level functions for major players involved in chloroplast metal trafficking. These studies will serve to improve our model of micronutrient dynamics in bioenergy crops and to provide a foundation for establishing how metal scarcity and excess affect the biomass yield of bioenergy crops.

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

Molecular Regulation of Cell-type Specific Responses to Abiotic Stresses in Poplar

Amir H. Ahkami [1] (amir.ahkami@pnnl.gov), Vimal K. Balasubramanian [1], Maria Del Mar Rubio Wilhelmi [4], Tanya Winkler [1], Samuel Purvine [1], Ying Zhu [1], Sarah Williams [1], Lye Meng Markillie [1], Hugh Mitchell [1], Dusan Velickovic [1], Yongil Yang [2], C. Neal Stewart, Jr.[2], Jonathan Cumming [3], Stephen DiFazio [3] and **Eduardo Blumwald [4]*** (eblumwald@ucdavis.edu)

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

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.

Plants react to abiotic stress with a combination of physiological, biochemical, and developmental changes. These responses include alterations in signaling components, gene transcription, synthesis of proteins and metabolites which occur in a cell-type and tissue-specific manner [1]. 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 an understanding of the cell/molecular properties of specific single cell-types within a tissue.

In this work, 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 water deficit, salinity, heat, and the combination of all three stresses were monitored. Leaf and root tissues were collected at different time points, fixed and embedded for cell-type specific omics analyses. We targeted distinct poplar cell types and tissues including leaf mesophyll, xylem/phloem, root epidermis and cortex cells using cryo-sectioning and laser-capture microdissection (LCM) techniques. Samples from same plants were also collected for whole tissue omics (RNA-seq, metabolite) analysis.

To decode the structural and regulatory gene networks that mediate spatiotemporal specialization, we employed a cell-type specific gene expression profiling approach. To this end, RNA was extracted from 400-500 cells per cell-type. Full length cDNA and template libraries were generated, quantified and sequenced. Our data revealed that under all investigated water-deficit stress conditions including Early Water-Stress (EWS), Late Water-Stress (LWS), and Recovery (R), 585 and 138 transcripts were significantly regulated in leaf palisade and vascular cell types, respectively. The observed transcriptional changes were further assessed at the protein translation level. For proteomics analysis, total protein was extracted from 300-800 cells per cell-type, and samples were processed within ultrasmall-volume “nanowells” (nanoPOTS technology)

[3]. In response to water-deficit stress and water recovery conditions, a total of 2,384 and 3,490 proteins were identified as leaf palisade mesophyll and vascular-specific proteins, respectively. Among these, 498 (in palisade cells) and 270 (in vascular cells), 184 (in palisade cells) and 336 (in vascular cells), and 392 (in palisade cells) and 428 (in vascular cells) proteins were exclusively identified at EWS, LWS, and R phases, respectively, as unique candidate cell-type specific drought-responsive proteins. Our pathway enrichment analysis revealed a number of proteins involved in amino acid and carbohydrate metabolism, photosynthesis, sucrose biosynthesis/degradation and antioxidant metabolism as responsive proteins exclusively in palisade mesophyll or vascular cells in water-deficit stress and post-recovery phases. These findings are being correlated with whole tissue metabolite profiles and linked to the spatiotemporal accumulation of selected metabolites using a MALDI-MS imaging approach[4]. These results will be presented and discussed.

Overall, our results indicate that specific cell-types in poplar tissues are defined by distinctive transcriptional and protein profiles under abiotic stress. This information is being used for motif discovery, the results of which will be used for engineering of cell-, tissue- and stress-specific promoters. Subsequent poplar transformation with constructs of stress-mitigating genes driven by these promoters will be performed. These efforts will contribute to the design of stress-tolerant poplar trees, a strategic bioenergy crop, and to a better understanding of the roles of different poplar cell types on the response to abiotic stress.

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. Velickovic D., R.K. Chu, G.L. Myers, A. Ahkami, and C.R. Anderton. An Approach for Visualizing the Spatial Metabolome of an Entire Plant Root System Inspired by the Swiss-rolling Technique. (2019). *Journal of Mass Spectrometry*. doi:10.1002/jms.4363.

The SyPro Poplar project is supported by the U. S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), Genomic Science Program.

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

Yongil Yang[1,2], Jun Hyung Lee[1,2], Yuanhua Shao[1,2], Magen R. Poindexter[1,2], Cristiano Piasecki[2], Amir H. Ahkami[3], Stephen DiFazio[4], Eduardo Blumwald[5], **C. Neal Stewart, Jr**.[1,2 *]

[1]*Center for Agricultural Synthetic Biology, University of Tennessee Institute of Agriculture, Knoxville, Tennessee, USA*
[2]*Department of Plant Sciences, University of Tennessee, Knoxville, TN*
[3]*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA*
[4]*Department of Biology, West Virginia University, Morgantown, WV*
[5]*Department of Plant Sciences, University of California, Davis, CA*

*nealstewart@utk.edu

Crops are frequently subjected to abiotic stresses including water deficit, salinity, and heat. Here we show the performance of a set of rationally-designed osmotic stress-inducible synthetic promoter in hybrid poplar, a strategic bioenergy crop. We mined poplar transcriptome data for signature promoter motifs that are putatively drought- or salt-responsive. *De novo* motif-detecting algorithms yielded 30 water stress- and 34 salt stress candidate DNA motifs using *E*-value threshold of 0.001 from the respective promoters of drought- and salt-responsive co-expressed genes from poplar data sets. In drought responsive motifs, a novel domain was comprised of 3 to 9 conserved motifs (SD1-9) found in 16 co-expressed gene promoters. A newly-identified well-conserved motif (SS16) for salt-response was discovered. Fifteen synthetic promoters using mined sequence were fused to a green fluorescent protein (GFP) gene. These promoters were screened by transient expression assays using poplar leaf mesophyll protoplasts and agroinfiltrated *Nicotiana benthamiana* leaves under osmotic stress condition. Twelve of these synthetic promoters induced GFP expression in both transient expression systems. Especially, two SD (SD3-1 and 6-2) and three SS (SS16-1, 16-2, and 16-3) synthetic promoters responded significantly to low water content and high salinity, respectively, in agroinfiltration test. These five synthetic promoters were then selected for generating stable transgenic *Arabidopsis* to validate the inducibility in plants. SD3-1 and 6-2 responded to water deficiency, while SS16-1, SS16-2, and SS16-3 responded to high salinity in *Arabidopsis*. Orthogonal synthetic promoters for multiple crops is a 'grail' of biotechnology. The current results appear to provide multiple expression profiles and choices to deploy osmotic stress-inducible promoters that may be effective in multiple plant species. The design-build-test strategy appears fruitful for tuning abiotic-stress tolerance responses in the next steps of the research in engineered poplar trees.

The SyPro Poplar project is supported by the U. S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), Genomic Science Program.

Microbial environmental feedbacks and the evolution of soil organic matter

Stephany Chacon[1], Aizah Khurram[1], Hoi-Ying Holman[2], Liang Chen[2], Lee Dietrich[3], Daniela Cusack[3], **Nicholas J Bouskill[1]**

[1]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720.

[2]Berkeley Synchrotron Infrared Structural Biology Program, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720.

[3]Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, 80523.

The vast majority of Earth's organic matter is stored in soil. The products of microbial metabolism as well as dead microbes (necromass), along with residues from plants and other organisms at different stages of decomposition, constitute a large fraction of that soil organic matter (SOM). The ability of microbes to modify and degrade SOM depends on physicochemical characteristics of the soil, affecting SOM stability and persistence. While the contributions of microbes to the decomposition and loss of SOM have been intensively studied, their role in maintaining terrestrial SOM is poorly understood. Specifically, how fungi, bacteria, and archaea participate in SOM production, its interaction with minerals, and the formation of soil aggregates remains a significant gap in our understanding of the terrestrial carbon cycle. Herein, we employ field and laboratory experiments to further understand the role of microbial communities in stabilizing SOM under drought and fluctuating wet-dry cycles in clay-rich tropical soils. We begin by identifying traits characteristic of the single cell physiological response to drought stress through real-time and non-destructive Fourier-transform infrared spectroscopy at the Advanced Light Source at LBNL. This dataset serves as a baseline for the interpretation of molecular data generated across several different scales in complexity, beginning with simple community responses to drought within an abstraction of the soil environment. We scale up through a mesocosm experiment with intact soil cores collected from Isla Buena Vista, a kaolinite rich soil within a tropical rainforest, which also serves as the location of a long-term drying experiment (termed *Parched.Panama*), that serves as the final experimental scale. We use a combination of metagenomic sequencing and metabolomic profiling to examine how traits expressed within the laboratory and field experiments feedback on to the composition of soil organic matter. These three interacting scales will be used to parameterize and benchmark a mechanistic model coupling above- and belowground processes (the *ecosys* model), to predict the fate of SOM on longer time scales.

Department of Energy Early Career Research Program Grants to N.J. Bouskill, #FP00005182 and D.F. Cusack #DE-SC0015898. This research used resources of the Berkeley Synchrotron Infrared Structural BioImaging (BSISB) Program at the Advanced Light Source, which is a DOE Office of Science User Facility under contract no. DE-AC02-05CH11231.

Enabling Predictive Metabolic Modeling of Diurnal Growth Using a Multi-Scale Multi-Paradigm Approach

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 are currently using circadian gene expression data from *Chlamydomonas reinhardtii* to cluster gene expression, convert discrete data into continuous functions and implement these as additional constraints on our metabolic model. We will present the result of this work on how these constraints are able to further constrain 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

Harnessing Metagenomic Stable Isotope Probing to Uncover the Carbon Cycling Capacity of Soil Microbes

Samuel E. Barnett* (seb369@cornell.edu) and **Daniel H. Buckley**

School of Integrative Plant Science, Cornell University, Ithaca, NY

Project goals: We aim to recover metagenome assembled genomes of bacteria active in soil carbon cycling by enriching for isotopically labeled DNA from ^{13}C-labeled substrate treated soils with metagenomic-SIP. The goal of this project is to investigate these genomes for signatures of bacterial life history strategies that drive dynamic function in soil carbon cycling.

Soil dwelling microorganisms are an essential part of soil carbon cycling yet their biology and ecology are still poorly understood. There is a lack of representative, cultured isolates from this system and diverse, high-quality genomes are only now being recovered from large metagenome studies. DNA-stable isotope probing (DNA-SIP) has recently allowed us to more thoroughly explore the ecology of bacteria active in the soil carbon cycle. Previously, we ran a high-resolution DNA-SIP experiment to characterize bacteria involved in breaking down nine chemically distinct carbon substrates mimicking natural components of plant litter. By analyzing ^{13}C-labeling patterns of community members over 48 days, we uncovered discrete ecological groups which we believe define important life history strategies. We hypothesize that these ecological groups have distinct genome characteristics explaining their carbon assimilation dynamics.

To elucidate the biology behind these ecological groups and better understand their functional potential, we performed metagenomic-SIP on a subset of samples. Metagenomic-SIP enriches for isotopically labeled genomic DNA from treated environmental samples and has been shown to improve assembly and binning of ^{13}C-labeled genomes from diverse communities. Preliminary analysis indicates that metagenomes were enriched for taxonomic groups corresponding to ^{13}C-labeled OTUs. For example, we observed an enrichment of genes from the *Verrucomicrobia* genus *Chthoniobacter*, a prominent assimilator of carbohydrates, with few cultured representatives. By comparing abundance of COG orthologues between ^{13}C-treatment and ^{12}C-control samples, we also observed substrate and timepoint specific enrichment of relevant gene categories and pathways. For instance, cell mobility genes were highly enriched in early timepoint samples but less so in later timepoints. Similarly, mobile elements including prophages and transposons were enriched in early timepoint samples and those treated with highly labile substrates. We further found characteristic enrichment of CAZy-classified glycoside hydrolases families across samples. Through assembly and binning of metagenome assembled genomes (MAGs) we aim to further identify genomic signatures of life history strategies important to soil carbon cycling. This study is part of ongoing research with the goal of unearthing the role of the soil microbial community in global carbon cycling.

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

Phenolic acid-degrading populations of *Paraburkholderia* prime decomposition in forest soils

Roland C. Wilhelm[1*] (rwilhelm@cornell.edu), Christopher DeRito[2], James Shapleigh[2], Eugene Madsen[2] and **Daniel H. Buckley**[1]

[1] School of Integrative Plant Sciences, Bradfield Hall, Cornell University, Ithaca, NY, 14853, USA
[2] Department of Microbiology, Wing Hall, Cornell University, Ithaca, NY, 14853, USA

Project Goals

Our project aimed to characterize the microbial populations involved in the metabolism of plant-derived carbon (C) and their influence on soil carbon cycling. We targeted the ecological and functional traits of phenolic-acid degrading bacteria, since phenolic acids are a major component of plant root exudates and lignin. These populations have also been recently identified as contributing to the soil priming effect, in which exogenous C stimulates the mineralization of endogenous soil organic carbon (SOC). We used stable isotope probing, metagenomics and culturing to link the ecology of phenolic-acid degrading bacteria with soil C-cycling.

Abstract

Plant-derived phenolic acids are metabolized by soil microorganisms whose increased activity can prime the decomposition of SOC. We characterized bacteria that enhanced SOC mineralization in forest soils when primed with ^{13}C-labeled *p*-hydroxybenzoic acid (PHB). We investigated whether PHB-induced priming could explain differences in SOC content among mono-specific tree plantations in a 70-year-old common garden experiment. A set of closely related *Paraburkholderia* and *Caballeronia* phylotypes dominated PHB degradation in all soils despite large differences in community composition with respect to tree species and soil type. We isolated the principal PHB-degrading phylotype (*Paraburkholderia* sp. RP11^{T}) and found it encoded a large number of oxidative enzymes (laccase, peroxidase and dioxygenase) and confirmed its ability to degrade phenolics. RP11^{T} uniquely encoded paralogs of the enzyme responsible for PHB oxidation (*pobA*). The RP11 phylotype (RP11^{ASV}) increased dramatically in relative abundance (23-fold) after PHB amendment, corresponding with the priming of 3 - 13 µmols C g^{-1} dry wt soil of native SOC. In contrast, glucose amendment reduced SOC mineralization by -3 to -8 µmols C g^{-1} dry wt soil. RP11^{ASV} abundance and *pobA* expression correlated with PHB respiration rates and were inversely correlated to *in situ* SOC accumulation. The metabolic state of RP11^{T} cells was critical for priming, which occurred solely during PHB respiration. We conclude that the metabolism of plant-derived phenolic acids stimulates soil priming with potential impacts on SOC cycling.

Funding Statement

This material is based upon work supported by the Department of Energy, Office of Biological & Environmental Research Genomic Science Program under Award Numbers DE-SC0010558 and DE-SC0004486.

RCSB Protein Data Bank: Making connections from genes to ecosystems

John D. Westbrook[1]* (john.westbrook@rcsb.org), Jasmine Young,[1] Robert Lowe,[1] Christine Zardecki,[1] Jose Duarte,[2] and **Stephen K. Burley**[1,2]

[1]Rutgers, The State University of New Jersey, Piscataway, NJ; [2]University of California San Diego, La Jolla, CA

http://www.rcsb.org and http://pdb101.rcsb.org

Project Goals: The Vision of the RCSB PDB is to enable open access to the accumulating knowledge of 3D structure, function, and evolution of biological macromolecules, expanding the frontiers of fundamental biology, biomedicine, and biotechnology.

Protein Data Bank (PDB) was established as the 1st open access digital data resource in biology and medicine. Today, the PDB is one of two global resources for experimental data central to science as a public good (the other key Primary Data Archive being the International Nucleotide Sequence Database Collaboration). PDB currently houses >160,000 atomic level biomolecular structures determined by crystallography, NMR spectroscopy, and 3D electron microscopy. It is managed by the Worldwide Protein Data Bank partnership (wwPDB; wwpdb.org) according to the FAIR principles (*i.e.,* Finability, Accessibility, Interoperability, and Reusability).

Through an internet information portal and downloadable data archive, researchers and educators can access 3D structure data for large biological molecules, such as proteins and DNA. These are the molecules of life, found in all organisms on the planet. Knowing the 3D structure or shape of a biological macromolecule is essential for understanding the role the molecule plays in health and disease of humans, animals, and plants, food and energy production, and other topics of concern to global prosperity and sustainability.

The RCSB PDB operates the US data center for PDB, serves as Archive Keeper for the global PDB archive, and makes PDB data available at no charge to all Data Consumers without limitations on usage. Studies of website usage, bibliometrics, and economics demonstrate the powerful impact of the PDB data on basic and applied research, clinical medicine, education, and the economy.

During calendar 2019, >800 million structure data files were downloaded from the RCSB PDB by Data Consumers working worldwide. During this same period, the RCSB PDB processed >5,530 new atomic level biomolecular structures plus experimental data and metadata coming into the archive from Data Depositors working in the Americas and Oceania. In addition, RCSB PDB served millions RCSB.org users worldwide with PDB data integrated with ~40 external data resources providing rich structural views of fundamental biology, biomedicine, and energy sciences, and supported >650,000 PDB101.rcsb.org educational website users around the globe.

RCSB PDB serves a rich collection application programmable interfaces (APIs) that enable searching of and access to PDB data content. These APIs enable search by key citation, biological and structural features, and retrieval of individual PDB structure entries, reports, and chemical and molecular reference data. Building on this functionality in 2020, RCSB PDB will deploy new API services that integrate structure and sequence comparison in combination with more granular access to the corpus of PDB data.

Access to PDB data and services contribute to patent applications, drug discovery and development, publication of scientific studies, innovations that can lead to new product development and company formation, and STEM education.

Publications

1. The Protein Data Bank (2000) Nucleic Acids Research 28: 235-242. doi: 10.1093/nar/28.1.235.
2. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy (2019) Nucleic Acids Research 47: D464–D474. doi: 10.1093/nar/gky1004.
3. RCSB Protein Data Bank: Enabling biomedical research and drug discovery (2019) Protein Science 29: 52-65 doi: 10.1002/pro.3730.
.

RCSB PDB is funded by the National Science Foundation (DBI-1832184), the US Department of Energy (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM133198.

Spatiotemporal Mapping of Lignocellulose Decomposition by a Naturally Evolved Fungal Garden Microbial Consortium

Paul D. Piehowski[1], Ying Zhu[1], Lily Khadempour[2], Jennifer E. Kyle[3], Dušan Veličković[1], Rosalie K. Chu[1], Lisa M. Bramer[3], Bobbie-Jo M. Webb-Robertson[3], Cameron R. Currie[2], **Kristin E. Burnum-Johnson**[1] (Kristin.Burnum-Johnson@pnnl.gov)

[1] The Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA; [2] Department of Bacteriology, University of Wisconsin-Madison, Madison, WI; [3] Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA

This project will carry out a multi-omics approach 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, microscopic hyphae called gongylidia produced by the fungus to feed the ants, and spatially-resolve regions of the fungus garden at initial to advanced stages of leaf degradation. The lipid species identified in the sample types and garden regions varied significantly. Lipids containing alpha-linolenic acid from the leaves were enriched in the top of the gardens, but not dominant in the middle or bottom regions. Furthermore, gongylidia were dominated by lipids containing linoleic acid. These microscale lipidomic measurements deciphered biomass lipid metabolism and bacterial-fungal lipid syntheses by spatially resolving those regions where specific deconstruction activities are enriched, thus functionally reducing the heterogeneity of the system.

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.

Identifying Plant Genes Associated With Pathogen Antagonism in *Populus trichocarpa*

Devin R. Leopold[1], Piet Jones,[2] Dan Jacobson,[2] **Posy E. Busby**[*1] (posy.busby@oregonstate.edu)

[1]Oregon State University, Corvallis, OR; [2]Oak Ridge National Laboratory, Oak Ridge, TN

Project Goals: Within the leaves of *P. trichocarpa*, antagonistic interactions among non-pathogenic endophytes and pathogens can result in reduced plant disease severity. The overarching goal of our project is to use genome-wide-association-studies (GWAS) to identify and validate plant SNPs/genes associated with the abundance of pathogen antagonists and effective disease suppression in the *P. trichocarpa* leaf microbiome. **Aim 1:** Identify plant SNPs/genes associated with *P. trichocarpa* leaf microbiome composition and the abundance of putative pathogen antagonists in contrasting common garden environments. **Aim 2:** Identify plant SNPs/genes associated with the abundance of known fungal antagonists of *Melampsora* leaf rust in *P. trichocarpa,* and with fungal antagonism of *Melampsora* leaf rust, in a controlled greenhouse environment. **Aim 3:** Validate the effect of selected genes associated with fungal antagonism of *Melampsora* leaf rust using the CRISPR/cas9 system to introduce loss-of-function mutations.

Abstract: Managing plant microbiomes to facilitate pathogen antagonism could complement traditional methods of combating disease (e.g., breeding for resistance and fungicide), enhancing the productivity and sustainability of *Populus* feedstock production for biofuels. Because plants play an active role in shaping the composition of their microbiome, breeding or genetic modification aimed at promoting pathogen antagonism could provide one approach to microbiome management for disease protection. However, before the promise of enhanced pathogen antagonism through these methods can be realized, many basic questions must be answered regarding the genomic basis of host control over endophyte colonization. Toward this end, our research seeks to identify plant genes that influence the species composition of the *P. trichocarpa* leaf microbiome across contrasting environments, with a focus on known and putative antagonists of *Melampsora* leaf rust.

This abstract focuses on the results of our first study aim. We used ITS metabarcoding to characterize fungal communities in the leaves of > 500 *P. trichocarpa* genotypes planted in two common garden environments (Corvallis OR, Boardman OR), in both early (June) and late (September) season. These gardens represent a dramatic contrast in the abiotic environment – approx. 110 cm rain/yr in Corvallis vs. 20 cm rain/yr in Boardman – and dissimilar regional propagule pools of plant pathogens and endophytes. Greater moisture availability is known to positively impact fungal growth and survival. We found large differences in fungal species composition between sites (Fig 1), and between the early and late season samples within each garden (Fig 1). Species richness and diversity also varied across space and time. Observed richness was greater in the wet site, though differed minimally between early and late season

sampling points in both gardens. However, evenness increased over time in Corvallis and decrease over time in Boardman.

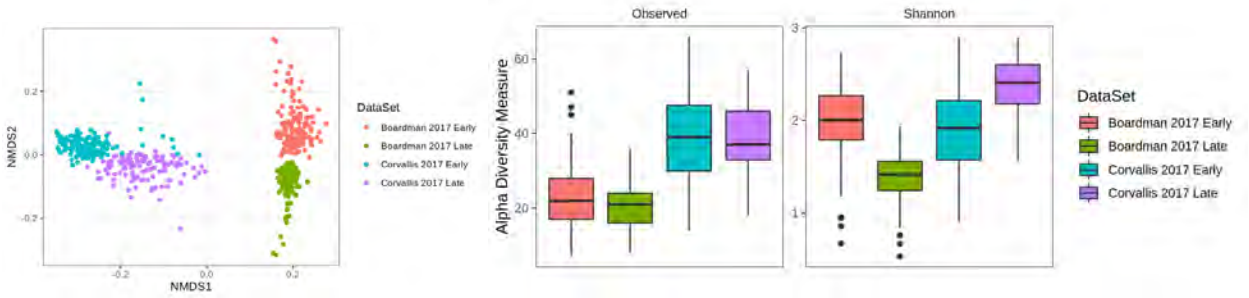


Figure 1. Species composition and diversity of foliar fungi sampled from identical *Populus trichocarpa* genotypes in two contrasting common gardens (Boardman and Corvallis, Oregon) and at two time points (June and September). Left: Non-metric multidimensional scaling (NMDS) reveals large differences in species composition between gardens and at early and late season time points. Right: Observed species richness was consistently higher in the wetter environment (Corvallis). Shannon diversity in the two gardens was similar early in the growing season but diverged over time.

In our preliminary GWAS analysis of the four datasets, the majority of plant SNPs/genes associating with differential relative abundances of individual fungi varied among the datasets. However, across all datasets we identified fungi significantly associated with genes involved in phytohormone action, cell wall modification, lipid metabolism, and protein modification (among others). Boardman shows a larger number of genes associated with vesicle trafficking, while Corvallis has a higher prevalence of solute transport genes. In comparing early versus late season samples within gardens we see modest differences in the prevalence of genes associated with protein homeostasis, protein modification, and solute transport. Some fungi were consistently associated with SNPs/genes across all four datasets, including putative pathogen antagonist species in the genus *Cladosporium*. Though about half the fungi for which we found significant associations with SNPs/genes were identified in just one dataset.

Funding Statement: This work was supported by the Departments of Energy's Office of Science, Office of Biological and Environmental Research, and the Department of Agriculture, National Institute of Food and Agriculture, Agriculture and Food Research Initiative (DE-SC0019435).

Elucidating Principles of Bacterial-Fungal Interactions

Samuel Britton[1], Francesco Pancaldi[1], Mark Alber[1], Dale Pelletier[2], Jessy Labbe[2], and **William R. Cannon[3]**

[1] Department of Mathematics, University of California, Riverside, Riverside, CA 92521
[2] Oak Ridge National Laboratory, Oak Ridge, TN, 37830
[3] Pacific Northwest National Laboratory, Richland, WA, 99352;

Project Goals: The goals of this new 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 will:

- Evaluate 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.
- Characterize *Laccaria*-stimulated chemotaxis of *P. fluorescens* by coupling trehalose signaling and metabolism to chemotaxis *P. fluorescens*.
- Experimentally investigate (1) *Pseudomonas fluorescens* chemotaxis and metabolism of *Laccaria* produced metabolites, and metabolism of *P. fluorescens* produced metabolites in *Laccaria*.

Abstract. In comparison to bacterial-bacterial interactions, there is very little known about bacterial-fungal interactions even though these interactions are thought to be fundamentally important to DOE missions in sustainability, crop biofuel development and biosystem design. In biofuel crops, many crop root systems live in mutualistic symbiosis with fungi and bacteria. Mycorrhiza helper bacteria (MHB) increase host root colonization by mycorrhizal fungi, which in turn act as a micro-root system to provide the plant with soil nutrients. Recent work on the *Populus* root microbiome has determined that the interactions between the mycorrhizal fungus *Laccaria bicolor* and the bacterium, *Pseudomonas fluorescens* are key to fitness of the plant. These organisms, *Laccaria* and *P. fluorescens*, are the focus of this project to understand fundamental principles of interactions between fungi and bacteria from the perspective of material exchange and energetics, and how material and energetics are linked in inter- and intra-microbial subsystems.

The first task of this project is to develop a *hybrid simulation-machine learning model of Laccaria*. We are constructing a hybrid simulation-reinforcement learning model for *Laccaria* that combines metabolic control analysis, physics-based mass action kinetics with experimental data to predict experimentally faithful dynamics of the organism. The model will include all of central metabolism and secondary metabolism including protein production and the production

of biosynthetic enzymes. Under this task, we are using the *Laccaria* model to understand the growth energetics of *Laccaria* under various conditions pertaining to the questions of interest. We will evaluate the material and energy flow in *Laccaria* when sub- inhibitory levels of phenazines are present, with the naïve hypothesis is that material flow will be redirected away from respiration, possibly towards filament growth or branching. We will evaluate the impact of thiamine uptake from the environment on metabolism, with the naïve hypothesis that filament growth and branching will increase. We will predict conditions in which trehalose synthesis and export are favorable, and conditions in which water generated from metabolism is exported out of the cell. Currently the model includes central metabolism, cell wall synthesis and trehalose synthesis, as shown in the figure.

The second task is to construct a similar hybrid simulation-machine learning model of *P. fluorescens*, but to also couple the metabolism of this model with chemotaxis. We will use the *P. fluorescens* model to evaluate the change in growth and thermodynamic cost of thiamine production, and to understand the mechanism of trehalose stimulated chemotaxis of *P. fluorescens.*

The thiamine biosynthesis pathway is regulated by a thiamine riboswitch in *P. fluorescens*. When thiamine taken up from the environment, the riboswitch turns off the production of the enzymes needed to synthesize thiamine. The thermodynamic cost of thiamine production can be determined by understanding the changes in enzyme expression/activity, which the model will predict, when thiamine is present or not. The cost of enzyme synthesis can be estimated from either average cost of protein synthesis, or specifically from the amino acid sequences of the specific enzymes.

In order to understand the interplay and feedback between metabolism and chemotaxis, we are coupling metabolism to chemotaxis in *P. fluorescens.* Stimulation of the two-component chemoreceptor system alone is not sufficient to drive chemotaxis. Energy must also be redirected from other cellular processes to drive the rotation of flagella. Virtual experiments are helping us to develop hypotheses on how metabolism might be redirected for this purpose.

The third task is to experimentally investigate (1) *P. fluorescens* chemotaxis and metabolism of *Laccaria* produced metabolites and (2) biofilm formation between *P. fluorescens* and *Laccaria*.

We are initially focusing on several *Pseudomonas* strains that have differential responses on *Laccaria* (e.g. GM41, GM18, GM17). Bacteria utilize a chemoreceptor-phosphorelay system to sense and respond to chemical gradients. The binding of attractants to membrane bound methyl-accepting chemotaxis proteins (MCPs) initiates this process which leads to effects on motility behavior.

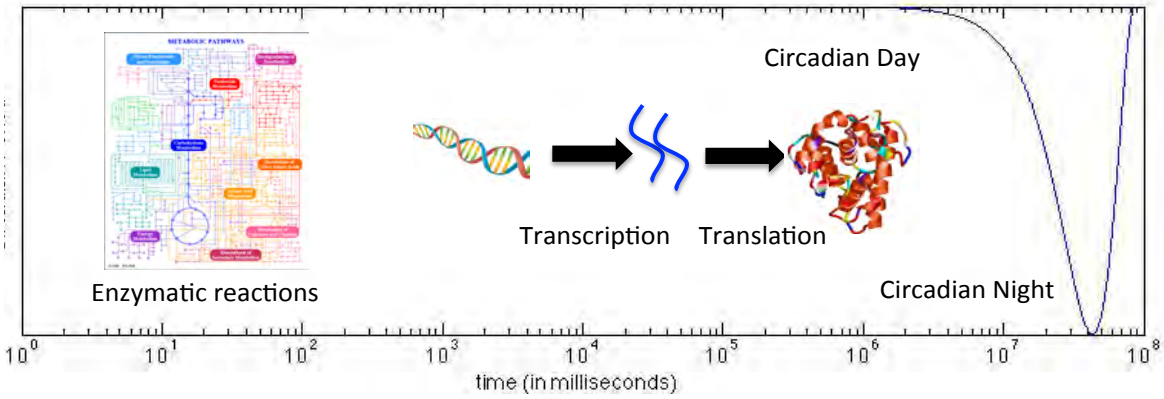
This project is supported by the U.S. Department of Energy's Office of Biological and Environmental Research.

LEARNING REGULATION AND OPTIMAL CONTROL OF ENZYME ACTIVITIES AND APPLICATION TO SYSTEMS BIOLOGY DATA OF NEUROSPORA CRASSA.

William R. Cannon[1], Samuel Britton[2], Mark Alber[2], Jennifer M. Hurley[3], Meaghan S. Jankowski[3], Tina Keliher[4], Jeremy D. Zucker[1], Scott E. Baker[5], Jay C. Dunlap[4]

[1] Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99352;
[2] Department of Mathematics, University of California, Riverside, Riverside, CA 92521
[3] Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY, 12180
[4] Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH 03755
[5] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, 99352

Project Goals: The goal of this research is to develop and implement a new computational and theoretical method for modeling biological systems that fills a gap in modeling mass action dynamics. Based on statistical thermodynamics, the method bridges data-poor scales (parameters for mass action kinetics) and data-rich scales (chemical potentials of metabolites, and metabolite, protein & transcript data) to enable predictive modeling from enzymatic reactions (10^{-3} to 1 s^{-1}) to gene and protein regulation (~20 minutes) to circadian rhythms (24 hours).

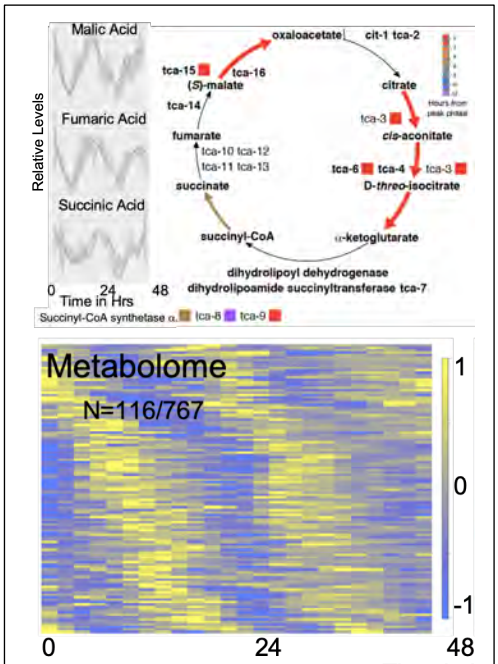


Timescales that the simulations using statistical thermodynamics will cover. Enzymatic reactions occur on the millisecond to second timescale while gene and protein expression occur on the minute to ~30-minute scale and the circadian rhythm occurs over a period of 24 hours.

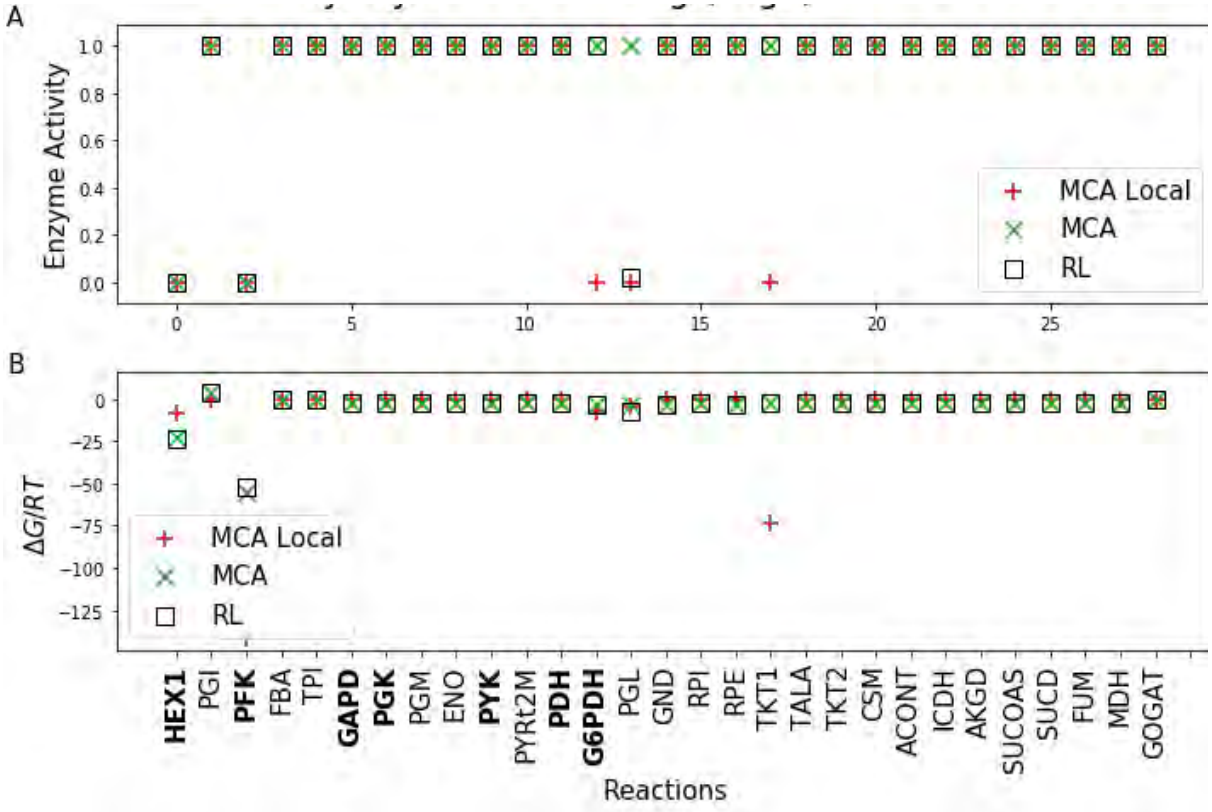
Abstract. Experimental measurement or computational inference/prediction of the enzyme regulation needed in a metabolic pathway is hard problem. Consequently, regulation is known only for well-studied reactions of central metabolism in model organisms. In the last year, we use statistical thermodynamics and metabolic control theory as a theoretical framework to determine the enzyme activities that are needed to control metabolite concentrations such that they are consistent with experimentally measured values. A reinforcement learning approach is utilized to learn optimal regulation policies that match physiological levels of metabolites while maximizing the entropy production rate and minimizing the work needed to maintain a steady state. The learning takes a minimal amount of time, and efficient regulation schemes were learned that agree surprisingly well with known regulation. The learning is facilitated by a new approach in which steady state solutions are obtained by optimization rather than ODE solvers, making the time to solution seconds rather than days. The optimization is based on the Marcelin-De Donder formulation of mass action kinetics [1]. Consequently, a full ODE-based, mass action simulation with rate parameters and post-translational regulation is obtained.

We demonstrate the process on the central metabolism of *Neurospora crassa* which requires different regulation schemes under different nutrient conditions. While many reasons have been proposed as to why enzyme activities are regulated, very few of the proposals lead to specific hypotheses that can be tested. We hypothesize that the post-transcriptional regulation of enzymes is at least in part driven by the need to maintain the solvent capacity in the cell, a directly testable hypothesis.

We investigate this hypothesis computationally by combining experimental metabolomics data with steady state concentrations predicted computationally from equations for reformulated mass action kinetics, which can be solved either by simulation or by optimization. Using quantitative metabolomics data, including both absolute quantitation and relative quantitation over circadian time (right) as well as physical and biological principles, we can predict the control of activity required to bring metabolite levels down to observed values using reinforcement learning (RL) or metabolic control analysis (MCA, local MCA). The predicted activities (below) agree with known regulation of central metabolism in model organisms (highlighted in bold in figure axis). Moreover, the results show that regulated enzymes have higher free energies of reaction precisely because of the regulation, turning common wisdom about enzyme regulation upside-down. Instead of highly non-equilibrium reactions being the targets for regulation in metabolic pathways [2, 3], regulation results in reactions being much further from equilibrium than non-regulated reactions. Being further away from equilibrium than other reactions is an effect, not a cause, of regulation.



Key metabolite levels in the glycolysis pathway in *Neurospora* oscillate in a circadian manner in time with the circadian oscillation of the enzymes that are responsible for producing them.



(Top) Plot of predicted enzyme activity vs enzyme for enzymes of glycolysis, the pentose phosphate pathway and the TCA cycle. Enzymes in bold on x-axis are known to be post-translationally regulated. Activities with values of 1.0 are not regulated while values of 0.0 indicate full regulation. (Bottom) Free energy of reactions for the reactions catalyzed by the respective enzymes.

References

1. Cannon, W.R., et al., *Prediction of metabolite concentrations, rate constants and post-translational regulation using maximum entropy-based simulations with application to central metabolism of Neurospora crassa.* Processes, 2018. **6**(6).
2. Newsholme, E.A. and C. Start, *Regulation in Metabolism*. 1973, London: John Wiley & Sons. 349.
3. Lehninger, A.L., et al., *Lehninger principles of biochemistry*. 5th ed. 2008, New York: W.H. Freeman.

This project is supported by the U.S. Department of Energy's Office of Biological and Environmental Research and the National Institute of Biomedical Imaging and Bioengineering.

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

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

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

Project Goals: Our primary goals are to accelerate the development of superior switchgrass cultivars and to expand the range of switchgrass cultivation in the Northeast. We are building on the cultivars, populations, and field trials resulting from previous projects by expanding progeny trials and creating higher levels of associations from genomics information. To this end we are addressing the following specific objectives: 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 for development 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 previous research, populations and genomics tools 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 is specifically focusing on improvement of resistance to anthracnose (caused by *Colletotrichum navitas*), *Bipolaris* leaf spot (caused by *Bipolaris oryzae*), and environmental stress-resistance. In this poster, the results of the metagenomics and phenotypic data collections and analyses in year 1 will be presented.

In year 1 (December 2018 – November 2019), progress was made on all project objectives. This included seeding new plots of advanced experimental populations and standard check cultivars at all 3 test locations, developing a uniform set of methods for phenotyping and conducting disease ratings across the project, and identifying superior breeding lines for growth and disease tolerance from mature cultivar nurseries at Cornell University and Rutgers University. New progeny trials were established with 5,760 seedlings of progeny from 180 advanced breeding lines selected in the USDA NEWBio project. These progeny trials will be evaluated for disease resistance and growth phenotypes in years 2 and 3. The [Lu et al. 2013] association panel was well established at all 3 trial sites by the start of the project, and growth was abundant at the field sites during the 2019 growing season. The Cornell, Rutgers, and PSU groups completed data collection for plant growth (height and circumference in cm), plant vigor (rated from 1 to 5), and

anthracnose severity (rated from 1 to 5) from replicates for all 552 genotypes in the panel. Bipolaris infections were not observed in 2019. Values for plant volume were computed as a proxy for biomass yield. Heritability estimates for each trait in 2019 were calculated from the phenotypic data. All replicates at the 3 trials will be scored for disease ratings and growth traits again in project year 2 (the 2020 growing season). We observed substantial, normally-distributed quantitative phenotypic variation between the sites and among genotypes for all of the traits studied. A trend in site-productivity of Rutgers>Cornell> PSU was observed for height, circumference and volume. Anthracnose incidence was severe at all 3 sites. Furthermore, heritability levels were high for all of the traits at all 3 trial sites. The results in year 1 indicate strong genetic and site effects on variation in traits that should allow GWAS and GxE analyses to uncover genes and alleles important in biomass productivity and disease-resistance.

A third year of phenotypic data was collected from Rutgers University's QTL mapping family, including scores for anthracnose severity and plant vigor. The 3 years of phenotypic data will next be used to construct a genetic linkage map and to identify QTL. Most of the plants were too severely infected in the 2019 growing season to permit high quality and contamination-free DNA to be purified from field-collected samples. Thus, after disease ratings were taken, the mapping population was cut back and all 240 plants were brought into the greenhouse for re-growth under controlled conditions for the collection of fresh tissues for DNA isolations.

In pre-award work, triplicate soil and root samples were collected from all of the association population plants within 2 days after the clonal "plugs" of each genotype were collected in triplicate from the switchgrass provenance trial common garden site at Cornell University near Ithaca, NY [Lu et al. 2013]. DNA was isolated from all rhizosphere soil samples. The 382 samples with highest DNA quality and representing all wild population groups [Lu et al. 2013] were selected for the initial "pre-transplant" metagenome sequencing. Amplicons for the 16S rRNA gene and ITS loci were generated, and separate NGS libraries produced, for each of the 382 samples. From the pooled libraries, app. 60Gb of sequence was produced on an Illumina HiSeq 2000. The pre-transplant metagenome data revealed an established rhizosphere microbial community with a rich composition of at least 493 bacterial genera and 57 fungal genera. This provides the composition of rhizosphere microbiomes representing past interactions between the switchgrass genotypes and both aboveground and belowground environmental conditions at the original nursery at Cornell, representing the common starting point of our project. From the 382 samples at the transplant stage, we selected 128 association panel genotypes in which to conduct detailed monitoring of changes in rhizosphere microbiome composition that may occur as the association population becomes re-established in the 3 different field trial sites in PA, NJ, and NY. The 128 genotypes were selected to maximize the range of switchgrass source populations represented. In July 2019, triplicate rhizosphere soil samples were collected from all 3 replicates of the 128 core set of genotypes at all 3 sites, to assess changes in the microbiomes that may have occurred during the establishment of the genotypes at the 3 trial sites. DNA isolations, amplifications, and sequencing will be completed early in year 2 (2020).

References cited:

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.

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.

Examinations of the Fungal Genus *Monosporascus* Reveal its Potential as an Experimental Model for Studying Bacterial-Fungal Interactions

Aaron J. Robinson[1,2]*, (arobin@lanl.gov), Donald O. Natvig,[2] Demosthenes Morales,[1] Julia M. Kelliher,[1] Geoffrey L. House,[1] LaVerne A. Gallegos-Graves,[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://genomicscience.energy.gov/research/sfas/lanlbfi.shtml

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.

The genus *Monosporascus* represents an enigmatic group of fungi important in agriculture and widely distributed in natural arid ecosystems. Of the eight described species of *Monosporascus*, two (*M. cannonballus* and *M. eutypoides*) are important pathogens on the roots of members of the Cucurbitaceae in agricultural settings. The remaining six species are capable of colonizing roots from a diverse host range without causing obvious symptoms of disease. To explore evolutionary relationships within the genus and gain insights into potential ecological functions, we sequenced and assembled the genomes of three *M. cannonballus* isolates, one *M. ibericus* isolate and six phylogenetically distinct New Mexico isolates. The assembled genomes were significantly larger than what was expected for the Sordariomycetes, despite having predicted gene numbers similar to other members of the Class. Differences in predicted genome content and organization were observed between endophytic and pathogenic lineages of *Monosporascus*,

The sequencing results from several of these *Monosporascus* isolates contained a significant 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. Phylogenetic comparisons indicate that the *R. pickettii* sequences found within *Monosporascus* represent multiple distinct lineages that are closely related to previously identified environmental isolates. Broad-scale evolutionary comparisons conducted with the limited sequencing data also suggest differences between the endosymbiotic and environmental *R. pickettii* lineages. The diversity

Integrating read-based microbiome taxonomy classification tools into KBase

Mark Flynn (mflynn@lanl.gov),[1*] Chien-Chi Lo,[1] Paul Li,[1] Karen Davenport,[1] Bin Hu,[1]and Patrick S. G. Chain [1]

[1]Los Alamos National Laboratory, Los Alamos, NM

http://bioedge.lanl.gov
https://edgebioinformatics.org

Project Goals: Our goal is to add a new capability, read-based taxonomy, to KBase. These classification tools will allow KBase users to analyze bacterial-fungal interactions by determining which organisms are present in a microbiome sample.

Interactions between bacteria and fungi play a critical role in soil ecosystems. To facilitate the study of these interactions we have added a new capability to KBase, read-based taxonomic classification. Currently KBase has only one taxonomy classifier, the amino acid-based Kaiju algorithm. While it is highly sensitive and able to detect all organisms in a sample that are in its database, it has very low specificity which results in many false positive classifications. To help overcome this problem, we have added four taxonomic classifiers to KBase: Centrifuge, Kraken2, as well as two profilers developed by our team at LANL: GOTTCHA2 and PanGIA. These nucleotide-based classifiers each make different trade-offs between sensitivity and specificity. GOTTCHA2 uses unique signatures within genomes to significantly reduce the number of false positives while maintaining high sensitivity. PanGIA strikes a balance between sensitivity and specificity. Centrifuge and Kraken2 have higher sensitivity and are considerably faster than GOTTCHA2 and PanGIA but have lower specificity. We chose to add all four classifiers to KBase since there may be circumstances where several tools in conjunction would provide a more comprehensive analysis.
We are in the process of creating an app that allows the generation of a custom database of unique genomic signatures from any user-supplied set of references. This app will be used to power the dynamic updating of a database of unique fungal sequence signatures as the JGI adds additional fungal genome assemblies to their publicly available datasets.
Adding these new capabilities will allow users to make use of KBase's flexible platform to: 1) perform enhanced metagenomic data analytics; and 2) and determine unique genomic signatures at each taxonomic rank for fungal genomes (similar to what exists for bacterial genomes) to allow unambiguous assignment and to reduce false discovery.

This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59C

Investigating Chloroplast Signal within the Hyphae of Diverse Fungi

Julia M. Kelliher[1]* (jkelliher@lanl.gov), Aaron J. Robinson,[1] Demosthenes P. Morales,[1] Geoffrey L. House,[1] La Verne A. Gallegos-Graves,[1] Hang N. Nguyen,[3] Simone Lupini,[3] Debora F. Rodrigues,[3] Saskia Bindschedler,[2] Pilar Junier,[2] and **Patrick S.G. Chain[1]**

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico; [2]Insitute of Biology, University of Neuchâtel, Neuchâtel, Switzerland; [3]Civil and Environmental Engineering, University of Houston, Houston, Texas

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

Project Goals: The vast taxonomic diversity and the complexity of interactions within the soil microbiome presents a unique challenge. Many of the interactions between soil-dwelling bacteria and fungi are not yet well understood, and a more comprehensive understanding 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 fields. The aim of this Science Focus Area (SFA) is to better understand the diverse and abundant interactions within the soil rhizosphere and decipher the mechanisms behind their communication. Herein, we explored an unexpected endo-hyphal signal that was revealed following an amplicon screen of fungi from diverse culture collections.

The endo-hyphal microbiome contains many uncharacterized inhabitants and interactions. This fungal microbiome, coupled with an extensive network of extracellular interactions with bacteria and plants within the soil, contribute to the complex ecosystem services facilitated by fungi. We sought to characterize the members of the intracellular fungal microbiome as a way to better understand the roles of fungi and their associated endosymbionts. Based on a 16S rRNA amplicon screen of four distinct fungal collections from different geographic locations, we identified taxonomic signatures of many bacteria not previously known to be associated with fungi. Rather unexpectedly, one of the amplicon sequence signatures that was found across all culture collections, was a recurrent signal for various plant chloroplasts. Several techniques were utilized in validating the potential associations between fungi and the chloroplasts (which we anticipate are represented by these signatures), including FISH staining, phylogenetic analyses, qPCR and PCR amplifications, as well as broader bioinformatic screens. This discovery is leading to several new avenues of research to explore the acquisition, maintenance, evolution, as well as any physiological or ecological function of chloroplasts within fungi.

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

LA-UR-20-20696

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

of both the host fungus and these closely related endobacterial partners suggest this relationship could be valuable in establishing an experimental 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

Coupling Metabolic Source Isotopic Pair Labeling And Genome Wide Association For Metabolite And Gene Annotation In Plants

Jeff Simpson,[1,2] Cole Wunderlich,[1,2] Rajdeep Khangura,[1,2] Brian Dilkes,[1,2] and **Clint Chapple,[1,2]* (chapple@purdue.edu)**

[1]Department of Biochemistry, Purdue University, West Lafayette, IN; [2]Purdue Center for Plant Biology, West Lafayette, IN

Project Goals: Improving our understanding of plant genomes and metabolomes is critical to unlocking higher plant productivity, developing new strategies to protect crops from stress, and identifying sources of new plant-based products. Progress towards these goals is currently 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 are addressing these challenges using our recently developed stable isotope feeding/LC-MS/genome wide association (GWA) strategy to identify functional gene-metabolite relationships for metabolites that are derived from amino acids in Arabidopsis and sorghum. The aims of this work are to establish precursor-of-origin annotations for plant metabolites; use metabolite annotations to identify functional gene-metabolite relationships using GWA; and authenticate these associations by reverse genetics and biochemical assays of enzyme activity.

Enhanced sensitivity and throughput for analytical tools in metabolomics, principally liquid chromatography-mass spectrometry (LC-MS), permit the measurement of thousands of metabolites in complex plant extracts in a single analytical run. Parallel reductions in the cost of nucleic acid sequencing technologies and informatics has identified all of the genes encoded by many plant genomes and provided data about their expression. Despite these technical leaps, we cannot rationally design plant products because we do not know the functions of most genes in plant genomes nor the identity of most plant metabolites. We propose to provide fundamental understanding of metabolic gene functions and leverage “omics” datasets to annotate gene functions and identify metabolites.

Whole genome sequencing has identified millions of single nucleotide polymorphisms (SNPs) segregating in populations of many plant species, including Arabidopsis and sorghum. It is now possible to use these nucleotide polymorphisms to identify the genetic mechanism of natural variation in traits, including metabolite accumulation. The advantage of combing metabolite profiling and this Genome Wide Association (GWA) analyses lies in the ability of metabolite levels to report on enzyme function and genetic mapping to reveal gene-metabolite associations without prior knowledge that any such association exists.

We have developed a robust isotopic labeling/ LC-MS/ GWA approach to meet the challenges associated with identifying metabolites and annotating genes within plant metabolic pathways. We initially focused on identifying phenylalanine-derived metabolites in Arabidopsis, and genes

associated with their production. The experimental approach and bioinformatic pipeline is now being adapted to identify metabolites derived from any metabolic pathway for which isotopically labeled precursors are available. Briefly, metabolites were extracted from Arabidopsis stems fed [6-13C] ring-labeled Phe or unlabeled Phe and metabolite profiles were obtained by LC-MS. The chromatograms were aligned and queried for the presence of unlabeled and isotopically labeled "peak pairs" in 13C-Phe-fed samples to both detect and measure the relative abundances of unlabeled metabolites and their isotopologues. This method identified more than 500 phenylalanine-derived LC-MS features (the "phenylalanome"), many of which represent compounds that had not been described previously. Matches for these 500 phenylalanome mass features were identified in an unlabeled LC-MS dataset collected from an association mapping panel of Arabidopsis. GWA identified strong associations between these mass features and multiple genes encoding known enzymes in the phenylpropanoid pathway as well as enzymes with no validated functions. Many of these associations were confirmed by mutant analysis when the phe-derived metabolite was gained or lost in a mutant corresponding to the gene predicted by GWA. Thus, this approach can be used to simultaneously discover the pathway of origin for metabolites and to definitively link genes of unknown function to pathways, metabolic products, and reactions.

This work will synthesize our capacity to sequence plant genomes and classify metabolites based on their pathway of origin to provide functional annotation to genes associated with plant metabolism. Here we propose to classify metabolites produced from metabolic pathways that require an amino acid as their precursor, then utilize GWA and mutant resources in Arabidopsis and Sorghum to inform species-specific annotations for genes associated with plant metabolism.

This work is supported by the U.S. Department of Energy under Award DE-SC0020368.

Characterizing the portability of RecT-mediated oligonucleotide recombination

Gabriel T. Filsinger[a,b] (filsinger@g.harvard.edu), Timothy M. Wannier[b,c,1], Felix B. Pedersen[d,1], Isaac D. Lutz[e,1], Julie Zhang[f,1], Devon A. Stork[b,g], Kevin Gozzi[h], Helene Kuchwara[c], Verena Volf[b,i], Stan Wang[b,c], Seth L. Shipman[j,k], John Aach[c], Michael T. Laub[h], **George M. Church**[b, c, 2].

[a]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA.

[b]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts, USA.

[c]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.

[d]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark.

[e]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

[f]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

[g]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA.

[h]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[i]Harvard University John A. Paulson School of Engineering and Applied Sciences, Cambridge, Massachusetts, USA.

[j]Gladstone Institutes, San Francisco, CA

[k]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA

1. These authors contributed equally

Project Goals: Characterize the requirements for the portability of recombineering methods

Abstract: Improved methods for genome editing in bacteria would enhance our ability to engineer human commensals, pathogens, and chassis for biosynthesis. Currently, methods of homologous recombination are limited in microbes since Cas9 is lethal in a majority of species. Phage encoded RecT proteins improve the efficiency of homologous recombination, but these proteins have been established in few species and are not broadly portable. Here, we reveal that this host-limitation is minimally defined by a requirement for compatibility between phage RecTs and the host's single-stranded DNA-binding protein (SSB). We characterize the RecT-SSB interaction, finding that it is mediated by 7 amino acids on the SSB C-terminal tail, and provide evidence that RecTs are portable between species where a host SSB interaction is maintained. Co-expressing cognate RecT-SSB pairs broadens recombineering activity, and in certain species improves recombination efficiency up to ~1000 fold even when the RecT proteins alone are non-functional. We use both rational selection of RecT proteins, and the screening of RecT-SSB pairs to establish

oligonucleotide recombination in *L. rhamnosus* and *C. crescentus*. We then show how dominant negative versions of the host MutL protein can enable efficient single-nucleotide mutagenesis in species beyond *E. coli*. In *L. lactis* we use an optimized recombineering method with dominant negative mismatch repair to generate libraries of variants at selected genomic loci, and show that the method far surpasses the capabilities of error-prone or mutagenic methods. Specifically, we make millions of mutations within the spectinomycin binding pocket, and characterize a unidentified landscape of epistatic effects that include 2-5 mutations. Due to the requirement for many simultaneous mutations, these variants are inaccessible through error-prone genomic diversification methods, but have the highest fitness in the presence and absence of selection. This work elucidates requirements for the portability of recombineering methods and emphasizes recombineering as the pre-eminent technique for generating genomic variants beyond what is accessible using evolution.

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

Deep Mutational Learning of Protein Function for New Intracellular Biosensors

Alexander S. Garruss[1,2,3]* (garruss@fas.harvard.edu), Katherine Collins[2], and **George Church**[1,2]

[1]Harvard Medical School, Department of Genetics, Boston, MA,[2]Wyss Institute for Bioinspired Engineering at Harvard University, [3]Program in Bioinformatics and Integrative Genomics, Harvard University, Cambridge, MA

Project Goals: Develop high-throughput computational and experimental approaches to deeply characterize natural allosteric transcription factor function so to accelerate the design and engineering of new factors that respond to molecules of special interest to the Department of Energy.

Recent progress in DNA synthesis and sequencing technologies have enabled systematic studies of protein function at a massive scale. We apply deep mutational scanning, a process whereby a library of protein variants modified at each position to every alternative amino acid are assembled and assayed in parallel, to study the sequence-structure-function relationships of allosteric transcription factors in bacteria. Such transcription factors are interesting since they intrinsically couple the binding of a small molecule to the binding of DNA and have emerged as useful tools in synthetic biology as intracellular biosensors. We constructed deep mutational scanning libraries for several allosteric transcription factors, including lacI, and screened these libraries to determine variants that can no longer bind DNA or no longer release DNA under induction from their native inducers. We also screend protein variant response to several new ligands to gain an understanding of protein-ligand binding relationships. Furthermore, we utilize advanced deep learning methodologies, integrating experimental results with large-scale measures of protein conservation and associated predictions from molecular modeling/simulation tools, to reveal essential regions of protein function and to provide useful predictions for custom biosensor design.

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

Genetic Code Expansion in *Bacillus subtilis*

Devon Stork[1]* (dstork@g.hmc.edu), Erkin Kuru,[1] Aditya M. Kunjapur,[1,2] Ethan Garner,[1] and **George Church**[1]

[1]Harvard University, Boston;[2]University of Delaware, Newark

Our goals are to establish broad genetic code expansion tools in the primary gram-positive model bacterium *Bacillus subtilis*. We aim to transfer and characterize most noncanonical-amino acid incorporation systems present in *E. coli* to *B. subtilis* and utilize them for several applications.

Encoding nonstandard amino acids (nsAAs) into proteins allows for expansion of the genetic code beyond the standard 20 amino acids for probing, labelling, or controlling proteins in a minimally disruptive manner. However, genetic code expansion has been unavailable in many bacterial model systems, such as the primary gram-positive model and common industrial organism, *Bacillus subtilis.* Here we describe the use of several classes of genome-integrated synthetases to incorporate a variety of different nsAAs into proteins in *B. subtilis* (figure 1) including nsAAs used for biorthogonal labelling, fluorescence imaging, and photo-crosslinking. We also demonstrate a nsAA-titratable protein expression system in this bacterium. Unlike *E. coli* codon expansion systems, where nsAAs were not incorporated into native UAG codons even before recoding efforts, *B. subtilis* nsAA systems incorporate nsAAs into many genomic proteins at native UAG codons. This feature presents both challenges and opportunities for follow-up work in *B. subtilis* nsAA research and genome modification. The general and effective expansion of nsAA technology to *B. subtilis* can facilitate new experiments in this important model bacterium and enable industrial protein production of nsAA-containing proteins.

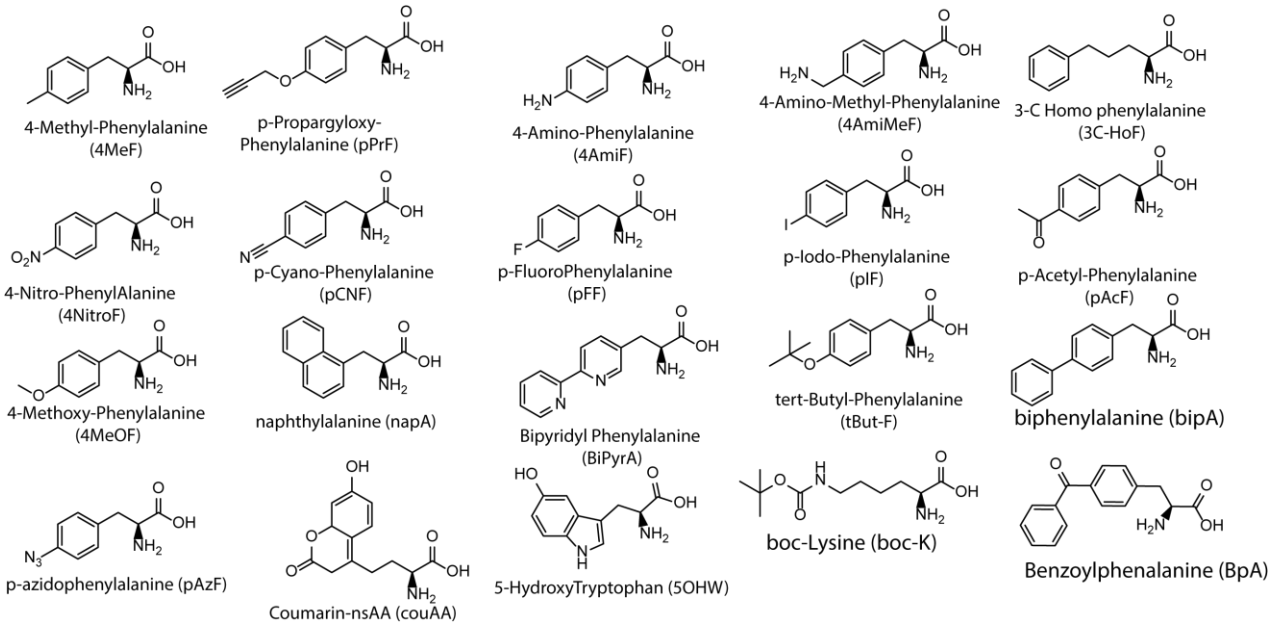


Figure 1. nsAAs so far incorporated into proteins in the organism *B. subtilis*

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 Functional Variant Screens *via* In-vivo Production of Single-stranded DNA

Max G. Schubert[1]* (mschubert@g.harvard.edu), Daniel B. Goodman[2], Fahim Farzadfard,[3] Timothy K. Lu,[3] Seth L. Shipman[2], **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

http://arep.med.harvard.edu

Project Goals: Explore recombineering using substrates produced *in-vivo*, via specialized bacterial retro-elements. Construct a pooled Functional Genomics system, combining efficient editing and NGS-based tracking of mutants. Deploy this tool toward Energy-related goals, and work toward expanding its functionality to diverse bacteria.

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. 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). 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. Pooled genomic editing using ssDNA produced *in vivo* thus presents new avenues for creating and exploring variation at the whole genome scale. We also introduce our future directions, in which RLR could enable unique applications in Bio-Energy and Negative Carbon Emissions, if adapted for use outside of *Escherichia coli.*

This project has been funded by DOE grant DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related fincancial 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.

Parallelized *in vivo* Construction of a Synthetic 57-Codon *E. coli* Genome

Nili Ostrov[1], Akos Nyerges[1*] (Akos_Nyerges@hms.harvard.edu), Maximilien Baas-Thomas[2], 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 assembling a fully recoded, 3.97 Mb *Escherichia coli* genome, in which seven codons are replaced with synonymous alternatives in all protein-coding genes. For this aim, the full recoded genome was *de novo* synthesized and assembled *in vivo* into 87 segments. In the final steps of genome construction, we combine these 87 segments *in vivo* to assemble the fully recoded genome.

We present the synthesis of a fully recoded, 57-codon *Escherichia coli* genome, in which seven codons are replaced with synonymous alternatives in all protein-coding genes. To this aim, the entirely synthetic recoded genome was assembled into 50 kb episomal segments which were then individually tested for functionality. The genome is constructed by CRISPR/Cas9-mediated *in vivo* recombineering, in which each synthetic segment replaces its corresponding wild-type sequence. Multiplex Automated Genome Engineering (MAGE)[1] and directed evolution with random genomic mutations (DIvERGE)[2] are further used to identify alternative recoding schemes. Replacement efficiency was enhanced up to 100% by implementing a novel, three-plasmid CRISPR/Cas9 knock-in technique. Cycle time was reduced to 11 days by extensively streamlining the replacement procedure and accelerating sequencing-based quality-control steps. Importantly, no significant decrease in growth rate has been observed in eight recoded clusters (total up to 500 kb). In parallel with genome construction, we are optimizing conjugative assembly (CAGE)[3] for combining recoded clusters. As we approach the final assembly of a virus-resistant *E. coli* genome, intermediate strains are also used to implement dependency on non-standard amino acids and encode modules for self-destruction for stringent biocontainment of the final strain. Our work expands the toolkit available for large scale engineering in living cells and opens a new avenue for the bottom-up synthesis and refactoring of organismal genomes.

References

1. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature, 460(7257):894–898. https://doi.org/10.1038/nature08187
2. 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

3. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB, Emig CJ, Bang D, Hwang SJ, Jewett MC, Jacobson JM, Church GM (2011) Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. Science, 333(6040):348–353. https://doi.org/10.1126/science.1205822

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.

Uncovering spatial taxonomic structures of synthetic microbial communities using subcellular RNA sequencing.

Andrew C. Pawlowski[1,2]* (Andrew_Pawlowski@hms.harvard.edu), Jenny M. Tam[1,2], and **George M. Church**[1,2].

[1]Harvard Medical School, Boston; [2]Wyss Institute for Biologically Inspired Engineering, Boston, Massachusetts

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

Project Goals:

This study aims to develop methods for characterizing microbial communities with single-cell and spatial resolution. We are employing fluorescent *in situ* sequencing (FISSEQ) technology for acquiring genomic and transcriptomic information to better understand the spatial arrangement of microbes in natural and synthetic microbial communities.

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 the environment (*e.g.*, polymicrobial metabolism, biofilms, horizontal gene transfer). 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 use a synthetic community of 3 microbes and target a conserved region of the small ribosomal subunit that differs by a single nucleotide. We find that anchoring probes to a hydrogel matrix provides superior nucleotide specificity compared to *in situ* crosslinking for maintaining spatial localization of signal. Furthermore, we extend this method from requiring target-specific reverse transcription primers to a mixture of randomers that achieves the same signal and specificity but reduces the cost of probe synthesis by half. We also demonstrate simultaneous identification of each community member with a unique barcode, which we readout with sequential nucleotide labelling using fluorescent reversible terminators. Finally, we have created software for the automated design of probes for unbiased taxon discrimination at any given taxonomic level and apply this for discrimination on the Domain and Class level for the tree-of-life, including discriminating mitochondria from bacteria and bacterial classes. This work forms the framework for investigating the taxonomic structure of natural microbial communities with unknown compositions, and forms the foundation for future whole-transcriptome sequencing of microbes *in situ*.

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.

Revealing the Prevalence and Diversity of a "Rare" Phosphorus Metabolism through Selective Enrichments and Genome Resolved Metagenomics

Sophia Ewens*(sdewens@berkeley.edu)[1, 2], Alexa Gomberg [1], Tyler P. Barnum[1], and **John D. Coates[1, 2]**

[1] Department of Plant and Microbial Biology, University of California, Berkeley
[2] Energy & Biosciences Institute, University of California, Berkeley

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.

Phosphorus (P) is an essential biological nutrient, but the majority of P is biologically unavailable because it is trapped in mineral deposits as oxidized phosphate (P^{+5}). Alternative P redox states are often ignored in global P cycle models, despite the fact that reduced P species have been identified in diverse environments[1]. Phosphite (P^{+3}) is a highly soluble, reduced P compound that can account 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 in cellular energy metabolism[3]. This organism, *Desulfotignum phosphitoxidans* FiPS-3, is an autotrophic acetogen for which DPO activity was attributed to the *ptx-ptd* gene cluster[4]. However, the biochemical mechanism of DPO is still poorly understood. In FiPS-3, the *ptx-ptd* gene cluster occurs on a genomic island, and its closest relatives were incapable of DPO, suggesting that this metabolism was acquired via horizontal gene transfer. This gene cluster has since been identified in two instances of *Ca. Phosphitivorax anaerolimi*: strain Phox-21 was identified in an active DPO enrichment metagenome; and strain F81 was identified in a full scale anaerobic digester metagenome[5,6]. Both strains of *Ca. P. anaerolimi* harbor six of the seven *ptx-ptd* genes, and neither strain shows classical signatures of horizontal gene transfer. Both *Ca. P anaerolimi* strains are missing genes for any known electron accepting pathway, host only partial Wood Ljungdahl pathways for

carbon fixation, and have the genomic capacity to perform syntrophic butyrate oxidation[5,6]. Furthermore, *Ca. P. anaerolimi* is in a distinct taxonomic clade from FiPS-3, with no characterized close relatives[6]. DPO is the most energetically favorable chemotrophic electron donating process known to date, yet *D. phosphitoxidans* and *Ca. P. anaerolimi* are the only identified representatives, thereby limiting our understanding of the phylogenetic and taxonomic diversity of DPO. In an effort to understand the diversity and prevalence of this metabolism, we established 42 DPO enrichments from six different wastewater facilities throughout the San Francisco Bay area. Over 70% of our enrichments showed DPO activity, with representative enrichments from each of the six sample sites. No phosphite oxidation occurred in heat-killed controls. Genome resolved metagenomics of 11 active enrichments yielded 237 high quality MAGS, 19 of which hosted *ptx-ptd* marker genes (DPO MAGS). The DPO MAGS dominated our selective enrichments, and the taxonomy of DPO MAGS spanned six phylogenetic classes (including both gram positive and negative bacteria) and yielded 8 novel DPO hosts, 6 of which are classified as well-characterized syntrophic bacteria. Given the diversity of DPO hosts in our enrichments, we used the ptxD as a marker gene to identify DPO members in global metagenomic databases. We found that the DPO ptxD occurs worldwide, in diverse environments, and forms a distinct phylogenetic clade. Furthermore, when the ptxD from our enrichment DPO MAGS were placed into this database, we discovered that the phylogenetic relationships of the ptxD followed the same evolutionary pattern as the host taxonomy. Prior to this work, DPO was considered a rare metabolism. Our work now shows that DPO is environmentally prevalent, phylogenetically diverse, and hosted by a broad range of organisms with a propensity for syntrophic metabolism. Future work will involve comprehensive analysis of DPO MAG metabolism, including studies on energy conservation, carbon assimilation, and syntrophic partnerships within each community.

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. *Proc. Natl. Acad. Sci.* **111**, 15468–15473 (2014).
3. Schink, B. & Friedrich, M. Bacterial metabolism: Phosphite oxidation by sulphate reduction. *Nature* **406**, 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*. *J. Bacteriol.* **192**, 5237–5244 (2010).
5. Hao, L. *et al.* Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *ISME J.* (2020) doi:10.1038/s41396-019-0571-0.
6. Figueroa, I. A. *et al.* Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO_2 fixation pathway. *Proc. Natl. Acad. Sci.* 201715549 (2017) doi:10.1073/pnas.1715549114.

The Lineages of Dissimilatory Phosphite Oxidation Genes Indicate an Ancient, Vertically Transferred Metabolism

Sophia Ewens[1, 2], Alexa Gomberg* (agomberg@berkeley.edu)[1], Tyler P. Barnum[1], and **John D. Coates[1, 2]**

[1] Department of Plant and Microbial Biology, University of California, Berkeley
[2] Energy & Biosciences Institute, University of California, Berkeley

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 2.

Phosphorus (P) is essential for all life, and predominantly exists on Earth as oxidized phosphate (P^{+5}). The majority of this is trapped in mineral deposits, resulting in biological P limitation for most ecosystems. Global P cycle models consistently overlook alternative P redox states, yet reduced P species have been identified in diverse environments and frequently serve as a biological P source when soluble orthophosphate is limited[1]. In 2000, Schink et al. identified a novel microbial metabolism in which phosphite was used as a chemolithotrophic electron donor in cellular energy metabolism[2]. This metabolism, denoted dissimilatory phosphite oxidation (DPO), was first identified in *Desulfotignum phosphitoxidans* FiPS-3, where DPO activity was attributed to the *ptx-ptd* gene cluster. The *ptx-ptd* gene cluster of FiPS-3 is composed of seven genes (ptxDE-ptdCFGHI), whose functions have been assigned based on proteomics, heterologous gene expression, and homology assignments[3]. However, the biochemical mechanism and evolution of DPO is still poorly understood. In FiPS-3, the *ptx-ptd* gene cluster occurs on a genomic island, and its closest relatives are incapable of DPO, suggesting that this metabolism was acquired via horizontal gene transfer. However, this gene cluster has since been identified in two instances of *Ca. Phosphitivorax anaerolimi*: strain Phox-21, which was identified in an active DPO enrichment metagenome; and strain F81, which was identified in a full scale anaerobic digester metagenome[4,5]. Both strains of *Ca. P. anaerolimi* only harbor six of the seven *ptx-ptd* genes, and neither strain shows classical signatures of horizontal gene transfer. Furthermore, *Ca. P. anaerolimi* is in a distinct taxonomic clade from FiPS-3, with no characterized close relatives. DPO is the most energetically favorable chemotrophic electron donating process known to date, yet *D. phosphitoxidans* and *Ca. P. anaerolimi* are the only

identified representatives, thereby limiting our understanding of the diversity and evolutionary history of DPO. In this work, we have used the ptxD as a marker gene to probe metagenomic databases for the *ptx-ptd* operon. We found that the DPO ptxD is prevalent in global metagenomes, is phylogenetically diverse, and forms a monophyletic clade with the ptxD from known DPO representatives. We evaluated the gene neighborhoods of the contigs within the DPO ptxD clade and found that the *ptx-ptd* gene cluster also showed diversity in synteny and gene inclusion. In addition to the ptxD identified in global metagenomes, Ewens et al. (unpublished) recently enriched for several new DPO representatives whose ptxD fell within the database-generated DPO clade. The host taxonomy of these enrichment representatives spanned gram-positive and gram-negative bacteria, and also corresponded with the evolutionary lineages of the ptxD, suggesting that ptxD is vertically inherited. To test whether this pattern was true for the *ptx-ptd* operon as a whole, we repeated our database search using the ptdC and ptdF as probes, and we found that the phylogenetic relationships of those genes corresponded with those of the ptxD. We also identified divergent ptxD with unique *ptx-ptd* operon structures, indicating a potentially greater diversity of ptxD than was previously recognized. This data collectively suggests that DPO is a vertically transferred metabolism, and that the genomic island found in FiPS-3 is an exception. Given the broad taxonomic diversity of DPO hosts, DPO is likely to be an ancient metabolism that dates back to the last common ancestor of gram-positive and gram-negative organisms, with present-day metabolic diversity attributed to subsequent adaptive radiation. These hypotheses are corroborated by work from Pasek et al., suggesting that early metabolisms may have relied on reduced phosphorus compounds[6]. Today's DPO microorganisms may therefore be the fossils of an ancient metabolism that has otherwise been lost since the oxidation of Earth's phosphite following the great oxygenation event[7].

References

1. Pasek, M. A., Sampson, J. M. & Atlas, Z. Redox chemistry in the phosphorus biogeochemical cycle. *Proc. Natl. Acad. Sci.* **111**, 15468–15473 (2014).
2. Schink, B. & Friedrich, M. Bacterial metabolism: Phosphite oxidation by sulphate reduction. *Nature* **406**, 37 (2000).
3. 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. *J. Bacteriol.* **192**, 5237–5244 (2010).
4. Figueroa, I. A. *et al.* Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO_2 fixation pathway. *Proc. Natl. Acad. Sci.* 201715549 (2017) doi:10.1073/pnas.1715549114.
5. Hao, L. *et al.* Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *ISME J.* (2020) doi:10.1038/s41396-019-0571-0.
6. Pasek, M. A role for phosphorus redox in emerging and modern biochemistry. *Curr. Opin. Chem. Biol.* **49**, 53–58 (2019).
7. Pasek, M. A., Gull, M. & Herschy, B. Phosphorylation on the early earth. *Chem. Geol.* **475**, 149–170 (2017).

Discovery and characterization of disease resistance loci using a unique gene copy number variant population

Héloïse Bastiaanse[1,2*], Maria Marlin[4], Abigail Ferson[4], George Newcombe[4], Randall Rousseau[5], Isabelle Henry[3], Andrew Groover[1,2], **Luca Comai***[2,3](lcomai@ucdavis.edu)

[1] US Forest Service, Pacific Southwest Research Station, Davis CA; [2] UC Davis, Department of Plant Biology; [3] UC Davis, Genome Center; [4] University of Idaho, Department of Forest, Rangeland and Fire Science; [5] Mississippi State University, Department of Forestry

http://comailab.genomecenter.ucdavis.edu/index.php/Poplar

1. Identification of genomic regions that control response to two damaging poplar diseases, leaf spot and leaf rust. In field trials in two US locations and in controlled infection experiments, genome-wide scans will identify poplar loci underlying disease susceptibility.

2. Characterization cell functions governing the response of poplar to leaf pathogens. We will integrate gene expression studies with developmental and microbial responses to identify mechanisms that contribute to disease resistance.

3. Identification of genes that regulate these responses and whose manipulation could result in sustainable, long-term tolerance to the target pathogens. Candidate genes will be evaluated based on their molecular and cellular properties. Individual candidate genes will be tested using transgenic approaches and genome editing.

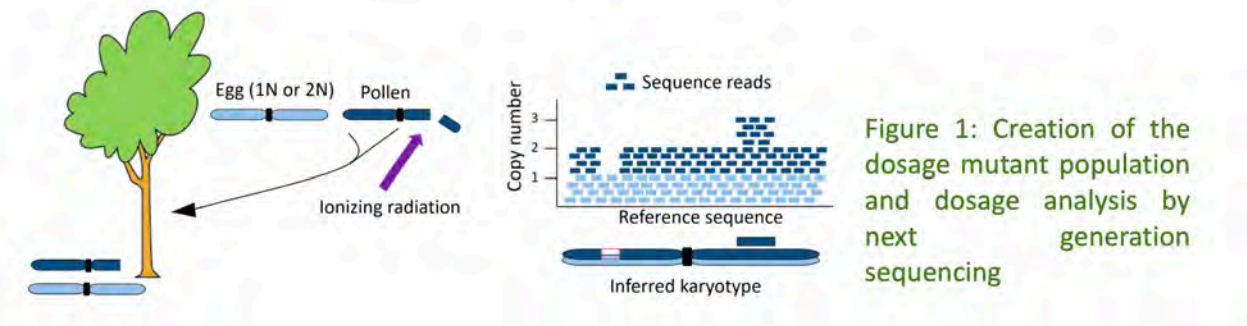


Figure 1: Creation of the dosage mutant population and dosage analysis by next generation sequencing

Pathogenic fungi that colonize poplar leaves and stems reduce yield and can cause failure of industrial bioenergy plantations. Despite extensive study of poplar pathosystems, the genetic control of poplar resistance to pathogens is still poorly understood, underscoring the need for new approaches. We developed a unique functional genomics resource based on gene dosage variation in an elite biomass poplar hybrid (1). We pollinated *Populus deltoides* with gamma irradiated *P. nigra* pollen to produce ~ 800 F1 seedlings. These contain large-scale deletions and insertions that together tile each chromosome multiple times. This resource, developed through previous funding from USDA-DOE Plant Feedstock Genomics for Bioenergy Program, has been used to define loci affecting phenology and biomass (2), and, more recently, leaf shape. Under

natural infection in the field, as well as under controlled inoculations in the greenhouse, we observed a wide variation in disease resistance within our population and were able to identify dosage QTLs influencing resistance of poplar to some of its most important fungal diseases: leaf rust and leaf spot (Melampsora sp., Septoria sp. and *Sphaerulina musiva*). Next, time-course analysis of gene expression during progression of disease symptoms will be performed for selected genotypes and used to develop predictive models of transcriptional networks underlying disease susceptibility. A final set of experiments will aim to identify candidate genes for functional analysis by manipulation using CRISPR-Cas9. Such dosage-sensitive candidate genes with significant effects on disease resistance phenotypes could be exploited in breeding programs through the selection of germplasm with naturally-occurring allelic variation or indels/copy number variation covering resistance loci.

References

1. Henry IM, Zinkgraf MS, Groover AT, Comai L. A System for Dosage-Based Functional Genomics in Poplar. Plant Cell [Internet]. 27(9), 2370–2383 (2015). http://dx.doi.org/10.1105/tpc.15.00349
2. Bastiaanse H, Zinkgraf M, Canning C, et al. A comprehensive genomic scan reveals gene dosage balance impacts on quantitative traits in Populus trees. Proc. Natl. Acad. Sci. U. S. A. (2019). http://dx.doi.org/10.1073/pnas.1903229116..

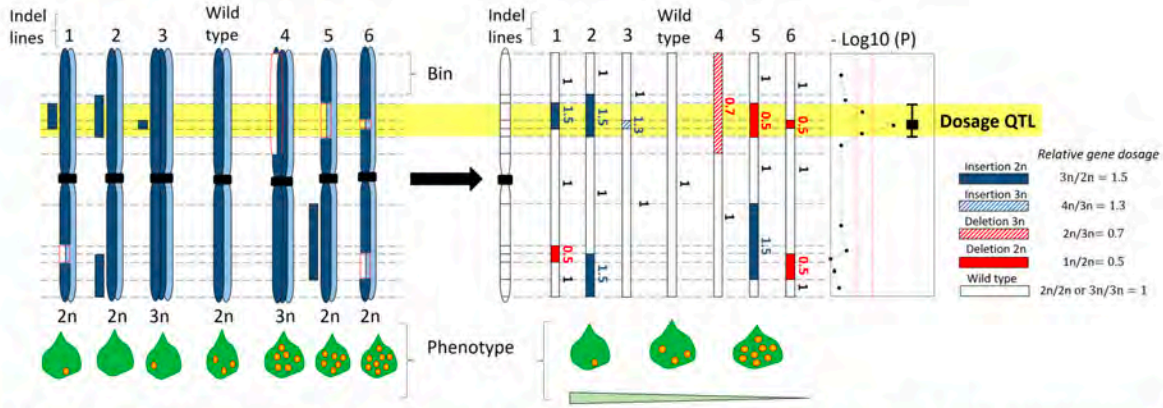


Figure 3: Left: chromosomal bins are defined by the breakpoints of the indels tiled to each chromosome. Right: Genotypes within each bin are assigned a relative dosage value reflecting both the background ploidy level and indel type. Correlations among relative dosage value and phenotypes can then be calculated.

Funding statement.

This project was supported by National Institute of Food and Agriculture, U.S. Department of Agriculture, under proposal number 2017-05632, through the USDA-DOE Plant Feedstock Genomics for Bioenergy Program

EvoNet: A Phylogenomic and Systems Biology approach to identify genes underlying plant survival in marginal, low-Nitrogen soils

Gil Eshel, [1*] (ge30@nyu.edu), Kranthi Varala,[2] Rodrigo Gutiérrez,[3] Dennis Stevenson,[4] Robert DeSalle,[5] W. Richard McCombie,[6] Jean-Michel Ané,[7] Heidi Kaeppler[7] and **Gloria Coruzzi**[1]

[1]New York University, New York; [2]Purdue University, West Lafayette; [3]Pontifica Universidad Católica de Chile, Santiago, Chile; [4]New York Botanical Garden, New York; [5]American Museum of Natural History, New York; [6] Cold Spring Harbor Laboratory, New York and [7]University of Wisconsin, Madison

http://evonet.org

Project Goals: This DOE BER sustainability project aims to identify the key genes and gene regulatory networks that enable "extreme survivor" plants to adapt and grow in marginal, nitrogen (N)-poor soils in the hyper-arid Atacama Desert in Chile. These "extreme survivor" species cover the main branches in flowering plants and include 7 grass species of particular interest for biofuels. We focus on 28 "extreme survivor" Atacama species and compare their encoded genes to Californian "sister" species that live in a N-replete conditions in arid (27 species) or mesic (27 species) environments. Deep RNA-sequencing of these "triplet species" was used to fuel a phylogenomic analysis that can identify genes that support the evolutionary divergence of the extreme survivors in Atacama Desert from their sister species in California. The genes thus identified will help to discover the mechanisms underlying physiological and developmental processes that allow plant survival in nitrogen-poor, dry soils. The genes and network modules so uncovered can potentially be translated to biofuel crops to greatly increase biomass and nitrogen use efficiency in marginal, low-fertility soils.

This collaborative project exploits the genomes of "extreme survivor" plants adapted to thrive in marginal, Nitrogen (N) poor soils in the hyper-arid Atacama Desert in the Chilean Andes. It uses a previously validated phylogenomic pipeline we developed called PhyloGeneious [1], which can identify genes that provide support to species divergence. By applying this phylogenomic pipeline to the gene sequences of "triplet species", we can identify the genes that distinguish these "extreme survivors" in Atacama from their related "sister" species adapted to similarly dry regions in California (CA) not constrained by N and/or water availability and to mesic "sister" species growing in N and water (W) replete conditions. These "extreme survivor" species from the Atacama desert broadly cover the main branches in flowering plants, and therefore offer a wide range of genomic backgrounds within which the survival traits repeatedly arose i.e., multiple independent origins of trait.

To maximize our ability to separate the trait-relevant signature from overall speciation events, our "triplet species" sampling covers multiple independent origins of the low-N adaptive trait. In published studies, we showed that our phylogenomic pipeline could; i) identify genes that underlie convergent evolution of antioxidant synthesis in Rosids in a study of 150 plant genomes [1]; and ii) identify 100+ genes associated with the loss of Arbuscular Mycorrhizal (AM) symbiosis in the *Brassicaceae* [2]. We further extend this phylogenomic approach to the study of "extreme survivor strategies" as follows:

Aim 1. **Species collection and deep transcriptome sequencing**: (NYU, NYBG, Chile). ***Progress:*** We sequenced all 28 species collected in the Atacama Desert and 50 of their CA "sister" species, with a resulting average gene coverage of 87%, based on the BUSCO single-copy orthologs assessment.

Aim 2. **Phylogenomic Analysis**: Perform phylogenomic analysis of 82 "triplet species" to identify genes that repeatedly support nodes that distinguish the "extreme survivors" in the Atacama Desert from their sister species in CA (AMNH, NYU). ***Progress:*** We performed phylogenomic analysis that includes each of the major plant lineages (*Poaceae* - 13 taxa; *Caryophyllales* – 10 taxa; *Lamiids* – 12 taxa; *Campanulids* – 17 taxa; *Fabaceae* – 12 taxa). This analysis identified 1022 genes, with a strong exclusive signal of positive selection in multiple Atacama "extreme survivors" species.

Aim 3. **Network Analysis**: We used the Arabidopsis gene-network tool from VirtualPlant (virtualplant.bio.nyu.edu), and reveled a sub-network of 35 positively selected genes, that are connected through a transcription factor (TF) hub, bZIP63 (Figure 1). This transcription factor is known to regulate primary metabolism in order to boost cellular energy in response to starvation, as part of the SnRK1-TOR signaling pathway [3]. We are currently using the *TARGET* system, developed in our lab, to validate the targets of bZIP63, to complement our in planta studies in Aim 4.

Aim 4. **Functional Validation**: To functionally validate top-ranked candidate genes for low-N adaptation in Arabidopsis and Brachypodium (NYU, Chile, U Wisconsin). ***Progress:*** We have begun to transform Brachypodium with the 28 most promising candidates from our phylogenetic analysis using our Atacama set and their closest sequence available sister species.

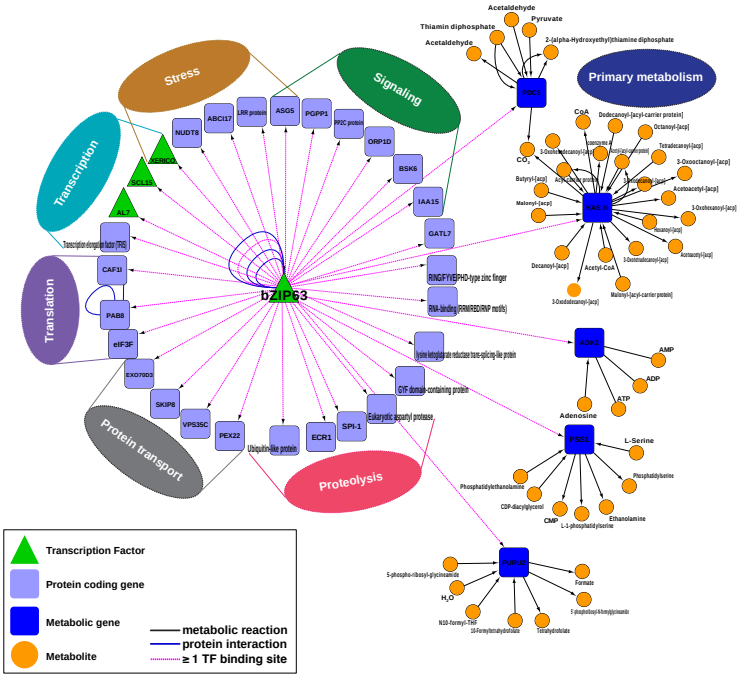


Figure 1. Sub-network of 35 positively selected genes found exclusively in multiple Atacama "extreme survivors" species. *Based on the Arabidopsis orthologous genes, we found a set of positively selected genes in Atacama that are connected through the transcription factor bZIP63, known to regulate primary metabolism in order to boost cellular energy in response to starvation, as part of the SnRK1-TOR signaling pathway [3].*

References

1. Lee E et al. (2011) A functional phylogenomics view of the seed plants. PLoS Genet 7(12):e1002411.
2. Delaux PM et al. (2014) Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. PloS Genet 10(7):e1004487.
3. Mair A et al. (2015) SnRK1-triggered switch of bZIP63 dimerization mediates the low-energy response in plants. eLife. 4:e05828.

The EvoNet grant is supported by the Genomics Science program within the Office of Biological and Environment Research in the DOE Office of Science.

Genome-wide association analysis of anthracnose resistance response in the NPGS sweet sorghum collection

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:
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 sorghum accessions on chromosomes 5 and 9, however, these resistance sources are in grain sorghum germplasm. **The identification of novel anthracnose resistance sources present in sweet sorghum germplasm will expedite the development of new resistant sweet sorghum cultivars and hybrids by avoiding time-consuming introgression breeding approaches with non-sweet sorghums serving as donor of the resistance alleles**.

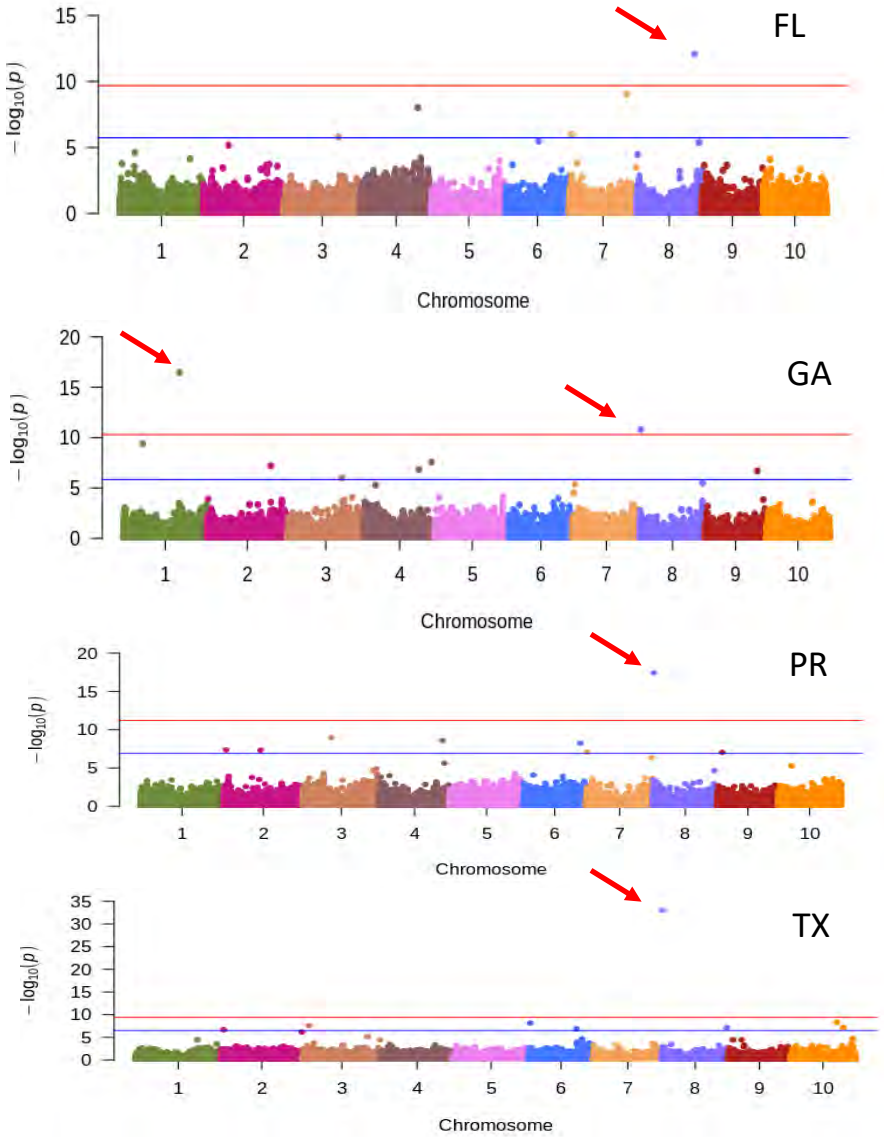
The sweet sorghum collection of the USDA-ARS National Plant Germplasm Systems (NPGS) is the primary source of genetic diversity for the development of new sweet sorghum varieties for biofuel production. Recently, we established a sweet sorghum diversity panel (SWDP) of 233 accessions that includes advanced breeding materials and represent ~15% of the NPGS sweet sorghum collection. A two-year replicated field trial of the SWDP for anthracnose resistance response in Texas, Georgia, Florida and Puerto Rico identified 29 accessions resistant across locations, while another 145 accessions showed variable resistance response against location pathotypes. Phylogenetic analysis among resistant accessions suggests that the SWDP contains multiple sources of resistance. However, it is yet unknown if these resistance sources are identical to those identified in grain sorghum germplasm.

A genome-wide association study (GWAS) based on 157,843 single-nucleotide polymorphisms (SNPs) and employing the fixed and random model Circulating Probability (farmCPU) analysis identified three genomic regions associated with anthracnose resistance (Figure 1). The distal genomic region of chromosome 8 was associated with resistance response observed in Florida, and an associated SNP was located within an *R*-gene. Genomic regions on chromosomes 1 and 8 were associated with resistance responses observed in Georgia. Candidate gene analysis identified a cluster containing genes involved in salt stress/antifungal activity on chromosome 1 and an *R*-gene on chromosome 8. The candidate genes at both loci were in linkage disequilibrium and less than 15 kb downstream of the associated SNP. Resistance response in Texas and Puerto Rico were associated with a genomic region on chromosome 8 that includes an *R*-gene. Candidate *R*-genes among Puerto Rico, Texas and Georgia were 18 kb apart suggesting that each gene might be specific for each location's pathotype.

The results of this study indicate that the NPGS sweet sorghum germplasm collection contains multiple anthracnose resistance sources. A genomic scan revealed these resistance sources are

different from those known in grain sorghum germplasm. It is imperative that these resistance sources be used effectively to increase the durability of anthracnose resistance of new biofuel sorghum varieties. When the goal is to produce a particular cultivar across a large geographic area, multiple sources of resistance effective against the local pathotypes need to be combined.

Figure 1. Genome-wide association analysis for anthracnose resistance response among 233 NPGS sweet sorghum accessions evaluated in Florida (FL), Georgia (GA), Puerto Rico (PR), and Texas (TX) during 2017 and 2018.



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

Characterizing growth and metabolism of *Rhodococcus opacus* PD630 on real lignin breakdown products

Garrett Roell[1*](garrettroell@wustl.edu), Rhiannon Carr[1], Anirudh Krishnamurthy[1], Zeyu Shang[1], Young-Mo Kim[2], **Gautam Dantas**[3,4,5,6], **Marcus Foston**[1], **Tae Seok Moon**[1], and **Yinjie Tang**[1]

[1] Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, St. Louis, MO, 63130, USA; [2] Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99354, USA; [3] The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63110, USA; [4] Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108, USA; [5] Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, 63130, USA; [6] Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108, USA

Project Goals: The goal of this project is to determine the mechanisms by which *Rhodococcus opacus* PD630 is able to convert depolymerized lignin to lipid-based biofuels. Our previous work has examined how *R. opacus* metabolizes several aromatic model compounds, but it does not extend to how this bacterium is able to tolerate and utilize depolymerized lignin as a sole carbon source. The specific aim of this work is to tailor the depolymerization reaction conditions and catalysts to generate lignin breakdown products suited to the metabolism of *R. opacus*.

R. opacus is a soil bacterium that is naturally able to tolerate and utilize aromatic substrates as sole carbon sources, as well as accumulate high titers of lipids [1, 2]. To date, published studies have focused on how *R. opacus* is able to tolerate and consume model aromatic compounds (phenol, benzoate, etc.), and have provided many novel insights into this species' aromatic and lipid metabolism [3-6]. However, there is a gap in knowledge about how *R. opacus* consumes real-world lignin breakdown products (LBP), as well as how the specific thermocatalytic process conditions and catalysts used to produce LBP affect *R. opacus* metabolism. Our team has employed diverse lignin chemical conversion processes, cell cultivation, GC-MS analysis of LBP consumptions, ^{13}C-based dynamic labeling of intracellular metabolites, and metabolite pool size quantifications, which offer rich insights into LBP metabolism in *R. opacus*.

First, we characterized LBP composition and its effect on *R. opacus* growth. We found that the LBP composition is highly dependent on the biomass source (e.g., switchgrass or poplar), reaction conditions (e.g., solvent, temperature, and residence time), and the catalyst (e.g., palladium or nickel on activated carbon). Our palladium catalyst produced more aromatic compounds than the nickel catalyst did. *R. opacus* was able to consume LBP generated with either catalyst as the sole carbon source, although cultures reached higher cell density when grown using LBP from the palladium catalyst.

Second, we developed a convenient method for gas chromatography–mass spectrometry to analyze and compare the compounds present in LBP cultures. This method can identify which compound *R. opacus* is able to consume. By GC-MS analyses of culture samples with different LBP compositions, we found that the major fractions of LBP consumed in all *R. opacus* culture samples were a mixture of ketone, furfural, and phenolic compounds.

Third, we measured and compared the intracellular concentrations of ~20 central carbon metabolites when *R. opacus* was grown with glucose, benzoate, phenol, and three defined aromatic mixtures. We found that metabolic concentrations were highly variable based on the carbon substrates present. This finding shows that *R. opacus* has a flexible metabolism, and demonstrates how changes in LBP composition can affect central metabolism.

Fourth, we measured intracellular metabolite concentration from cells growing with LBP, and have carried out isotopic pulse-trace experiments with LBP. LBP-fed cells had higher intracellular sugar phosphate concentrations than glucose-fed *E. coli* did, and their intracellular TCA cycle metabolite concentrations were ~50% of those of *E. coli*, suggesting that *R. opacus* is able to co-consume the sugar and aromatic fractions of LBP. Isotopic pulse-trace experiments have indicated that there is a significant lag phase (~3 hours) for the uptake of LBP after medium switching from phenol to LBP, likely because the cells need lag phase to sense the LBP and to produce the appropriate transporters and degradation enzymes.

Future work will focus on optimizing the thermocatalytic reaction parameters to produce LBP more compatible with the metabolic capabilities of *R. opacus*. Additionally, transcriptomic analysis will be performed to establish how accurately growth on the model compounds simulates growth on real LBP. The findings from this work improve our understanding of *R. opacus*' metabolism for future lignin valorization, and they will inform development of a machine learning-based predictive model.

Publications (+ Corresponding author)

1. A Chatterjee, DM DeLorenzo, R Carr and TS Moon^{+}. Bioconversion of renewable feedstocks by *Rhodococcus opacus*. *Curr. Opin. Biotechnol.* 64, 10-16. (2020)

2. 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)

3. GW Roell, RR Carr, T Campbell, Z Shang, WR Henson, JJ Czajka, HG Martín, F Zhang, M Foston, G Dantas, TS Moon^{+}, YJ Tang^{+}. A concerted systems biology analysis of phenol metabolism in *Rhodococcus opacus* PD630. *Metab. Eng.* 55, 120–130 (2019)

4. 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)

5. 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)

6. 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)

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.

Elucidating aromatic tolerance and utilization in adaptively evolved *Rhodococcus opacus* strains for lignin valorization

Winston Anthony[1]* (winstoneanthony@wustl.edu), Rhiannon Carr[2], Garrett Roell[2], **Yinjie Tang**[2], Joonhoon Kim[7,9], Jose Martin Martin[6], Hector Garcia Martin[6,7,8,9], Marcus Foston[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] Pacific Northwest National Laboratory, Richland, WA, USA

Project Goals: We aim to combine adaptive evolution and multiple–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.

The oleaginous microbe *R. opacus* is naturally tolerant to aromatic compounds commonly found in lignin-derived mixtures. We have demonstrated the potential of *R. opacus* for increased survivability in higher concentrations of aromatic compounds 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 – which were not permissive to WT growth – 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 phenol to rigorously measure and compare how aromatic substrates were consumed [3, 4].

We found similar transcriptional profiles in the established beta-ketoadipate pathway and aromatic gene clusters for mutant and wild type strains grown on the same carbon source. Instead of wild type and mutant strains differing in aromatic transcription profiles, the carbon source seems to drive differences in expression. An amino acid labeling experiment performed with phenol supported this finding by showing only minor differences between WT and mutant strains, suggesting that enhanced aromatic tolerance in adapted strains might be the product of not only

increased rates of aromatic utilization but also other unknown factors. Further work is currently underway to characterize and define unknown mechanisms of aromatic tolerance and utilization through analysis of differentially expressed transcripts, use of machine learning-based transcript-to-flux prediction models, and recently developed synthetic biology tools [5-8]. 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 in which these strains are fully characterized.

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)

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.

Pectin – lignin interactions in plant cell walls and model composites

Riddhi Shah[1], Hugh O'Neill[2]* (oneillhm@ornl.gov), Yunqiao Pu[2], Ajaya Biswal[3], Sai Venkatesh Pingali[2], Arthur J. Ragauskas[2], Debra Mohnen[3], and **Brian H. Davison[2]**

[1]University of Tennessee, Knoxville, Tennessee; [2]Oak Ridge National Laboratory, Oak Ridge Tennessee; [3]University of Georgia, Athens, Georgia.

http://cmb.ornl.gov/index.php/research/bioenergy/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.

Lignin – carbohydrate complexes (LCCs) are hypothesized to form through interactions of lignin with polysaccharides such as pectin and hemicellulose in the plant cell wall. Here, we report on putative LCCs formed between lignin and the pectin homogalacturonan (HG). A recent study showed that HG is found in the middle lamella and cell wall corner region that is also reported to be the location where lignification is initiated.[1,2] In addition, a transgenic switchgrass with reduced pectin, formed by suppressing galacturonic acid synthesis (GAUT4-kd), was shown to have reduced recalcitrance and released higher sugar compared to the wild-type (WT) switchgrass.[3] These studies provided the motivation to investigate the potential for LCCs formed due to interactions between lignin and pectin.

We compared the structural properties of WT and GAUT4-kd mutant switchgrass using small-angle neutron scattering (SANS). No differences were found in either the cellulose microfibril structure or the arrangement of the cell wall matrix copolymers in the native plants indicating that this particular mutation has negligible effects on the structure of the cell wall. However, after hot water pretreatment the GAUT4-kd variant formed significantly more lignin aggregates that were greater in size compared to wild-type switchgrass. Our data shows that despite being subjected to the same pretreatment conditions, more lignin is re-distributed to form aggregates in GAUT-kd and suggests that interactions between lignin and HG could decrease lignin aggregation in switchgrass. To further investigate lignin – HG interactions, we synthesized a model composite by polymerizing coniferyl alcohol to form a lignin-like polymer (i.e., dehydrogenation polymer (DHP)) in the presence of HG. Small-angle X-ray scattering (SAXS) of the composite showed a network-like structure unlike that obtained from a physical mixture of the individual polymers. Fourier transform infrared spectroscopy showed a unique absorption band in the ester region (~1730 cm^{-1})[4] that was only present in the composites and not in the HG, DHP or physical mixture

of HG and DHP. Furthermore, solid state nuclear magnetic resonance (SSNMR) analysis provides dynamic evidence of interaction between the synthetic lignin polymer and homogalacturonan.

Overall, studies of intact mutant and wild-type switchgrass and model composite materials provide evidence for formation of a lignin-polysaccharide complex. Interactions of polysaccharides with lignin, and the role polysaccharides play in changing lignin morphology during pretreatment, are critical to understand to improve pretreatment regimes and for conversion of these plant cell wall polymers to bioproducts.

References

1. DeMartini, J. D et al, Application of monoclonal antibodies to investigate plant cell wall deconstruction for biofuels production. *Energy & Environmental Science* 2011, *4* (10), 4332-4339.
2. Li, Y. et al, Optically transparent wood from a nanoporous cellulosic template: combining functional and structural performance. Biomacromolecules 2016, 17 (4), 1358-1364.
3. Biswal, A. K et al, Sugar release and growth of biofuel crops are improved by downregulation of pectin biosynthesis. 2018, 36 (3), 249-257.
4. Haitao Yang et al, Elucidation of the bonds between cellulose and dehydrogenation polymer with carbon-13 isotopic tracer method. Cellulose Chemistry and Technology 2013, 47 (3-4), 143-151.

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 ERKP752.

Polymer and structural science behind valorizing lignin using solvents

Marcella C. Berg,[1] Derya Vural,[1] Samarthya Bhagia[1], Luna Liang[2], Zhi Yang[1], Xianzhi Meng[2], Nidia Gallego[1], Nathan Bryant,[2] Sai Venkatesh Pingali[1], Hugh M. O'Neill,[1] Yunqiao Pu,[1] Eugene Mamontov,[1] Arthur J. Ragauskas,[1,2] Jeremy C. Smith,[1] Loukas Petridis[1]*(petridisl@ornl.gov), **Brian H. Davison**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]University of Tennessee, Knoxville

http://cmb.ornl.gov/index.php/research/bioenergy/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.

Converting plant biomass to renewable chemicals, fuels and materials can have both economic and environmental impact: sustainable economic growth and reducing the carbon footprint from the industries. The plant polymer lignin is especially well suited as a precursor for valuable renewable biomaterials because of its abundance and high content of carbon. However, the rigidity of lignin makes it the most recalcitrant component of plant biomass. To facilitate deconstruction and conversion of biomass, it is therefore necessary to soften lignin. At an atomic level softening the lignin means increasing lignin atomic fluctuations, making the molecules more dynamic. This is usually done by using high temperatures[1] or pretreatment solvents. Elucidating the motions that give rise to enhanced lignin dynamics may yield fundamental insights that enable rational improvement of biomass pretreatment and processing. However, knowledge about these changes in the dynamics of lignin that arise from the solvent-induced effects is lacking to this day. This study aims to understand the underlying processes that cause the dynamical increase of lignin motion in tetrahydrofuran (THF) solvent, a novel promising multifunctional solvent for biomass pretreatment and fractionation.

Comparative studies of THF and the organic solvents ethanol and gamma-valerolactone (GVL) for extraction of lignin from switchgrass strains found differences in conservation of structural properties dependent on both solvent and lignin characteristics. We studied wild-type Alamo and two transgenic switchgrass strains, one with altered lignin composition (downregulation of caffeic acid 3-O-methyltransferase (COMT) gene) and one with decreased lignin content (overexpression of a lignin biosynthetic repressor MYB transcription factor) that reduce recalcitrance of switchgrass. The two-dimensional ^{13}C-^{1}H heteronuclear single quantum coherence (HSQC) nuclear magnetic resonance (NMR) was carried out on solubilized lignin to obtain structural information on lignin structures. Chemical and physical characterization showed that ethanol pretreatment preferentially cleaved beta aryl ether linkages, while S/G ratio was significantly lower in THF and GVL lignins obtained from COMT and MYB switchgrass, and p-coumarate of THF and GVL lignins was much higher in all three switchgrass samples.

In order to elucidate lignin molecular motions, we employed both experimental and computational methods. Quasi-elastic neutron scattering (QENS) experiments provided direct information on molecular dynamics and geometry of motions in a non-invasive manner that does not unduly perturb the material. The high incoherent scattering cross section of the hydrogen atoms makes the QENS ideal for extracting self-

correlations (dynamics) in materials such as lignin. Molecular Dynamics (MD) simulations access a broad range of time scales (see Figure 1) and provides a full atomistic model of the system. Additionally, MD probes similar length- and time-scales as QENS, so that a theoretical neutron scattering profile can be calculated from the atomic positions in the simulation trajectories, thus results from MD and QENS can be directly compared. For the QENS experiment lignin was extracted and purified from Poplar feedstock corresponding to a molecular weight of 4514 g/mol with composition ratios of S ~ 58%, G~ 30% and H ~ 12 %[2]. The lignin molecules and their solvation concentrations for the MD simulation were modelled to replicate the experimental samples.

Both techniques show that in the presence of the solvent there is a dynamical transition at which lignin purified from polar becomes more dynamic. This dynamical transition is found even at small solvent concentrations and temperatures as low as -20 °C (see Figure 2). We furthermore investigate how THF solvent concentrations affect the atomic fluctuations of the lignin. In conclusion, the study provides a molecular-level understanding of the technologically important motions of lignin that will facilitate the rational improvement of biomass deconstruction and lignin processing.

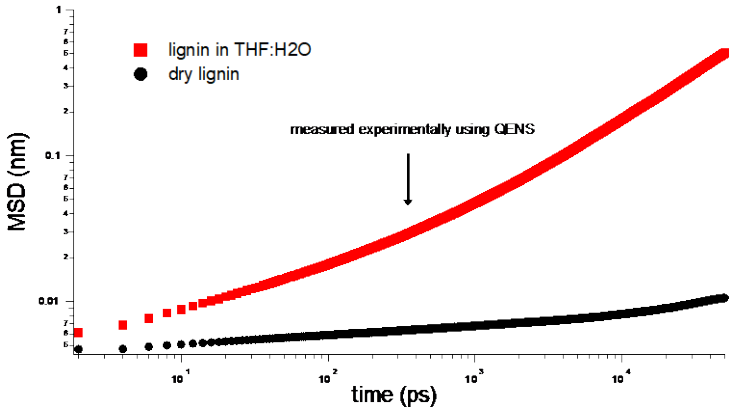


Figure 1: Simulated mean square displacement (MSD) for both dry and solvated lignin at 25 °C. MSD quantifies the amplitude of atomic fluctuations and is a measure of lignin softness. In the simulations, MSD is calculated over a broad time-range. The nano-second timescale, indicated by an arrow, is directly compatible with the experimentally obtained QENS spectra.

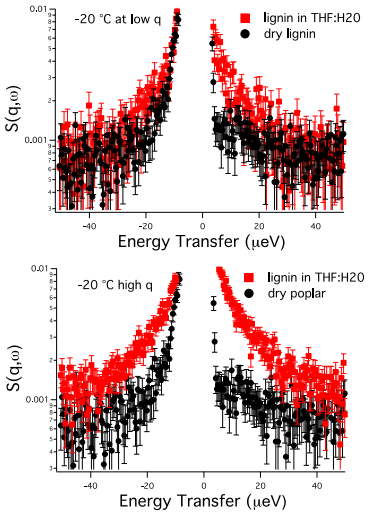


Figure 2: Experimentally measured QENS signal at -20 °C for dry and solvated lignin. The broadening of the signal indicates mobility and is a measure of lignin softness.

References

1. Derya Vural, Catalin Gainaru, Hugh O'Neill, Yunqiao Pu, Micholas Dean Smith, Sai Venkatesh Pingali, Eugene Mamontov, Brian H. Davison, Alexei P. Sokolov, Arthur J Ragauskas, Jeremy C. Smith, Loukas Petridis *Green Chem. R. Soc. Chem.* **2018 *20*, 1602-1611**
2. Xianzhi Meng, Aakash Parikh, Bhogeswararao Seemala, Rajeev Kumar, Yunqiao Pu, Phillip Christopher, Charles E. Wyman, Charles M. Cai, and Arthur J. Ragauskas *ACS Sustainable Chemistry & Engineering* **2018** *6* (7), 8711-8718

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.

Precision Labeling of Membrane Fatty Acids in *Bacillus subtilis* and the Impacts on the Cellular Proteome and Lipidome

Jonathan D. Nickels,[1] Suresh Poudel,[2] Sneha Chatterjee,[2] Abigail Farmer,[3,4] Destini Cordner,[1] Shawn R. Campagna,[3,4] Richard J. Giannone,[5] Robert L. Hettich,[5] Dean A. A. Myles,[6] Robert F. Standaert,[7] John Katsaras,[6,8,9] and James G. Elkins[2,10*] (elkinsjg@ornl.gov), and **Brian H. Davison[2]**

[1]Department of Chemical and Environmental Engineering, University of Cincinnati, Cincinnati, OH; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [3]Department of Chemistry, University of Tennessee, Knoxville, Knoxville, TN; [4]Biological and Small Molecule Mass Spectrometry Core, University of Tennessee, Knoxville, Knoxville, TN; [5]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [6]Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN; [7]Department of Chemistry, East Tennessee State University, Johnson City, TN; [8]Shull Wollan Center – a Joint Institute for Neutron Sciences, Oak Ridge National Laboratory, Oak Ridge, TN; [9]Department of Physics and Astronomy, University of Tennessee, Knoxville, Knoxville, TN; [10]Department of Microbiology, University of Tennessee, Knoxville, Knoxville, TN

http://cmb.ornl.gov/index.php/research/bioenergy/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.

In order to perform structural studies on the disruption of biomembranes additional tools to selectively manipulate biomembranes *in vivo* are needed. We previously developed a system to control both the chemical and isotopic composition of the hydrophobic region within the cellular membrane of the model bacterium, *Bacillus subtilis*.[1] Briefly, blocking *de novo* fatty acid (FA) biosynthesis using cerulenin as an irreversible inhibitor of beta-ketoacyl-ACP synthase II (FabF), in a strain that is unable to catabolize FAs (Δ*fadN*), forces growing cells to take-up exogenous FAs from the growth medium to construct their membrane. This feeding strategy has enabled the biophysical properties of biomembranes in viable bacterial cells to be studied using neutrons as a

non-destructive probe. If cells are grown in a high D_2O background while being fed protiated FAs, the membrane can be "visualized" using small-angle neutron scattering (SANS) techniques due to the differences in scattering length density between hydrogen and deuterium. By tuning isotopic contrast within the plane of the membrane, we have also shown that lateral structure exists in viable bacterial cells consistent with the presence of lipid domains (aka lipid rafts). Having demonstrated the potential of our approach to study bacterial membranes using SANS techniques *in vivo*, we now aim to expand the utility of our methods by determining how microbial membranes respond to second-generation biofuels and plant biomass pretreatment solvents. Amphiphilic solvents partition into biomembranes affecting their nanoscale structure and ability to form an effective barrier to the extracellular environment. A better understanding of how membranes respond to different solvents at the molecular scale will improve molecular dynamics simulations and potentially, the ability to engineer more resilient biomembranes.

While our strategy to introduce neutron contrast in the bilayer by feeding exogenous FAs has been effective, it is necessary to understand the systemic changes in *B. subtilis* cells induced by the labeling procedure itself. In this work, analysis of cellular membrane compositions was paired with shotgun proteomics to assess how the proteome changes in response to the directed incorporation of exogenous FAs into the membrane of *Bacillus subtilis*. Key findings from this analysis include an alteration in lipid headgroup distribution, with an increase in phosphatidylglycerol lipids and decrease in phosphatidylethanolamine lipids, possibly providing a fluidizing effect on the cell membrane in response to the induced change in membrane composition. Changes in the abundance of enzymes involved in FA biosynthesis and degradation are observed; along with changes in abundance of cell wall enzymes and isoprenoid lipid production. The observed changes may influence membrane organization, and indeed the well-known lipid raft-associated protein flotillin was found to be substantially down-regulated in the labeled cells – as was the actin-like protein MreB. Taken as a whole, this study provides a greater depth of understanding for this important cell membrane experimental platform and presents a number of new connections to be explored in regard to modulating cell membrane FA composition and its effects on lipid headgroup and raft/cytoskeletal associated proteins.

References

1. Nickels JD, Chatterjee S, Stanley CB, Qian S, Cheng X, Myles DAA, Standaert RF, Elkins JG, Katsaras J (2017) The in vivo structure of biological membranes and evidence for lipid domains. PLoS Biol 15(5): e2002214. https://doi.org/10.1371/journal.pbio.2002214

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 ERKP752.

Drivers and Mechanisms of Long-Term Soil Response to Chronic Warming

Luiz A. Domeignoz-Horta[1] (ldomeignozho@umass.edu), Grace Pold[2,3], Serita D. Frey[4], Jerry M. Melillo[5], Kristen M. DeAngelis[1*]

[1]University of Massachusetts, Amherst, MA; [2]Hampshire College, Amherst MA; [3]California Polytechnic State University, San Luis Obispo CA; [4]University of New Hampshire, Durham NH; [5]Marine Biological Laboratory, Woods Hole MA

Project Goals: The goal of this research is to quantify the effect of long-term warming on the temperature sensitivity of CUE and extracellular enzyme activity.

The predictions of how soil carbon stocks will change with chronic warming are sensitive to the assumptions made about microbial parameters such as carbon use efficiency (CUE). CUE is the fraction of carbon (C) taken up by the microbial cells and retained in biomass in relation to the fraction lost via respiration. Microbial CUE partitions the flow of C to the atmosphere, decomposer communities and potential soil C stocks. Because there are still uncertainties about what drives long- and short-term responses of CUE, more research is needed to better predict soil carbon stocks in a warming world. In this ongoing work, we sampled soils from two long-term experimental sites that have been heated for 28 or 13 years in a mid-latitude hardwood forest at the Harvard Forest LTER. We collected organic and mineral soils in summer and fall and measured the response of carbon use efficiency and extracellular enzyme activity to increasing temperature from 4°C to 30°C. Respiration, a CUE component, and extracellular enzyme activity both showed a smaller response to increasing temperature in heated than control soils. In the next steps of this work we will measure carbon quality. Our working hypothesis is that substrate quality plays a role in the adaptation of CUE to long-term warming. We expect to find depletion in the more labile soil carbon pool in heated compare to control soils which could account for the smaller CUE observed in these soils. This work will elucidate the effect of long-term warming to soil carbon pools and to microbial CUE response to temperature helping us better predict possible feedback mechanisms between soil and the atmosphere.

Publications supported by DE-SC0016590

Domeignoz-Horta, Luiz A., Kristen M. DeAngelis, and Grace Pold. (2019) "Draft genome sequence of Acidobacteria group 1 Acidipila sp. strain EB88, isolated from forest soil." *Microbiol Resour Announc* 8, no. 1: e01464-18.

Domeignoz-Horta, Luiz A., Grace Pold, Xiao-Jun Allen Liu, Serita D. Frey, Jerry M. Melillo, Kristen M. DeAngelis. (*in review*) "Microbial diversity drives carbon use efficiency in a model soil."

Li, Jianwei, Gangsheng Wang, Melanie A. Mayes, Steven D. Allison, Serita D. Frey, Zheng Shi, Xiao-Ming Hu, Yiqi Luo, and Jerry M. Melillo. (2019) "Reduced carbon use efficiency and increased microbial turnover with soil warming." *Global change biology* 25, no. 3: 900-910.

Melillo, Jerry M., Serita D. Frey, Kristen M. DeAngelis, William J. Werner, Michael J. Bernard, Francis P. Bowles, Grace Pold, Melissa A. Knorr, and A. Stuart Grandy. (2017) "Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world." *Science* 358, no. 6359: 101-105.

Pold, Grace, Erin M. Conlon, Marcel Huntemann, Manoj Pillay, Natalia Mikhailova, Dimitrios Stamatis, T. B. K. Reddy, T.B.K., Daum, C., Shapiro, N., Kyrpides, N. and Woyke, T., Kristen DeAngelis (2018) "Genome Sequence of Verrucomicrobium sp. Strain GAS474, a Novel Bacterium Isolated from Soil." *Genome Announc.* 6, no. 4: e01451-17.

Pold, Grace, Luiz A. Domeignoz-Horta, and Kristen M. DeAngelis. (2019) "Heavy and wet: evaluating the validity and implications of assumptions made when measuring growth efficiency using 18O water." *BioRxiv*: 601138.

Pold, Grace, Luiz A. Domeignoz-Horta, Eric W. Morrison, Serita D. Frey, Seeta A. Sistla, and Kristen M. DeAngelis. (2020) “Carbon use efficiency and its temperature sensitivity co-vary in soil bacteria .” *mBio* 11 (1): e02293-19. **doi:** 10.1128/mBio.02293-19

Pold, Grace, Marcel Huntemann, Manoj Pillay, Natalia Mikhailova, Dimitrios Stamatis, T. B. K. Reddy, Chris Daum, C., Shapiro, N., Kyrpides, N., Woyke, T. and DeAngelis, K.M. (2018) "Draft Genome Sequences of Three Strains of a Novel Rhizobiales Species Isolated from Forest Soil." *Genome Announc.* 6, no. 5: e01452-17.

Pold, Grace, Seeta A. Sistla, and Kristen M. DeAngelis. "Metabolic tradeoffs and heterogeneity in microbial responses to temperature determine the fate of litter carbon in simulations of a warmer world." (2019) *Biogeosciences* 16, no. 24: 4875-4888.

Funding statement. This work is supported by the DOE Genomic Sciences Program under Contract No. DE-SC0016590, and the National Science Foundation grants DEB 1237491 (Long-Term Ecological Research) and DEB 1456528 (Long-Term Research in Environmental Biology).

A role for differential Rca isoform expression in C4 bioenergy grass thermotolerance?

Sang Yeol Kim and Donald R. Ort
CARL R.WOESE INSTITUTE FOR GENOMIC BIOLOGY, UNIVERSITY OF ILLINOIS, URBANA-CHAMPAIGN

Project Goal:

Enhancement of crop production of C4 bioenergy grass by modifying posttranslational modification of Rubisco activase

ABSTRACT

Rubisco activase (Rca) facilitates the release of sugar-phosphate inhibitors at Rubisco catalytic sites during CO_2 fixation. Most plant species express two Rca isoforms, the larger Rca-α and the shorter Rca-β, either by alternative splicing from a single gene or from separate genes. The mechanism of Rubisco activation by Rca isoforms has been intensively studied in C3 plants; however, the functional role of Rca in C4 plants exposing Rca to much higher [CO_2] is less clear. In this study, we selected five C4 bioenergy grasses to investigate the role of Rca in C4 photosynthesis. All five C4 plants contained two *Rca* genes, one encoding Rca-α and the other encoding Rca-β, which are closely positioned in the genome. A variety of abiotic stress-related motifs exist in the *Rca-α* promoter of each grass, and while the *Rca-β* gene was constantly highly expressed at ambient temperature, Rca-α isoforms were expressed only at high temperature but surpassed 30% of Rca-β content. The pattern of Rca-α induction upon transition to high temperature and reduction upon return to ambient temperature was the same in all five C4 grasses but may be unique to C4 grasses. In sorghum (*Sorghum bicolor*), the induction rate of Rca-α isoforms was similar to the recovery rates of gas exchange and Rubisco activation from high temperature, which inferred a functional correlation between Rca-α isoform expression and maintenance of Rubisco activation at high temperature. Therefore, our research suggests Rca-α isoforms have a functional role in carbon fixation by supporting Rubisco activation at high temperature.

Funding: 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).

An Ecosystem-Scale Comparison of Sorghum, Maize, and Miscanthus as Three Bioenergy Crop Candidates

Caitlin E. Moore,[1]* (caitlinm@illinois.edu), Adam C. von Haden,[1] Mark Burnham,[1] Ilsa Kantola,[2] Christy D. Gibson,[1] Wendy H. Yang,[1] Evan H. DeLucia,[1,2] and **Carl J. Bernacchi**[1,3]

[1]Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, Urbana, IL; [2]Institute for Sustainability, Energy and Environment, University of Illinois at Urbana-Champaign, Urbana, IL; [3]Global Change Photosynthesis Research Unit, USDA/ARS, Urbana, IL

http://cabbi.bio/research/sustainability-theme/

Project Goals: The primary goal of our research team is to measure carbon, nitrogen, water, and energy fluxes from annual and perennial bioenergy crop systems across spatial and temporal scales. We achieve this using a suite of above and below ground measurement approaches, including eddy covariance flux towers, soil respiration chambers, and quantum cascade lasers for fluxes and regular survey sampling for pool quantification of each of these variables. Our efforts will provide a better understanding of how carbon, nitrogen, water, and energy partitioning change over time for annual and perennial crops, which is important for determining the long-term sustainability of bioenergy crop systems.

Abstract text. Perennial cellulosic crops have been the focus of bioenergy research and development due to sustainability advantages associated with their higher soil carbon storage and reduced nitrogen requirements compared to annual crops. However, perennial crops can take several years to fully establish and their sustainability benefits can be quickly negated if converted to other land uses. The development of a photoperiod sensitive, high yield energy *Sorghum bicolor* (sorghum) variety may offer an annual cellulosic crop alternative that could provide a better ecologically sustainable bioenergy source than its annual *Zea mays* (maize) counterpart and be more easily integrated into current crop rotation cycles. The University of Illinois has a rich history of using eddy covariance flux towers to quantify carbon, water, and energy fluxes over both maize and miscanthus (*Miscanthus x giganteus*) crop systems in the Midwest region of the United States. With support from CABBI, a new flux tower was installed in a sorghum field to compare against the existing maize and miscanthus records. We present results from the first growing season where all three crops were measured with the aim of characterizing ecophysiological similarities and differences in carbon, nitrogen, water, and energy flux between these three key bioenergy crop systems.

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.

Application of Spatially Adjusted Machine Learning Approaches to Improve Sorghum Biomass Prediction Using Unmanned Aerial Vehicles

Sebastian Varela[1]* (sv79@illinois.edu), **Andrew D. B. Leakey**[1,2]

[1] DOE Center for Advanced Bioenergy and Bioproducts Innovation; and [2] Department of Plant Biology, University of Illinois at Urbana Champaign, Urbana, IL

https://cabbi.bio/

Project Goals: In the context of design and engineer environmentally resilient crops with maximum yield and resource use efficiency, this project implements the use of UAV-acquired multispectral and thermal imagery to screen water use efficiency (WUE) and productivity traits in sorghum/Miscanthus field trials at scale.

Abstract Text:

Sorghum bicolor is a model C4 feedstock grass that has significant potential for improved production of biofuels and bioproducts. Phenotyping biomass production is a key element of field trials for breeding, quantitative genetics, and bioengineering that will increase productivity, resilience, sustainability, and value of crops. Traditional harvest methods for evaluating biomass production (yield) are very labor intensive and destructive, limiting the scale and speed of research. Remote sensing using cameras mounted to unmanned aerial vehicles (UAV) allows rapid, non-destructive, high-frequency data collection. Nevertheless, analytical methods are needed to exploit the high spatial and temporal resolution in images to predict final yield. Data has an explicit spatial nature that suggests that the true underlying relationship among dependent and independent variables can be spatially varying. Models that do not take this under consideration can draw limited inference. The application of Geographically Weighted Regression (GWR) as a spatially adjusted approach to improve yield prediction has been tested in wheat [1] and rice [2].

In this study, we introduce the Geographical Random Forest (GRF) concept. The principle idea is similar to that of GWR [3], in which we move to local computation (plot vicinity) rather than global (whole extent of experimental area) one, but relaxing normality assumption and non-linear dependencies between UAV-based features, and yield prediction. First, we identify the most informative features in image data from UAVs to predict yield of 870 diverse bioenergy sorghum accessions. Second, we then test GWR and GRF as local spatially adjusted approaches, and compare these spatially adjusted approaches with Principal Component Regression (PCR) as an “aspatial” global benchmark. Yield of biomass across a diverse population of accessions of biomass sorghum was predicted from UAV imagery with high precision and accuracy (r=0.93, RMSE = 3.06). The most important data descriptors predicting yield were: (1) UAV-derived height of the canopy 49 days after planting (DAP) associated with a period of fast vertical growth in mid-season; (2) canopy ground cover prior to canopy closure 43 DAP; and (3) canopy nitrogen status via Normalized Difference Red edge (NDRE) prior to canopy closure 43 DAP. PCR performance

was moderate, but with significant bias given its “aspatial” global fit. It overestimates low yield and underestimated high yield plots. GWR predictability outperforms PCR, it slightly improves accuracy and reduces bias, suggesting the benefit of a “locally” computed model. GRF outperforms GWR, suggesting the superiority of GRF as a spatially corrected, and non-linear model calibrated and computed locally rather than globally. This methodological advance will be used to advance CABBI research goals by evaluating field trials for breeding, quantitative genetics, and bioengineering that will increase productivity, resilience, sustainability, and value of crops.

1. Haghighattalab, A.; Crain, J.; Mondal, S.; Rutkoski, J.; Singh, R.P.; Poland, J. Application of geographically weighted regression to improve grain yield prediction from unmanned aerial system imagery. Crop Sci. 2017, 57, 2478–2489.
2. Shiu, Y.-S.; Chuang, Y.-C. Yield Estimation of Paddy Rice Based on Satellite Imagery: Comparison of Global and Local Regression Models. Remote Sens. 2019, 11, 111.
3. Fotheringham AS, Brunsdon C, Charlton M. 2003. Geographically weighted regression: the analysis of spatially varying relationships. Hoboken, NJ: John Wiley & Sons.

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.

Assessing Biological Nitrification Inhibition in the Rhizosphere of Field-Grown Bioenergy Sorghum

Mark B. Burnham[1]* (mburnham@illinois.edu), Sandra J. Simon,[1] D.K. Lee,[1,2] Angela D. Kent,[1,3] **Evan H. DeLucia**,[1,4,5] **Wendy H. Yang**[1,4,5,6]

[1]Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, Urbana, IL; [2]Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL; [3]Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign, Urbana, IL; [4]Institute for Sustainability, Energy and Environment, University of Illinois at Urbana-Champaign, Urbana, IL; [5]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL; [6]Department of Geology, University of Illinois at Urbana-Champaign, Urbana, IL

https://cabbi.bio/research/sustainability-theme/

Project Goals: Agricultural activity causes significant nitrate (NO_3^-) leaching and nitrous oxide (N_2O) gas fluxes, polluting groundwater and contributing to global greenhouse gas emissions. Improving our understanding of microbial NO_3^- and N_2O production via nitrification in agricultural soils could lessen the detrimental effects of bioenergy crop production. Sorghum is one of a handful of grasses that exude secondary metabolites with the potential to inhibit nitrification, but the characterization of biological nitrification inhibition (BNI) has been limited to laboratory and greenhouse studies. It is likely that environmental conditions alter the extent of BNI in the field, so we leveraged two sorghum trials in 2018 and 2019 to characterize BNI in rhizosphere soil under varying environmental and management conditions.

The heavy use of nitrogen (N) fertilizer to maximize crop yield represents a major disturbance of the N cycle that causes nitrate (NO_3^-) leaching into groundwater and emissions of the greenhouse gas nitrous oxide (N_2O) into the atmosphere. As a result, implementing annual bioenergy cropping systems to reduce our dependence on fossil fuels has other detrimental environmental impacts. In the soil ecosystem, bacterial and archaeal nitrification of ammonium (NH_4^+) to NO_3^- and subsequent denitrification of NO_3^- to N_2O and N_2 are key steps leading to NO_3^- and N_2O production. Laboratory and greenhouse studies show that sorghum, a candidate annual bioenergy feedstock, exudes secondary metabolites into the rhizosphere that inhibit nitrification. However, conditions in the field are likely to alter biological nitrification inhibition (BNI). We implemented two sorghum field trials in 2018 and 2019 and compared bulk and rhizosphere soil potential nitrification rates, potential denitrification rates, and microbial communities. During both seasons, BNI was strongest during the period of maximum plant growth, but was limited early and late in the growing season. BNI was substantially stronger in 2019, when the drier growing season likely caused hydrophilic BNI compounds to accumulate in rhizosphere soil. Although fertilizer addition weakened BNI in 2018, the apparent greater strength exerted by

reduced precipitation in 2019 outweighed any fertilizer effect. Potential denitrification was stimulated in the rhizosphere, suggesting that carbon (C) exudation by sorghum roots stimulates heterotrophic microbes and thus indirectly inhibits nitrification by creating competition for NH_4^+. Shifts in the rhizosphere microbial community relative to bulk soil, and the association of nitrifier relative abundance with potential nitrification rates when BNI was strongest mid-season, show that sorghum has some capacity to alter its rhizosphere community in a way that reduces NO_3^- production. Overall, we show that plant growth stage, precipitation variability, fertilization, and rhizosphere competition interact to control BNI in the field. Thus, although BNI has the potential to be an environmental benefit of bioenergy sorghum, its magnitude will strongly depend on growth conditions and the seasonal timing of NO_3^- production and leaching.

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.

Building a Comprehensive Set of Genetic Tools for Metabolic Engineering of *Issatchenkia orientalis*

Mingfeng Cao[1]*, Zia Fatma[1], Vinh G. Tran[1,2]*, Xiaofei Song[1,3], Ping-Hung Hsieh[1,4], Zengyi Shao[1,5], Yasuo Yoshikuni[1,4], **Huimin Zhao[1,2,6]** (zhao5@illinois.edu)

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation
[2]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign
[3]Department of Microbiology, Nankai University, Tianjin, China
[4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA
[5]Department of Chemical and Biological Engineering, Iowa State University, Ames, IA
[6]Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL

cabbi.bio

Project Goals: We aim to develop efficient genetic tools for metabolically engineering the acid-tolerant yeast *Issatchenkia orientalis*, including a stable episomal plasmid, a set of constitutive promoters and terminators, a CRISPR/Cas9-based gene deletion capability, and an *in vivo* DNA assembly method to produce biofuels and chemicals.

Abstract:

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, lack of efficient genetic tools, including a stable episomal plasmid and precise genome editing tool, remains a major bottleneck in metabolic engineering of this organism. Here we isolated one functional 0.8 kb centromere-like (CEN-L) sequence through *in silico* centromere prediction and library-based screening from the *I. orientalis* genome, which greatly improved the stability of the episomal plasmid[1]. We also discovered and characterized a set of constitutive promoters and terminators under different culture conditions by using RNA-Seq analysis and a fluorescent reporter[1]. In addition, the optimized CRISPR/Cas9 system consisting of an improved Cas9 and single guide RNA (sgRNA) expressed by a fusion *RPR1'-tRNA* promoter could achieve efficiencies of 97%, 90%, and 47% for single, double, and triple gene disruptions[2]. Lastly, we developed an *in vivo* assembly method in *I. orientalis*, which could achieve ~100% fidelity for assembly of a 7.4 kb-plasmid using 7 DNA fragments ranging from 0.7 kb to 1.7 kb[1]. These genetic tools should be generally applicable for rapid strain development and metabolic engineering of this organism for production of biofuels and chemicals.

References:

1. Cao M#, Fatma Z#, Song X, Hsieh P, Tran V, Lyon W, Sayadi M, Shao Z, Yoshikuni Y, Zhao H. Accepted in *Metabolic Engineering*
2. Tran V#, Cao M#, Fatma Z, Song X, Zhao H. *mSphere*, 2019, e00345-19

Funding Statement: *This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and*

Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

Coarse-grained Modeling of *Saccharomyces cerevisiae* Physiology

Viviana Nguyen[1,2]* (viviana2@illinois.edu), Yifei Li,[1,3] and **Ting Lu**[1,2,3,4]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois; [3]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois; [4]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, Illinois

Project Goals: We have developed a coarse-grained, mechanistic, and dynamic model of yeast physiology that integrates key processes of metabolic, gene regulatory, and signaling networks underlying cellular metabolism. The proposed model describes the cell's dynamic resource partitioning (proteome allocation) under varying glucose-replete and glucose-limited conditions. The model successfully captures key metabolic characteristics of yeast, including the Crabtree effect and diauxic shift observed during batch growth and the critical dilution rate seen in chemostat cultures.

S. cerevisiae has proven itself to be an industrial workhorse, where metabolic engineering has turned it into a cell factory for sustainable production of biofuels, pharmaceuticals, and other valuable chemicals. However, there have been experimental challenges in reprogramming the cells for production of heterologous chemicals due to the complexity of the underlying metabolic network. Mathematical modeling provides a potential solution to this problem. Genome-scale models have guided previous engineering attempts by predicting how mutations of a specific gene target or a group of targets could alter chemical production. Here, we provide a complementary approach to genome-scale modeling. We propose a coarse-grained, mechanistic, and dynamic model of yeast physiology that describes global resource allocation under varying nutrient conditions. The model is simple but is still able to capture key characteristics of yeast metabolism.

As protein synthesis is one of the most energetically expensive cellular processes and because cellular growth rate is proportional to protein synthesis rate[1], we chose to investigate resource allocation at the level of the proteome. Based on previous work of coarse-grained modeling of *E. coli*[2], we categorize the approximately 6000 proteins of *S. cerevisiae* into nine major proteome sectors: R (including gene expression machinery), 8 E proteins (including proteins that promote forward metabolic flux), and Z (housekeeping proteins and the remainder of the proteome). We found this to be the minimal number of coarse-grained enzymes required to reproduce the correct phenotype. Next, we developed a biophysical model that combines transcription and translation processes to describe protein synthesis. This module describes resource partitioning through competition for ribosomes among the different proteins. We then consider the cell's resource partitioning among amino acids, energy (ATP), ethanol, and storage carbohydrates. We chose PKA and SNF1 as coarse-grained global regulators that adjust the fermentation and respiration rates in response to varying glucose levels and TORC1 as the global regulator that responds to amino acid levels. These regulators adjust the resource allocation between the R and E sectors to meet metabolic demands.

We see that the model successfully captures key characteristics of yeast metabolism. Simulations return different proteome partitions for varying nutrient compositions. This changing proteome allocation provides a possible mechanistic explanation for the Crabtree effect as well as the diauxic shift and critical dilution rate that are observed in our simulations of batch and chemostat cultures. Under good nutrient conditions (glucose-replete), PKA and TORC1 promote cell growth and synthesis of R proteins, which minimizes the resource allocation to the E sector proteins. However, once glucose is depleted or in glucose-limited conditions, SNF1 shifts the metabolism toward ATP generation through remodeling of the proteome.

This work advances the fundamental understanding of microbial physiology by capturing and providing a quantitative and mechanistic explanation of the metabolic characteristics of *S. cerevisiae*. With addition of specific metabolic pathways to this simple framework, it has the potential to predict the effects of perturbations of the network on cell growth and chemical production to offer insight into the rational design of metabolic pathways.

References

1. Scott, M., Gunderson, C. , Mateescu, E., Zhang, Z., & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science*, **330**(6007), 1099–1102 (2010).
2. Liao, C., Blanchard, A., & Lu, T. An integrative circuit-host modeling framework for predicting synthetic gene network behaviors. *Nature Microbiology*. **2**, 1658–1666 (2017).

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.

Development of High Throughput MALDI-TOF-MS Workflow for Profiling Medium-Chain Fatty Acids from Microbial Colonies

Kisurb Choe[1], Pu Xue[2], **Jonathan Sweedler**[3] (jsweedle@illinois.edu)

[1] Department of Biochemistry, University of Illinois Urbana Champaign, Urbana, IL, USA; [2] Department of Chemical Engineering, University of Illinois Urbana Champaign, Urbana, IL, USA; [3] Department of Chemistry, University of Illinois Urbana Champaign, Urbana, IL,USA

Project Goals: The creation of a new mass spectrometry workflow for high throughput screening of microbial biocatalysts producing free fatty acids.

Abstract

Free fatty acids are value-added chemicals used as biofuels and industrial chemicals. Especially, there has been an increased need for medium chain fatty acids (MCFA) due to greater utility and little supply. Currently, there has been a lack of a high throughput analytical method of fatty acids profiling for biocatalyst research. We developed a MALDI-TOF based high throughput screening workflow for measuring fatty acids produced by microbial colonies. First, using shorter acyl-chain phosphatidylcholine (PC) as a proxy for MCFA from baker's yeast colonies, we isolated the variants of fatty acid synthase (FAS II) from mutant libraries that produce increased MCFA in liquid culture. Also, we developed a MALDI method based on N-Phenyl-2-naphthylamine (PNA) matrix showing sensitivity and robustness in measuring free fatty acids from *E. coli* colonies. For high throughput colony screening by these methods, we have developed a workflow using a colony picking robotics of an automation facility called Illinois Biological Foundry for Advanced Biomanufacturing capable of screening up to 10,000 colonies per day. Finally, we further developed an optically guided MALDI-TOF microMS workflow[1] as a robotics-free high throughput colony screening tool, where the updated software has been launched as a web application for collaborators to use for HT colony screening.

Publications

[1]Si T, Li B, Comi TJ, et al. Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. J Am Chem Soc. 2017;139(36):12466-12473.

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.

Development of High Throughput Primer Design and Quantification for Nitrogen Cycle Genes in Bioenergy Crop Soils

Jaejin Lee[1,2]* (jaylee@iastate.edu), Jia Liu,[2] Paul Villanueva,[2] Santosh Gunturu,[3,4] Jim Cole,[3,4] 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, IA; [3]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI; [4]Great Lakes Bioenergy Research Center

Project Goals: To improve understanding of how plant-soil-microbe interactions influence nitrogen cycling in bioenergy crop soils, this project aims to relate microbial functional gene diversity and abundance to rates of gross nitrogen cycling and gaseous nitrogen fluxes in bioenergy crop soil and to develop cost-effective, scalable tools to capture the phylogenetically diverse nitrogen cycle genes in environmental samples.

The development of molecular biological tools in combination with advanced sequencing technologies (e.g., 16S rRNA amplicon sequencing) has enabled high throughput characterization of community composition and structure. However, these approaches are often restricted to characterizing microbial community structure and cannot reliably provide information on the functional potential of genes. Metagenomic sequencing can be an improved approach to investigating diverse functional genes in environmental samples. However, these functional groups often comprise only a small fraction of the environmental DNA, resulting in high costs and low sequencing coverage. Another method to characterize functional genes leverages PCR-based methods. For PCR-based methods to target functional genes, the reliability of primer sets is a prerequisite, as the primer sets ultimately determine what is amplified in the environmental samples. Unfortunately, conventional PCR primers are known to detect a limited range of the diverse genes involved in nitrogen cycle. Further, the majority of currently available primers have been designed mainly for isolated strains instead of environmental biodiversity, and there is a lack of cost-effective and scalable platforms to cover the high diversity of target genes. As high throughput qPCR has become applicable for environmental samples, we can now assay hundreds of primer sets and genes in a single run. Thus, we have developed a pipeline to perform high throughput primer design based on abundant nitrogen cycle genes based on their abundance in over 1,900 existing soil metagenome samples. Through the pipeline, >400 novel primer sets were designed targeting denitrification (*nap*A, *nar*G, *nir*K, *nir*S, *nor*B, *nos*Z) and nitrification (*amo*A-AOA/AOB) genes. We have optimized BioMark HD, a high throughput qPCR system, for studying nitrogen cycling Miscanthus soil samples from the DOE CABBI LAMPS site.

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.

Effectiveness of Payments for Greenhouse Gas Mitigation to Induce Low Carbon Bioenergy Production

Fahd Majeed[1,2]* (fmajeed2@illinois.edu), Madhu Khanna[1,2], Ruiqing Miao[3], Elena Blanc[1], Tara Hudiburg[4], and Evan DeLucia[1]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation, [2]Department of Agriculture and Consumer Economics, University of Illinois, Urbana, IL; [3]Department of Agricultural Economics and Rural Sociolog, Auburn University, Auburn, AL; and [4]Department of Forest, Rangeland and Fire Sciences, University of of Idaho, ID

Project Goals: This study provides a comprehensive analysis of the potential impact of incentivizing greenhouse gas (GHG) mitigation from cellulosic ethanol feedstocks in a risk-averse crop production context. We examine the impact of payments linked to GHG reduction on cropping decisions of risk-averse farmers in the rain-fed area of the United States at various biomass prices with a focus on aggregate GHG mitigation, aggregate cost to reduce emissions, and the spatial distribution of feedstock. Additionally, the project explores whether the manner of payment influences cropping decisions by comparing lump-sum upfront payments for total GHG mitigated to annual payments for GHG mitigated each year. Additionally, we explore the cost-effectiveness of a GHG reduction payment approach for achieving a given emission reduction target compared to a uniform establishment cost payment.

Abstract text:

Motivation: Cellulosic ethanol feedstocks such as corn stover and energy crops have the potential to provide substantial greenhouse-gas reduction benefits, first, through aboveground emission reduction as cellulosic ethanol can displace less carbon-intensive gasoline and, second, via carbon mitigation through soil carbon sequestration. These feedstocks vary greatly in terms of their GHG intensities, production costs, and risks due to differing and spatially varying input requirements, yields, and soil carbon sequestration effects. For instance, corn stover is a low cost but low yielding source of biomass readily available to farmers planting corn but harvesting stover also results in the removal of soil carbon that would have otherwise remained sequestered in the ground. In contrast, energy crops like miscanthus and switchgrass provide spatially varying GHG reduction benefits through relatively high annual yield over the mature period of the crop, substantial soil carbon sequestration through the life of the crop, and their ability to grow on marginal land. Long establishment periods with high upfront costs and uncertain yields due to weather variations, however, reduce incentives for risk-averse, impatient, and credit-constrained farmers to produce these GHG mitigating energy crops.

Previous research shows that upfront subsidies to reduce establishment costs can incentivize the adoption of energy crops by risk-averse decision-makers but have not yet considered directly incentivizing the spatially varying GHG mitigation that these feedstocks provide that make them appealing in the first place. Areas that have the lowest cost for energy crop cultivation, for instance, may not necessarily be the areas with the highest GHG mitigation potential due to spatial differences in high sequestration and high yield areas, spatial differences in high miscanthus and switchgrass yield areas, and payment for GHG mitigation benefits of corn stover. Earlier research finds heterogeneity in emission reduction potential from different sources and sites. Uniform payments schemes such as those under the Renewable Fuel Standard and the Biomass Crop Assistance Program that do not take variation in GHG reduction potential into account may not be incentivizing those crops that provide the most GHG mitigation. Providing payments based on emission reduction would take advantage of spatial variability in GHG reduction potential, yields, and costs by creating differing incentives for

feedstocks in different locations and increase the production of cellulosic ethanol where greenhouse gas intensity of biofuel produced and overall cost of production is lowest.

The purpose of this paper is, first, to examine the impact of payments linked to GHG reduction on the cropping decisions of risk-averse farmers in the rain-fed area of the United States at various biomass prices with a focus on aggregate GHG mitigation, aggregate cost to reduce emissions, and the spatial distribution of feedstock. Second, we explore whether the manner of payments influences cropping decisions by comparing lump-sum upfront payments for total GHG mitigated to annual payments for GHG mitigated each year. Third, we explore the cost-effectiveness of a GHG reduction payment approach for achieving a given emission reduction target compared to a uniform establishment cost payment. To the best of our knowledge, this study is the first one that provides a comprehensive analysis of the potential impact of incentivizing GHG mitigation from cellulosic ethanol feedstocks in a risk-averse crop production context.

Methodology: We first present a theoretical framework under which a representative farmer optimally chooses her land allocation between conventional and energy crops. We examine how GHG mitigation payments will affect the farmer's optimal land allocations under various risk and time preferences and credit constraint specifications. We show that crop allocations will depend in part on yield riskiness, the temporal profile of returns, and diversification of the farmer's crop portfolio.We then use a stylized integrated model simulation framework that links an economic model with a biogeochemical model, DayCent, to analyze farmers' cropping allocation under GHG reduction payments while accounting for spatial and temporal heterogeneity in crop yields and GHG intensities. We use a Copula method to generate joint distributions in all crop yields and conventional crop prices to measure the risks of feedstock yields and conventional crop prices for each county. Each county is considered a representative farmer maximizing expected utility under various exogenously determined GHG reduction payment scenarios and carbon and biomass prices for an exogenous degree of risk aversion, time preferences, and credit constraints. The model is simulated for a fifteen-year cropping cycle for energy crops, where farmers can choose to allocate crops on cropland and marginal land and choose whether to harvest a portion of the corn stover from areas under corn production.

Data: Our numerical simulation focuses on the 1,919 rainfed counties in the eastern United States. In order to reflect the stochastic nature of prices and yields, county-level yields for thirty years for corn grain, corn stover, soybean, miscanthus, and switchgrass from DayCent and thirty-year harvest and futures corn and soybean Chicago Board of Trade prices are used to estimate a joint distribution for all yields and conventional crop prices. Belowground sequestration rates for conventional and energy crops are calculated from DayCent. Aboveground displacement is calculated as the sum of electricity co-product credits, material input use, emissions generated through electricity, diesel, transportation, and ethanol production, each of which are calculated using input intensities taken from the Greenhouse gases, Regulated Emissions, and Energy use in Transportation (GREET) model, model, nitrogen application rates from DayCent, and simulated yields. Additionally, land acreage by county for both cropland and marginal land is taken from National Agricultural Statistics Service (NASS). Input quantities for energy crops are compiled from Ohio State Extension documents, with input costs from NASS. Input quantities and costs for conventional crops are compiled from state extension documents and NASS. Parameters for the landowner's risk and time preferences are obtained from the literature. The model is solved for cropping shares under exogenous payment scenarios with exogenously varying biomass payment and an exogenous cost of carbon.

Results: Results show that without GHG reduction payments farmers require high biomass prices (at least $50 per ton in a not risk-averse, not credit constrained, and low discount scenario, and $70 per ton in the risk-averse, credit constrained, and high discounting scenario) to induce them to plant energy crops. GHG reduction payments cause farmers to adopt energy crops at low carbon prices when biomass price is between $30 and $60 per ton, with the potential to incentivize up to 200 million tons of biomass and displace or mitigate 350 million metric tons of carbon.

The manner in which GHG reduction is paid for influences feedstock adoption. Yearly payments are more cost-effective in driving farmer adoption when farmers are not risk-averse, not credit constrained, and have a low discount rate. For example, at a biomass price of $40 per ton, an $80 dollar price per ton of carbon drives farmers to produce 100 million metric tons of energy crops per year compared to 50 million metric tons when payment is upfront. Upfront payments being better are driving adoption when farmers are risk-averse, credit constrained, and have a high discount rate, with farmers producing 80 million metric tons from upfront payments but only 25 million metric tons from yearly payments.

GHG reduction payments are able to induce farmers to mitigate GHG more cost-effectively than uniform payments. For example, a uniform establishment cost share subsidy at a biomass price of $40 per ton in a less constrained scenario would cost 350 million dollars and result in 30 million metric tons of carbon displaced or mitigated. An upfront GHG reduction payment can achieve the same GHG mitigation goal for a cost of 250 million dollars by incentivizing just 46 counties that are able to mitigate GHG the most with the emissions reduction goal being met through a mix of energy crops and corn stover use.

Potential for generating discussion: Current bioenergy programs such as BCAP and RFS incentivize the production of energy crops but do not directly incentivize GHG mitigation and the impact of incentivizing GHG mitigation on biomass feedstock production by risk-averse farmers has not yet been studied. This study provides comprehensive economic analysis of the impact of GHG reduction payments on aggregate GHG mitigation, energy crop production, cropping choices, and geographical configuration of land allocation. Therefore, it should be of interest to economists and policy analysts who are interested in programs that support GHG mitigation or bioenergy production.

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.

Effects of Natural Variation in Self-Shading on Photosynthetic Traits in Sorghum

Nikhil S. Jaikumar[1,2*] (jaikumar@illinois.edu), Andrew D. B. Leakey[1,3] Pat Brown[4], Sam Fernandes[3] and Stephen P. Long[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL ; [3]Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL; [4] Department of Plant Sciences, University of California at Davis, Davis, CA.

[*] Presenting author.

Project Goals:

This project aims to identify natural variation in canopy structure and self-shading among over 800 sorghum accessions and to investigate the effect of self-shading on photosynthetic traits in the lower canopy with the eventual goal of optimizing light distribution and photosynthesis throughout the canopy.

Abstract:

Crops are being grown at ever higher densities, resulting in increasing pressure to maintain photosynthetic function under low light . While many C_3 plants have the ability to partially acclimate to low light levels over the course of a leaf lifespan[1,2], it is unclear whether C_4 plants have this same capacity, as previous studies in maize[3] and sugarcane[4] have yielded conflicting results. We compared upper and lower canopy leaves in diverse accessions of *Sorghum bicolor* with heavy self-shading (planophile accessions) and with less self-shading (erectophiles), over two field seasons at two sites. Parameters measured included maximum quantum yields of CO_2 assimilation (Φ_{CO2}) and whole chain electron transport (Φ_J), light-adapted respiration R_D, maximal photosynthetic capacity A_{max}, maximal electron transport rate J_{max},, and activity of three C_4 specific enzymes. We hypothesized that lower canopy leaves would show less of a decrease in photosynthetic efficiency relative to sunlit leaves in naturally occurring accessions with greater self-shading.

We found that:

- Shaded leaves exhibited lower Φ_{CO2} and Φ_J than sun leaves, but this decline was less in erectophile accessions (12-19%) than in planophiles .

- Due to combined effects of increased light penetration and improved photosynthetic efficiency in the lower canopy, we estimated that erectophile accessions support higher carbon assimilation in the lower canopy and on a whole-canopy basis than planophiles, with the advantage in lower canopy photosynthesis being greater in late season.
- Unlike the effect on photosynthetic rates, decreases in C_4 enzymatic activity and in R_D were not ameliorated by reducing the degree of self shading.

Breeders have found that erectophile lines of maize and sorghum are associated with higher yields, which has been attributed to a more optimized distribution of light. This study shows an additional major contributor, maintenance of photosynthetic efficiency under low light conditions.

References

1. Syvertsen JP. 1984. Light acclimation in citrus leaves. II. CO_2 assimilation and light, water and nitrogen use efficiency. *Journal of the American Horticultural Society.* 109: 812-817.

2. Beyschlag W, Barnes PW, Ryel R, Caldwell MM, Flint SD. 1990. Plant competition for light analyzed with a multispecies canopy model. II. Influence of photosynthetic characteristics on mixtures of wheat and wild oat. *Oecologia.* 82: 374-380.
3. Pignon CP, Jaiswal D, McGrath JM, Long SP. 2017. Loss of photosynthetic efficiency in the shade. An Achilles heel for the dense modern stands of our most productive C_4 crops? *Journal of Experimental Botany*. 68: 335-345.
4. Sales CRG, Ribeiro RV, Hayashi AM, Marchiori PER, Silva KI, Martins MO, Silveira JAG, Silveira NM, Machado EC. 2018. Flexibility of C_4 decarboxylation and photosynthetic plasticity in sugarcane plants under shading. *Environmental and Experimental Botany.* 149: 34-42.

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.

Energizing the Machinery of Storage Lipid Synthesis in Plant Vegetative Tissues

Yingqi Cai[1,2], Zhiyang Zhai[1,2], Jantana Keereetaweep[2], Hui Liu[1,2], Changcheng Xu[1,2], and **John Shanklin**[1,2]* (shanklin@bnl.gov)

[1]Center for Advanced Bioenergy and Bioproducts Innovation; [2]Biology Department, Brookhaven National Laboratory, Upton, New York

http://cabbi.bio

Project Goals: As part of the CABBI Feedstock Production theme, the overall goal of our research program is to enhance the production of vegetative lipids in bioenergy grass stems for industrial purposes.

Abstract text.

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 growing demand for fossil oil. Substantial progress has been made to enhance the production of vegetative lipids in bioenergy grass stems for industrial purposes by engineering factors directly involved in TAG biosynthesis and degradation. To further enhance TAG biosynthesis in vegetative tissues, we hypothesized that supplying the lipid synthesis machinery with additional energy may help to overcome some intrinsic limits. Our study has identified three adenosine triphosphate (ATP)-related factors (PAP2, NTT1 and NTT2) that, when overexpressed, have positive effects on TAG accumulation in plant cells. Purple Acid Phosphatase 2 (PAP2) has been shown to elevate ATP content and vegetative biomass when overexpressed in Arabidopsis, probably by dephosphorylating proteins and affecting their transport into chloroplasts and mitochondria. Arabidopsis plants overexpressing PAP2 showed increased FA synthesis rates in both leaves and siliques. The TAG content was also increased in *Nicotiana benthamiana* leaves when co-expressing PAP2 with lipogenic proteins (WRI1, DGAT1 and OLE1). The other two factors tested are plastid envelope-localized nucleoside triphosphate transporters (NTT1 and NTT2) that transport ATP from the cytosol into plastids. Transient overexpression of NTT1 or NTT2 alone or in combination with WRI1 significantly increased TAG accumulation in *N. benth.* leaves. Our findings provide novel targets that can be stacked with lipogenic factors to enhance TAG accumulation in plant vegetative tissues and potentially overcome the yield drag associated with high levels of vegetative TAG accumulation.

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.

Establishment of Efficient Multiplex Genome Editing in Sorghum Using Green Calli

Praveena Kanchupati[1,3*] (praveena@illinois.edu), Tony Trieu[1,2], Kankshita Swaminathan[1,2] and **Stephen P. Moose**[1,3]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation
[2]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama
[3]University of Illinois Urbana-Champaign, Urbana, Illinois

Project goals: To establish an efficient CRISPR/Cas9 gene editing system in sorghum (*Sorghum bicolor*, genotype Tx430)

In many grasses, including sorghum, immature zygotic embryos are the preferred source tissue for the induction of embryogenic callus cultures capable of efficient DNA transfer, selection of transformed cells, and regeneration into fertile plants (Raghuwanshi and Birch, 2010). One advantage of embryogenic callus cultures is they form a compact structure of large numbers of rapidly dividing cells, offering more genomic targets for transformation. Current protocols, irrespective of the DNA delivery method, achieve 20% to 50% transformation efficiency (Guo et al., 2015; Liu et al., 2015; Belide et al., 2017). However, these protocols require the maintenance of robust stock plants for a continuous supply of immature embryos that have a short window (two days) of competence for callus initiation. An alternative approach adapted from Cho et al. (1998) produces highly generable green callus tissue that can be maintained for up to two years, and although high transformation efficiencies with microprojectile bombardment have been reported, it has not been widely used because of low efficiencies with *Agrobacterium* transformation. Many of the issues associated with DNA integration following microprojectile bombardment are not a concern in genome editing experiments, so we are revisiting whether the green callus system may be a more efficient protocol for CRISPR/Cas9 genome editing in the sorghum genotype Tx430. We have successfully produced multiple transgenic callus lines containing a monocot optimized Csy4-Cas9 gene fusion and four guide RNAs. Plants are currently in regeneration and being characterized for possible edits, and experiments are in progress targeting mutagenesis of several genes expected to alter nitrogen utilization and carbon partitioning.

References

1. Belide S, Vanhercke T, Petrie JR, Singh SP., 2017. Robust genetic transformation of sorghum (Sorghum bicolor L.) using differentiating embryogenic callus induced from immature embryos. Plant Methods,13:109.
2. Cho MJ, Jiang W, Lemaux PG., 1998. Transformation of recalcitrant barley cultivars through improvement of regenerability and decreased albinism. Plant Science 138, 229–44.
3. Guo X., Ge Z., Sato S.J., Clemente T.E. (2015) Sorghum (Sorghum bicolor). In: Wang K. (eds) Agrobacterium Protocols. Methods in Molecular Biology, vol 1223. Springer, New York, NY

4. Liu G, Edward K, Gilding B, Godwin ID., 2015. A robust tissue culture system for sorghum (Sorghum bicolor (L.) Moench). South African Journal of Botany 98, 157–60.
5. Raghuwanshi, A., Birch, R.G., 2010. Genetic transformation of sweet sorghum. Plant Cell Reports 29, 997–1005.

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.

Gene Targeting and Targeted Mutagenesis for Genetic Improvement of Oilcane

Tufan Oz[1,2], Baskaran Kannan[1,2*] (kbaskaran@ufl.edu), Chakravarthi Mohan[2], Sara Sanchez[1,2], Ayman Eid[1,2], Sofia Cano[1,2], Niki Koukoulidis[1,2], Ratna Karan[2], Aldo Merotto[2], John Shanklin[1,3] and **Fredy Altpeter**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]University of Florida, Gainesville, FL; and [3]Brookhaven National Laboratory, Upton, NY

https://cabbi.bio/research/feedstocks-theme/

Project Goals:

1) Developing and deploying enabling bio-technologies for precision breeding of oil producing sugarcane.
2) Elevating the yield of biomass and vegetative lipids in sugarcane.

Abstract:

Genome editing tools such as CRISPR/Cas9 and other RNA guided nucleases (RGNs) have been employed in several crops. They enable targeted mutagenesis or gene targeting following DNA break repair by Non-Homologues End Joining (NHEJ) or Homology Directed Repair (HDR), respectively. NHEJ generates an abundance of random insertions and deletions (InDels). Frameshift mutations associated with these InDels of unspecified size and sequence might result in loss of function phenotypes of agronomic importance. Gain of function mutations, on the other hand, generally require precise nucleotide editing in the target gene for the replacement of inferior alleles with superior alleles by template mediated HDR. We present data confirming highly efficient HDR mediated precision nucleotide editing in multiple alleles of the acetolactate synthase (ALS) gene in the highly polyploid sugarcane which confer herbicide resistance. Faithful transmission of superior ALS alleles with introduced target mutations at 574 and/or 653 amino acid locations to vegetative progenies was confirmed with Sanger sequencing, PacBio SMRT sequencing and resulted in herbicide resistance.

Creation of knockout phenotypes is challenging in highly polyploid crops like sugarcane. We will present a rapid readout system that accelerates identification of multi-allelic edits by targeting candidate genes in chlorophyll biosynthesis. This system with facilitate the comparison of alternative RGNs and delivery systems.

Following successful demonstration of both, gene targeting and targeted mutagenesis in sugarcane, we are currently generating sugarcane lines with CRISPR/Cas9 mediated genome editing to suppress flowering and hydrolysis of triacyl glycerol (TAG). Extending the vegetative phase in sugarcane is expected to increase the yield of biomass and vegetative lipids. Targeted mutagenesis of the TAG lipase should further elevate lipid accumulation.

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.

Leveraging Comparative Population Genomics to Dissect the Mechanisms of *Issatchenkia orientalis* Fluconazole Resistance

Ping-Hung Hsieh[1,2], Yusuke Sasaki[1,2], Jing Ke[2], Zhiying Zhao[2], and **Yasuo Yoshikuni[1,2*] (yyoshikuni@lbl.gov)**

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation, [2]Lawrence Berkeley National Laboratory, Berkeley, CA

https://cabbi.bio/

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

Issachenkia orientalis is an emerging non-model ascomycetes yeast with unparalleled ability to tolerate multiple stresses, including extremely low pH, high temperature, and high concentrations of lignocellulosic inhibitors, salts, alcohols, and organic acids. These unique characteristics may make *I. orientalis* an attractive chassis for producing biofuels and bioproducts directly from lignocellulosic hydrolysates. Understanding how *I. orientalis* evolved to tolerate multiple stresses may allow engineering of a strain more suitable for industrial use than natural isolates are. We performed a population genomics study of 162 strains collected from various habitats and identified 305,435 single nucleotide polymorphism (SNPs), 16,177 insertions and deletions (indels), and other genetic variations, including ploidy, gene copy number, and pan-genome variations. We are currently working on genome-wide association study (GWAS) to understand genetic variations underlying various phenotypes. Here we discuss the results for fluconazole resistance, an unfavorable characteristic for industrial utilization of microbes.

References

1. Mesquita, Vanessa A., et al. (2015) Impact of Multi-Metals (Cd, Pb and Zn) Exposure on the Physiology of the Yeast *Pichia Kudriavzevii*. *Environmental Science and Pollution Research.* 22(14): 11127–11136.

2. Isono, Naoto, et al. (2012) A Comparative Study of Ethanol Production by *Issatchenkia orientalis* Strains under Stress Conditions. *Journal of Bioscience and Bioengineering.* 113(1): 76–78.

3. Peter, Jackson, et al. (2018) Genome Evolution across 1,011 *Saccharomyces cerevisiae* Isolates. *Nature.* 556(7701): 339–344.

Funding statement.

This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

Linking Microbial Community Structure and Function for Sustainable Production of Bioenergy Crops

Alonso Favela,[1,3] Sierra Raglin[1,2], Niuniu Ji[1] , Sandra Simon[1], Rachel Waltermire[1], and **Angela Kent**[1,2,3]* (akent@illinois.edu)

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation; [2]Department of Natural Resources and Environmental Sciences and [3]Program in Ecology, Evolution, and Conservation Biology; University of Illinois at Urbana-Champaign, Urbana, IL

Project goals: Our overall objective is to identify the genomic and biochemical mechanisms by which plants influence their rhizosphere microbiome and its functions, enabling future germplasm improvement; and to determine the impact of altered rhizosphere nitrogen cycling functions on nutrient retention and sustainability in agroecosystems. Our interdisciplinary team employs a combination of metagenomics, measures of N cycling process rates, and plant breeding genetics to address these knowledge gaps.

The plant microbiome helps plants acquire scarce resources and is an essential target for improving agricultural sustainability, particularly for biofuel feedstocks where reduction of anthropogenic inputs is desirable. However, the ability of modern crops to recruit and structure their microbiome may be altered by domestication and breeding. Selection for crop plants based on aboveground traits in high nutrient environments may inadvertently lead to changes to belowground plant physiology and relationships with the soil microbiome. In doing so, we have likely altered microbiome functions that contribute to sustainability and environmental quality (e.g. nutrient acquisition, nutrient retention, and GHG production).

The first step to investigate this is to examine the structure and activity of the microbiome as a function of plant genotype, with a specific focus on N cycling functional groups. The functional capacity of the rhizosphere microbiome and the benefit to the host plant, as well as ecosystem services such as nutrient cycling and greenhouse gas emissions, vary with the composition and abundance of microbial assemblages. We hypothesized that plant genotypes differ in their ability to recruit microbial functional groups, and ultimately that the functional profile of the microbial community can be treated as a selectable plant phenotype and optimized through plant breeding. We used targeted functional metagenomic sequencing to survey the rhizosphere of diverse genotypes of several crops to compare their capability to recruit microbial nitrogen cycling functional groups, and examined rates of nitrogen transformations. Using maize as a model for bioenergy grasses, we observed significantly different N cycling microbial community structure among crop genotypes that represent the endpoints of directed evolution. In addition, differential abundance and composition of N cycling functional groups was associated with significant reduction in nitrification and denitrification in specific maize lineages. We are following up with examination of sorghum genotypes that strongly vary in nutrient use efficiency. Our results allow the linkage of host-associated microbial communities and ecosystem function and suggest that there is genetic capacity to optimize recruitment of N cycling functional groups, and improve

crop sustainability. Understanding the mechanistic underpinnings of this relationship will allow breeders and ecosystem scientists to select bioenergy crop cultivars that interact with the nitrogen cycle in ways that improve the efficiency and sustainability of agriculture, while protecting environmental quality.

Funding statement: *This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.*

Metabolic Engineering of Triacylglycerols with Specialized Fatty Acids in Sorghum

Kiyoul Park[1,2,3]* (kpark4@unl.edu), Truyen Quach[1,2,4], Hyojin Kim[1,2,3], Zhengxiang Ge[1,2,4], Tieling Zhang[1,2,4], Tara Nazarenus[1,2,3], Shirley Sato[1,2,4], Haejin Kim[1,2,3], Ming Guo[1,2,4], **Tom E. Clemente**[1,2,4], and **Edgar B. Cahoon**[1,2,3]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation

[2]Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, USA

[3]Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE, USA

[4]Department of Biotechnology, University of Nebraska-Lincoln, Lincoln, NE, USA

https://cabbi.bio/research/feedstocks-theme/

Project Goals: Our research aims to enhance production of triacylglycerols (TAG) with specialty fatty acids in sorghum (*Sorghum bicolor* (L.) Moench) stems. Metabolic engineering of TAG biosynthesis with specialty fatty acids (e.g., medium-chain and hydroxy fatty acids) can increase TAG value and functionality for renewable alternatives to petroleum-based oils. Sorghum is regarded as an ideal feedstock for bioenergy because of its capacity for high biomass productivity with low agronomic inputs. Applying synthetic biology for modular and multiple gene assembly, we are introducing multigene-expression vectors in sorghum and currently evaluating the most effective transgene combinations using tobacco systems to increase the throughput of the design-build-test-learn cycle of our synthetic biology efforts.

Plants store triacylglycerols (TAG) in their seed as energy sources for germination and seedling establishment. Although TAG is principally stored in seeds for this purpose, numerous studies indicate that vegetative tissues also have the ability for TAG synthesis that can serve as energy-dense, renewable alternatives to petroleum-based oils. Sweet sorghum (*Sorghum bicolor* (L.) Moench) is regarded as an attractive feedstock for bioenergy because of its capacity for high biomass production under rain-fed conditions on marginal soils. The goal of this research is the synthetic re-design of lipid metabolism in stalks for enhanced TAG production and storage and to generate TAG with specialty fatty acids (e.g., medium-chain and hydroxyl fatty acids) to increase TAG value and functionality. Accumulation of TAG is achieved by the three major processes: push, pull, and protect (3P). Fatty acid synthesis can be pushed using master transcription factors such as WRINKLED1 (WRI1), pulled into TAG assemply by acyltransferases such as diacylglycerol acetyltransferase (DGAT), or protected from degradation by

oil body coat proteins such as oleosin. In addition to this fundamental gene combination (WRI1 + DGAT + oleosin), specific acyltransferases and acyl-ACP thioesterases are required for specialty fatty acid synthesis. We have identified and characterized the specific acyltransferases and acyl-ACP thioesterases from various plant species. We currently focus on medium-chain and hydroxy fatty acids because they are suitable for jet fuel and industrial heat-tolerant lubricants, respectively. By applying synthetic biology for modular and multiple gene assembly, we generated diverse vector constructs for TAG biosynthesis with specialty fatty acids. We are currently evaluating the most effective transgene combinations using tobacco systems to increase the throughput of the design-build-test-learn cycle of our synthetic biology efforts. The most effective combinations are then used for sorghum transformation.

This work is supported by U. S. Department of Energy Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420.

Oleaginous Yeasts for the Production of Sugar Alcohols

Sujit S Jagtap[1,2]* (jagtap@illinois.edu), Ashwini A Bedekar[2], Jing-Jing Liu[1], Anshu Deewan[1,2], and **Christopher V Rao**[1,2]

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign, Illinois. [2]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Illinois.

https://cabbi.bio/

Project Goal: The goal of this project is to investigate sugar alcohol production from plant-based sugars and glycerol in the oleaginous yeasts *Rhodosporidium toruloides* and *Yarrowia lipolytica*. We are also interested in understanding the mechanism of sugar alcohol production and the key genes involved in the polyol synthesis process.

Sugar alcohols are commonly used as low-calorie, natural sweeteners. They have also been proposed by the Department of Energy as potential building blocks for bio-based chemicals. They can be used to produce polymers with applications in medicine and as precursors to anti-cancer drugs [1]. Production of these sugar alcohols by yeast often results, from redox imbalances associated with growth on different sugars, accumulation of toxic intermediates, and as a cell response to the high osmotic pressure of the environment [2-3]. The ability of yeast to naturally produce these sugar alcohols from simple sugars provides a potentially safer and more sustainable alternative to traditional chemical hydrogenation.

In our study, we found that the oleaginous yeast *R. toruloides* IFO0880 produces D-arabitol during growth on xylose in nitrogen-rich medium [3]. Efficient xylose utilization was a prerequisite for extracellular D-arabitol production. D-arabitol is an overflow metabolite associated with transient redox imbalances during growth on xylose. *R. toruloides* is also able to produce galactitol from galactose [2]. In addition, *R. toruloides* was able to produce galactitol from galactose at reduced titers during growth in nitrogen-poor medium, which also induces lipid production. These results suggest that *R. toruloides* can potentially be used for the co-production of lipids and galactitol

from galactose. We further characterized the mechanism for galactitol production, including identifying and biochemically characterizing the critical aldose reductase. Intracellular metabolite analysis was also performed to further understand galactose metabolism.

We also explored sugar alcohol production in *Y. lipolytica*. It is known to produce erythritol during growth on glycerol. The heterologous overexpression of a sugar polyol phosphatase increases erythritol production two fold in *Y. lipolytica*. Intracellular metabolite analysis was also performed to further understand glycerol metabolism in wild type and engineered strains. We also overexpressed the key genes involved in glycerol assimilation pathways and the pentose phosphate pathway.

R. toruloides and *Y. lipolytica* have traditionally been used for the production of lipids and lipid-based chemicals [4-5]. Our work demonstrates that these non-model yeasts can also produce arabitol, galactitol, and erythritol. Collectively, our results further establishes that *R. toruloides* and *Y. lipolytica* can produce multiple value-added chemicals from a wide range of sugars and glycerol.

References

1. Jagtap, S. S.; Rao, C. V., Microbial conversion of xylose into useful bioproducts. *Applied Microbiology and Biotechnology* **2018,** *102* (21), 9015-9036.
2. Jagtap, S. S.; Bedekar, A. A.; Liu, J.-J.; Jin, Y.-S.; Rao, C. V., Production of galactitol from galactose by the oleaginous yeast *Rhodosporidium toruloides* IFO0880. *Biotechnology for Biofuels* **2019,** *12* (1), 250.
3. Jagtap, S. S.; Rao, C. V., Production of d-arabitol from d-xylose by the oleaginous yeast *Rhodosporidium toruloides* IFO0880. *Applied Microbiology and Biotechnology* **2018,** *102* (1), 143-151.
4. Zhang, S.; Ito, M.; Skerker, J. M.; Arkin, A. P.; Rao, C. V., Metabolic engineering of the oleaginous yeast *Rhodosporidium toruloides* IFO0880 for lipid overproduction during high-density fermentation. *Applied Microbiology and Biotechnology* **2016,** *100* (21), 9393-9405.
5. Zhang, S.; Jagtap, S. S.; Deewan, A.; Rao, C. V., pH selectively regulates citric acid and lipid production in *Yarrowia lipolytica* W29 during nitrogen-limited growth on glucose. *Journal of Biotechnology* **2019,** *290*, 10-15.

The Center for Advanced Bioenergy and Bioproducts 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.

Quantifying the Plant-Microbial Interactions Controlling Soil Organic Matter Formation in Bioenergy to Improve Model Representations of Sustainability

Joanna Ridgeway (jrridgeway@mix.wvu.edu)[1], and **Edward Brzostek**[1]

[1]West Virginia University, Morgantown

https://cabbi.bio

Project Goals: The Center for Advanced Bioenergy and Bioproducts Innovation's (CABBI) mission is to develop efficient ways to grow bioenergy crops, transform biomass into valuable chemicals, and market the resulting biofuels and other bioproducts. A key part of this mission is to ensure the sustainability of a bioenergy economy by maintaining or enhancing ecosystem services (e.g., soil carbon sequestration, nitrogen retention). However, our ability to predict the extent to which different management strategies and bioenergy feedstocks impact ecosystem services is limited by simplified models that do not represent the diversity of microbial traits or the ability of plant-microbial interactions to feedback on soil biogeochemical cycling. Thus, the goal of this project within CABBI is to quantify these traits and interactions that regulate soil carbon and nitrogen cycling to develop and validate a plant-microbial interactions model that predicts impacts on ecosystem services.

Enhancing soil carbon sequestration may enable the bioenergy industry to achieve carbon neutrality. To realize this potential, we must develop a predictive understanding of how management and feedstock decisions impact carbon stabilization in soil organic matter (SOM). However, there remains uncertainty in how interactions between plant traits (i.e. litter chemistry, rhizosphere exudation) and microbial traits (i.e. carbon use efficiency (CUE), turnover) drive SOM formation for bioenergy feedstocks. This hinders modeling efforts to predict the long term effects of transitioning between bioenergy feedstocks on soil carbon sequestration.

To address this uncertainty, we investigated key plant and microbial trait interactions influencing SOM formation for traditional bioenergy corn and alternate Miscanthus x Giganteus (miscanthus) feedstock systems. Plant-microbial interactions influence SOM formation indirectly through litter chemistry with rapidly, decomposing litters driving efficient microbial biomass production that upon death is thought to preferentially form mineral-associated SOM over particulate SOM. Thus, we hypothesized that low carbon to nitrogen ratio (C:N) corn litter decomposes faster and forms more mineral associated SOM than high C:N miscanthus litter. Directly, plants influence SOM formation by exuding carbon to microbes in the rhizosphere (zone of soil proximal to plant roots) in exchange for nutrients, promoting microbial activity and mineral-associated SOM formation. Given differences in fertilization rates and root traits between feedstocks, we hypothesized that soil microbes in miscanthus systems have a greater capability to use root exudates to drive greater biomass production and mineral associated SOM formation than heavily fertilized corn systems. To test these hypotheses in the lab, we incubated

^{13}C isotopically labeled aboveground and belowground litter from each crop in soil collected from experimental plots at the University of Illinois Energy Farm to assay microbial traits and trace the fate of litter into microbial biomass, CO_2, and mineral vs. particulate SOM pools. We simulated root exudates by adding a cocktail of organic acids to half of the incubated samples.

In support of our first hypothesis, preliminary results show that corn litter initially decomposed very rapidly, prompting greater respiration losses of carbon compared to miscanthus litter. In the last four weeks, corn litter decomposition slowed as litter C was likely immobilized in microbial biomass and mineral associated SOM. By contrast, miscanthus litter C continued to decompose at a consistent rate by less efficient microbes with more of the litter likely remaining in particulate SOM forms. Supporting our second hypothesis, exudate carbon additions promoted decomposition of litter over SOM in corn. By contrast, miscanthus showed the opposite pattern with exudates primarily speeding up SOM decomposition. Collectively, these results indicate that there is an important interaction between litter chemistry and root traits that controls the formation and decomposition of SOM in bioenergy systems. In addition, these data can be used directly to reduce parameter uncertainty in microbial traits and the fate of bioenergy plant inputs in microbial-mediated decomposition models.

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.

Rewiring Metabolism to Construct a Yeast Strain Capable of Producing 2,3-butanediol Without Ethanol and Glycerol Production

Jaewon Lee[1] (jlee762@illinois.edu), Nurzhan Kuanyshev[1], Liang Sun[2], and **Yong-Su Jin[1,2]**

[1]DOE Center for Advanced Bioenergy and Bioproducts Innovation

[2]Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, Illinois

Project Goals

Complete use of glucose and xylose is a prerequisite for producing biofuels and chemicals from lignocellulosic biomass. We have constructed an engineered *Saccharomyces cerevisiae* strain (CT2) capable of rapidly and efficiently producing ethanol from glucose and xylose. To construct a platform yeast strain for the production of CABBI target molecules (organic acids, alcohols, and lipids), metabolic designs enabling the maximum production of target molecules with no byproducts were tested in the CT2 strain using 2,3-butanediol (2,3-BDO) as a target molecule. We envision that this new strain, which efficiently and rapidly produces 2,3-BDO without byproduct formation, can be further engineered to produce other CABBI target molecules via modifications in metabolic, redox balance, and energetic pathways.

Abstract

2,3-butanediol (2,3-BDO) is a versatile commodity chemical which can be used for making synthetic rubbers, anti-freeze, and liquid fuel additives. While production of 2,3-BDO by microbial fermentation has been demonstrated before, use of potential human pathogenic microorganisms, formation of byproducts, and low productivities limited the commercialization of the fermentation processes.

As an alternative, we engineered *Saccharomyces cerevisiae*, which is a GRAS (generally recognized as safe) and preferred microorganism by industrial biotechnology companies, and achieved efficient and rapid production of 2,3-BDO without byproduct formation. Key objectives were to 1) introduce a 2,3-BDO production pathway, 2) to eliminate the formation of byproducts, such as ethanol and glycerol, and 3) to re-wire redox-balancing metabolic pathways for the production of 2,3-BDO with a high yield (>80%) and productivity (>1 g/L·h).

To these ends, we first deleted major isozymes of *PDC* (pyruvate dehydrogenase) and *ADH* (acetaldehyde dehydrogenase) genes in *S. cerevisiae* to minimize ethanol production while maintaining sufficient acetyl-CoA pool for cell growth. Second, heterologous genes—*alsS,* coding for acetolactate synthase, and *alsD,* coding for acetolactate decarboxylase, from *B. subtilis* were overexpressed. Third, we eliminated glycerol accumulation by deleting both *GPD1* and *GPD2*, which code for glyceraldehyde-3-phosphate dehydrogenase. Fourth, we introduced a NAD^+ regenerating pyruvate-malate cycle to resolve the redox imbalance from deletion of the glycerol

producing pathway. Lastly, we enhanced the expression level of *PYC1* and *PYC2* to enhance the NAD^+ regenerating capability of the pyruvate-malate cycle. As a result, our best strain was able to produce 2,3-BDO with a much higher productivity (1.1 g/L·h) than previously constructed 2,3-BDO producing strains (0.1~0.2 g/L·h) in a batch fermentation with 100 g/L of glucose. In addition to the rapid production of 2,3-BDO, the best strain produced negligible amounts of glycerol and ethanol. As such, the engineered yeast offers the potential for economical downstream processing and efficient catalytic upgrading of 2,3-BDO.

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.

Towards a fully automated algorithm driven platform for biosystems design

Mohammad HamediRad[1,2,6], Ran Chao[1,2,6], Scott Weisberg[3,6], Jiazhang Lian[1,7], Shekhar Mishra[1,2]*(smishr10@illinois.edu), Saurabh Sinha[2,4] & Huimin Zhao[1,2,3,5]

[1]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[2]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[3]Departments of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[4]Departments of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[5]Departments of Chemistry and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

[6]Present address: LifeFoundry Inc., 60 Hazelwood Dr., Champaign, IL 61820, USA

[7]Present address: Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, 310027 Hangzhou, China

Project goals:
To achieve full automation of the Design-Build-Test-Learn process for biosystems design via an integrated robotic system coupled with machine learning algorithms.

https://www.igb.illinois.edu/DOEcenter

Abstract:
Large-scale data acquisition and analysis are often required in the successful implementation of the design, build, test, and learn (DBTL) cycle in biosystems design. However, it has long been hindered by experimental cost, variability, biases, and missed insights from traditional analysis methods. Here, we report the application of an integrated robotic system coupled with machine learning algorithms to fully automate the DBTL process for biosystems design. As proof of concept, we have demonstrated its capacity by optimizing the lycopene bio- synthetic pathway. This fully-automated robotic platform, BioAutomata, evaluates less than 1% of possible variants while outperforming random screening by 77%. A paired predictive model and Bayesian algorithm select experiments which are performed by Illinois Biological Foundry for Advanced Biomanufacturing (iBioFAB). BioAutomata excels with black-box optimization problems, where experiments are expensive and noisy and the success of the experiment is not dependent on extensive prior knowledge of biological mechanisms.

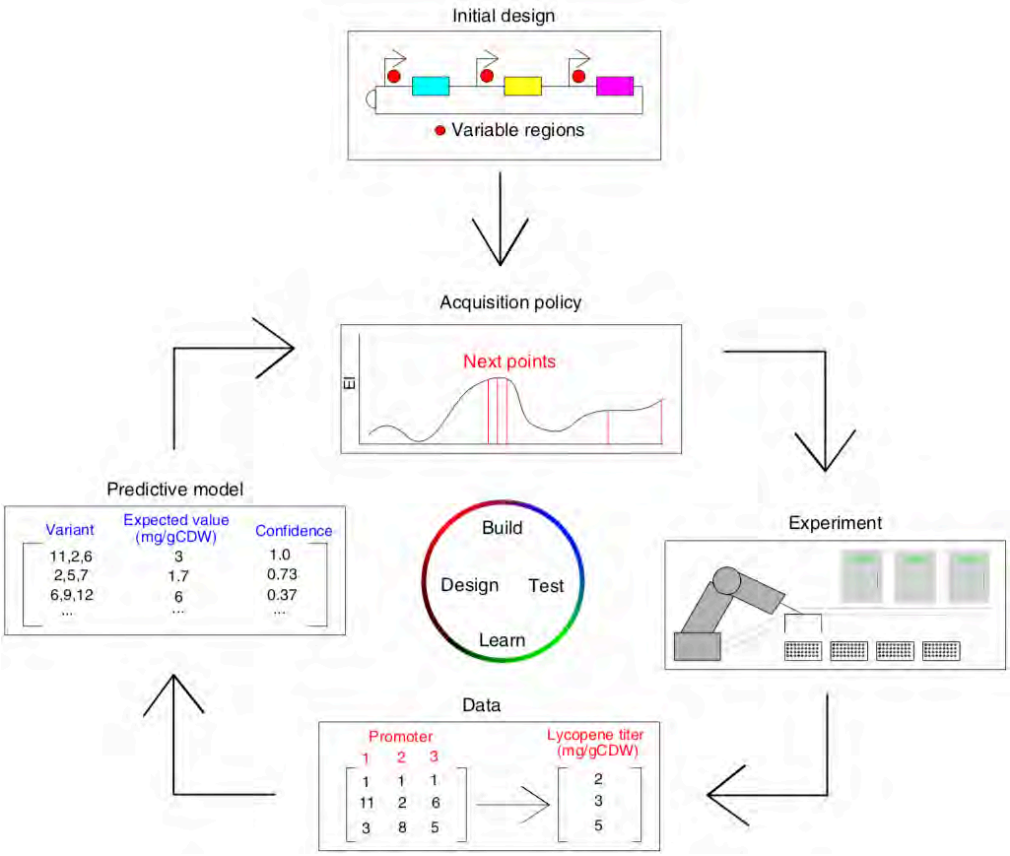


Figure 1: The overall workflow of BioAutomata. After setting the initial parameters, designing the sequence space of variable regions (such as promoter variants in a combinatorial pathway assembly), and defining the objective function, BioAutomata selects which experiments are expected to result in the highest improvement of yield, performs those experiments, generates data and learns from it, updating its predictive model given the newly presented evidence. It will then decide on the next experiments to perform to reach the goal set by the user while trying to minimize the number of experiments and the cost of the project

References:

1. HamediRad, M., Chao, R., Weisberg, S. *et al.* Towards a fully automated algorithm driven platform for biosystems design. *Nat Commun* **10,** 5150 (2019). https://doi.org/10.1038/s41467-019-13189-z

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-SC0018420.

Discovering Innovations in Stress Tolerance Through Comparative Gene Regulatory Network Analysis and Cell-Type Specific Expression Maps.

José R. Dinneny[1*] (dinneny@stanford.edu), Maheshi Dassanayake,[2] Dong Ha Oh,[2] Song Li,[3] and John Schiefelbein[4]

[1]Stanford University, Stanford, CA; [2]Louisiana State University, Baton Rouge, LA; [3]Virginia Tech, Blacksburg, VA; and [4]University of Michigan, Ann Arbor, MI

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.

The current post-genomics era is defined by a multitude of genome sequences that are available to study. Experimentally tractable molecular genetic model species were some of the first to have their genomes sequenced. Today, however, genomes of organisms that are difficult to culture, rare or even extinct exist in our data repositories. While their genome sequences have been unlocked, the nature of the genes contributing to this fascinating diversity in physiology and development is currently hidden due to a lack of methods available to extract significant functional meaning. Which genes allow cactus to survive the desert heat, or sea grass to grow in ocean water, or some ferns to tolerate desiccation? In agriculture, domestication has led to the breeding of rapidly growing cultivars that perform well when the climate cooperates, but often fail when water or nutrients are limiting. Bioenergy crops will be grown on soils of poor quality so as not to compete with other agricultural sectors, yet these environments will dramatically impact biomass accumulation. If the innovations that nature has selected for across plant species can be discovered, we have an opportunity to address these challenges and improve the sustainability of agriculture in ways that are simply impossible by traditional breeding.

Innovations in gene function allow wild plants to inhabit environments that are commonly stressful to domesticated crops. Our interdisciplinary team seeks to identify such innovations by defining the regulatory and physiological context that genes function across 11

sequenced Brassicaceae genomes, including bioenergy crops, and crop wild relatives, using recent advances in single-cell RNA sequencing (scRNA-seq) and DNA affinity purification sequencing (DAP-seq) technologies. Furthermore, machine-learning algorithms will utilize the evolutionary history of gene duplication events and functional genomics data to identify innovations in gene function associated with growth control under environmental stress. Putative genes associated with extremophyte resilience will be introduced into stress sensitive species to test whether extremophyte physiological properties can be transferred to naïve genomes. These studies will utilize a molecular genetic model that can be rapidly characterized using scalable and open source phenomics systems. Through our investigation, we will establish an experimental and data analytics pipeline that will be broadly applicable to the study of gene function at the plant family level and result in a road map for improving plant traits for bioenergy and beyond.

Funding statement: ***U.S. Department of Energy,*** *DE-SC0020358: Discovering innovations in stress tolerance through comparative gene regulatory network analysis and cell-type specific expression maps, Awarded September 2019*

Plant-Microbe Interfaces: Application of machine-learned protein-metabolite binding prediction models to plant-microbe interfaces

Omar N. Demerdash[1]* (demerdashon@ornl.gov), Amy L. Schaefer,[2] Caroline S. Harwood,[2] Dale A. Pelletier,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]The University of Washington, Seattle, WA

http://pmiweb.ornl.gov/

Project Goals: The goal of the PMI SFA is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. *Populus* and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, we focus on 1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, 2) defining the chemical environment and molecular signals that influence community structure and function, and 3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.

Computational prediction of the binding of small organic metabolites and other ligands to biological macromolecules has far-reaching implications for a range of problems, particularly metabolomics. Small metabolites are implicated in a host of roles, including symbiotic relationships between plant and microbe. Nonetheless, critical tasks such as predicting the bound structure of a protein-ligand complex along with its affinity have proven to be very difficult, owing largely to the inherent approximations in generating physically reasonable bound conformations of the ligand and an accurate free energy or proxy thereof. In recent years, machine learning-based methods have proven to be more robust than the standard linear sum of energetic terms, suggesting a complex, potentially non-linear interaction among terms. However, these methods are often trained on a small set of features, with a single functional form for any given energetic or physical effect, and often with little mention of the rationale behind choosing one functional form over another. Moreover, a systematic investigation of the effect of machine learning method is not undertaken, with a single method being favored for reasons that are often obscure. Here we undertake a comprehensive effort towards developing high-accuracy, machine-learned scoring functions, systematically investigating the effects of machine learning method and choice of features, and, providing insights into the relevant physics using methods that assess feature importance. Here, we show synergism among disparate features, yielding Pearson correlations (R^2) with experimental binding affinities of up to 0.865 and enrichment for native bound structures of up to 0.913 in an independent test set consisting of the well-known CASF-2013 benchmark. We deploy these models to predict the relative activity of metabolites in two systems of importance in plant-microbe symbiosis, one plant-bacterial (the LuxI enzyme and its potential substrates), and the other an enzyme that synthesizes plant defense hormones. We show the ability to discriminate

low- from high-activity substrates and describe further how these methods shall be deployed on a larger scale to screen larger sets of molecules.

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.

Plant-Microbe Interfaces: Systems biology approaches to understanding a plant's adaptation to and regulation of the phytobiome

Daniel Jacobson* (jacobsonda@ornl.gov), Jared Streich, David Kainer, Ben Garcia, Piet Jones, Timothy J. Tschaplinski and **Mitchel J. Doktycz.**

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.

We have created Parakraken, a parallelized kmer profile-based taxa identification approach, which uses full kmer profiles from every publicly available genome sequence (including bacteria, archaea, fungi, viruses, nematodes, insects, protists, etc.) to allow detection of the cryptic phytobiome in the DNA or RNA sequence of any host tissue. We are using Parakraken to identify taxa in different host tissues across a genome wide association study (GWAS) population of *Populus trichocarpa.* We also have other independent layers of 'omics data across this population and over 28,000 temporal-climactic phenotypes (climatypes) measured across the original locations of the genotypes in the population. In addition, we have developed new methods for GWAS using the presence/absence of genes in the pan genome of this population. Furthermore, we have developed supercomputing and explainable-AI approaches to find complex epistatic architectures responsible for the host's ability to detect and modulate its microbiome and other phenotypes. The result is a comprehensive systems biology model of a plant and its microbiome, its adaptation to its climactic environment and the metabolic intermediaries involved therein. Thus, we combine the results of these approaches with many orthogonal layers of information in order to score each hypothesis supported by multiple lines of evidence to prioritize specific mechanisms for experimental validation.

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

Plant-Microbe Interfaces: Experimental characterization of protein movement from plants to ectomycorrhizal fungus

Xiaohan Yang[1]* (yangx@ornl.gov), Hua Cheng,[1] Linling Li,[1] Haiwei Lu,[1] Xiaoli Hu,[1] Md Mahmudul Hassan,[1] Zhenzhen Qiao,[1] Guoliang Yuan,[1] Jin Zhang,[1] Degao Liu,[1] Jennifer Morrell-Falvey,[1] Jessy Labbé,[1] Wei Shen,[2] Shi-You Ding,[2] Muneeba Khalid,[1] Scott T. Retterer,[1,3] Julie Mitchell,[1] Jin-Gui Chen,[1] Wellington Muchero,[1] Gerald A. Tuskan,[1] and **Mitchel J. Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Department of Plant Biology, Michigan State University, East Lansing, MI; [3]Center for Nanophase Materials Sciences, Oak Ridge National Lab, 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.

Beneficial interactions between plants and fungi in the below-ground world are important for improving ecosystem stability and crop resilience. The establishment of plant-fungal associations in the rhizosphere requires complex molecular cross-talk between symbiotic partners *(Martin et al.*, 2017). Mycorrhiza-induced small secreted protein MiSSP7 can move from the mutualistic fungus *Laccaria bicolor* into the roots of *Populus* to facilitate the establishment of symbiosis with host trees (Plett *et al*., 2014; Plett *et al*., 2011). Yet, to our knowledge, there is no prior evidence showing that proteins can move from plant roots into the hyphae of their fungal partners. Recently, we predicted that more than 400 *P. trichocarpa* small secreted proteins (PtSSPs) could be responsive to symbiosis with *L. bicolor* (Plett *et al.,* 2017). We hypothesized that some of these PtSSPs can move from plant roots into *L. bicolor* hyphae. To test this hypothesis, we selected a subset of 14 PtSSPs (i.e., PtSSP1, PtSSP2, …, PtSSP14), based on computational analysis of DNA-binding capability and signal peptides for secretion, for experimental characterization of protein movement from plant roots into fungal hyphae. Transgenic *Arabidopsis thaliana* and poplar plants were created to overexpress these PtSSPs fused to green fluorescent protein (GFP). The transgenic plants were co-cultured with *L. bicolor* to assess the movement of the PtSSP-GFP fusion proteins. So far, we have found that PtSSP1, PtSSP5 and PtSSP8 could move from roots of transgenic *Arabidopsis* and/or poplar into the hyphae of *L. bicolor*. We are currently establishing a microfluidic platform to monitor the movement of PtSSP-GFP fusion proteins from

transgenic yeast cells into *L. bicolor* hyphae, and using computational approaches to predict the protein domains and 3D structures of these PtSSPs.

In summary, the results from our experiments support our hypothesis that PtSSPs can move from plant roots into the hyphae of *L. bicolor*. In the near future, we will study the impact of these mobile poplar small proteins on fungal gene expression and identify the potential fungal proteins interacting with the poplar small proteins.

References

Martin, F.M., Uroz, S. and Barker, D.G. (2017) Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science* **356**, eaad4501.

Plett, J.M., Daguerre, Y., Wittulsky, S., Vayssières, A., Deveau, A., Melton, S.J., Kohler, A., Morrell-Falvey, J.L., Brun, A. and Veneault-Fourrey, C. (2014) Effector MiSSP7 of the mutualistic fungus Laccaria bicolor stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proceedings of the National Academy of Sciences* **111**, 8299-8304.

Plett, J.M., Kemppainen, M., Kale, S.D., Kohler, A., Legué, V., Brun, A., Tyler, B.M., Pardo, A.G. and Martin, F. (2011) A secreted effector protein of Laccaria bicolor is required for symbiosis development. *Current Biology* **21**, 1197-1203.

Plett, J.M., Yin, H., Mewalal, R., Hu, R., Li, T., Ranjan, P., Jawdy, S., De Paoli, H.C., Butler, G., Burch-Smith, T.M., Guo, H.-B., Ju Chen, C., Kohler, A., Anderson, I.C., Labbé, J.L., Martin, F., Tuskan, G.A. and Yang, X. (2017) *Populus trichocarpa* encodes small, effector-like secreted proteins that are highly induced during mutualistic symbiosis. *Scientific Reports* **7**, 382.

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

Plant-Microbe Interfaces: Identification of gene products involved in plant colonization by *Pantoea* sp. YR343 using a plant-responsive diguanylate cyclase

Jennifer Morrell-Falvey[1,2,3*](morrelljl1@ornl.gov), Amber N. Bible,[2] Mang Chang,[3] and **Mitchel J. Doktycz[1,3]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville TN; [3]UT-ORNL Graduate School of Genome Science and Technology, 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.

Populus deltoides (poplar) hosts a diverse microbiome that influences its growth and productivity. The ability of plant growth promoting bacteria to exert beneficial effects on plant hosts is mediated through chemical and physical associations with plant tissues. *Pantoea* sp. YR343, a gamma-proteobacterium isolated from the rhizosphere of *P. deltoides*, forms robust biofilms along the root surfaces of *Populus* and possesses plant growth-promoting characteristics, such as phytohormone production and phosphate solubilization. The mechanisms governing biofilm formation along plant roots by bacteria, including *Pantoea* sp. YR343, are not fully understood and many genes involved in this process have yet to be discovered. Because the signaling molecule cyclic di-GMP plays an important role in biofilm formation, we employed a strategy for identifying putative colonization factors by modulating c-di-GMP expression in *Pantoea* sp. YR343. To this end, we identified three diguanylate cyclases, enzymes that synthesize c-di-GMP, that are expressed during colonization of plant roots. Overexpression of one of these diguanylate cyclases (encoded by PMI39_02884) significantly impacted exopolysaccharide production, motility, and biofilm formation. This overexpression strain was utilized for a genetic screen to identify genes that respond to high levels of c-di-GMP. Several genes were identified, including a UDP-galactose lipid carrier transferase (PMI39_01848) and a capsule polysaccharide transporter (PMI39_03059), which are predicted to function in EPS production. Transposon mutants affecting these genes were further characterized for their ability to colonize plant roots.

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.

Plant-Microbe Interfaces: Identification and characterization of Proteolytic Cleavage Product (PCP) peptides that function as key signaling molecules for Plant-Microbe Interactions

Him Shrestha,[1,2] Robert L. Hettich[2]* (hettichrl@ornl.gov), Manuel I. Villalobos Solis,[1,2] Suresh Poudel,[2] Clémence Bonnot,[3] Claire Veneault-Fourrey,[3] Francis Martin,[3] Paul E. Abraham,[2] and **Mitchel J. Doktycz**[4]

[1]Department of Genome Science and Technology, University of Tennessee-Knoxville, Knoxville, Tennessee 37996, United States; [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States; [3]UMR 1136 INRA-Université de Lorraine 'Interactions Arbres/Microorganismes', Laboratoire d'Excellence ARBRE, Centre INRA-Lorraine, 54280 Champenoux, France; [4]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

https://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 ectomycorrhizal fungus, *Laccaria bicolor,* forms mutualistic association with roots of *Populus* species, in which *L. bicolor* provides *Populus* access to mineral nutrients from the soil, such as complex organic nitrogen, in exchange for fixed carbon derived from photosynthesis. This mutualistic interaction is enabled by the extensive crosstalk between fungi and the plant host. Proteolytic cleavage products (PCPs) are emerging as key signaling molecules that mediate cell-to-cell crosstalk. PCPs are post-translationally processed products of proteins that are involved in various biological processes that occur between and within plants, fungi, and bacteria. The discovery and characterization of PCPs have been challenging because these signaling molecules usually function at extremely low concentrations and undergo extensive post-translational processing that are not well understood. Moreover, PCPs originating from small open reading frames are often overlooked in gene predication/annotation tools due to their small size, and thus are missing in the reference database. In this study, we utilized a molecular weight based selective enrichment strategy, combined with high-performance tandem mass spectrometry and *de novo*-assisted database searching to improve the identification of PCPs. We benchmarked the qualitative and quantitative performance of the purposed approach using reference synthetic peptides.

Initial work focused on evaluating this approach to identify PCPs from different tissues of *Populus* interacting with *L. bicolor*. In total, we identified 1660 *Populus* and 2870 *L. bicolor* PCPs. Besides

qualitative identification of well-known PCPs, the LC-MS/MS method was able to capture a total of 157 PCPs that were significantly more abundant in root tips with established ectomycorrhiza as compared to root tips without established ectomycorrhiza and extramatrical mycelium of *L. bicolor*. These PCPs mapped to 64 *Populus* proteins and 69 *L. bicolor* proteins, with several of them previously implicated in biologically relevant associations between plant and fungus, including a variant of the Mycorrhiza-Induced Small Secreted protein MiSSP7.6. We then extended this approach to identify the PCPs involved in regulating *Populus* and *L. bicolor* interaction at high nitrate concentrations. Even though ectomycorrhizal interactions are often prevalent in the presence of nitrate, little is known about this symbiotic interaction when soils are exposed to extremely high level of nitrate from anthropogenic sources like fertilizers. Experimental observation shows that the rate of *L. bicolor* colonization in *Populus* root is regulated at high nitrate concentration. To better understand this regulatory mechanism, PCPs were extracted and identified from the root tissue of *Populus* with/without *L. bicolor* interaction that were further treated with different concentrations of nitrate. In total, we identified 1443 PCPs in root tissue of *Populus* with *L. bicolor* interaction, out of which 126 PCPs were differentially abundant in different concentrations of nitrate. While the data-analyses are still underway, some of these differentially abundant PCPs mapped to 19 *Populus* proteins and 79 *L. bicolor* proteins, with several of them being oxidative stress-related proteins. Overall, PCPs identified in these studies help further our understanding of molecular progression involved in selecting and maintaining a symbiotic relationship between *Populus* and *L. bicolor*. Moreover, the method implemented in this study provides an avenue for identifying novel PCPs in other biological system.

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. Funding also provided by the Laboratory of Excellence ARBRE (ANR-11-LABX-0002-01).

Plant-Microbe Interfaces: Metabolomics of non-host switchgrass plants expressing a poplar lectin receptor-like kinase in response to the mycorrhizal fungus *Laccaria bicolor*

Zhenzhen Qiao,[1] Timothy Tschaplinski[1*] (tschaplinstj@ornl.gov), Jin-Gui Chen,[1] Wellington Muchero,[1] Nancy Engle,[1] Jerry Tuskan,[1] Jessy Labbé,[1] and **Mitchel J. Doktycz**[1]

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

https://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 black cottonwood poplar (*Populus trichocarpa*) lectin receptor-like kinase (PtLecRLK1) was recently identified that mediates the symbiosis between *P. trichocarpa* and *Laccaria bicolor* (Labbé *et al*., 2019). When PtLecRLK1 was heterologously expressed in *Arabidopsis thaliana*, a non-host species for *L. bicolor*, the transgene induced the ability of *Arabidopsis* to display interstitial hyphal growth and Hartig net-like extracellular structures created by *L. bicolor*, and suppressed the host's defense responses upon exposure to *L. bicolor*, a key mechanism initiating colonization. Given that *Arabidopsis* is not known to harbor ectomycorrhizal relationships, a new study was initiated to determine if heterologously expressing PtLecRLK1 in a grass species that is a known host of ectomycorrhiza can result in establishing a symbiotic relationship that, otherwise, would not occur. Four transgenic switchgrass (*Panicum virgatum*) lines expressing PtLecRLK1 were generated and gas chromatography-mass spectrometry (GC-MS)-based metabolomics were conducted on roots of transgenic plants in contrast with wild-type plants growing in the presence of *L. bicolor*. The largest metabolomic responses of transgenesis were associated with accumulation of numerous nitrogenous metabolites, which are likely associated with the observed decline in plant growth. Given that there were declines in fatty acids and organic acids, which have been observed previously with symbiosis, the metabolomic results suggest that many of the early steps in successful colonization occurred, but that later-stage events were lacking.

References

1. Labbé, J., W. Muchero, O. Czarnecki, J. Wang, X. Wang, A.C. Bryan, K. Zheng, Y. Yang, S.S. Jawdy, L.E. Gunter, W. Schackwitz, J. Martin, F. Le Tacon, T. Li, Z. Zhang, P. Ranjan, X. Yang, D.A. Jacobson, T.J. Tschaplinski, J. Schmutz, J-G. Chen, and G.A. Tuskan. 2019. Mediation of plant-mycorrhizal interaction by a lectin receptor-like kinase. Nature Plants 5:676-680.

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.

Plant-Microbe Interfaces: Pathway prediction and production through cell-free synthetic biology

David C. Garcia,[1,2] Jaime L Dinglason,[1,2] Benjamin P. Mohr,[1,2] Robert F. Standaert,[1] Him Shrestha,[1,3] Paul E. Abraham,[3] Robert L. Hettich,[2,3] and **Mitchel J. Doktycz**[1,2 *] (doktyczmj@ornl.gov)

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Department of Genome Science and Technology, University of Tennessee-Knoxville, Knoxville, Tennessee 37996, United States; and [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

https://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.

Collectively, *Populus* and its microbiome have the potential to produce a large number of highly diverse and unique metabolites. As our definition and understanding of the plant and its microbiome has progressed, so too has the backlog of information and open questions generated from the thousands of uncharacterized proteins and metabolites that comprise these systems. To address this problem, we are developing new tools in order to rapidly test and define the function of uncharacterized proteins and metabolic pathways. Cell-free systems have developed into a powerful tool for synthetic biology and metabolic engineering with applications across multiple disciplines. Developments in cell-free biology have remarkably improved its capacity for expressing proteins as well as created the field of cell-free metabolic engineering. We are advancing cell-free systems as a rapid means for exploring protein function and metabolite production using two complementary approaches. The first, uses crude cell-free extracts to produce proteins related to potential metabolic pathways in newly isolated organisms. Upon modular assembly of the pathway and metabolite analyses, active and inactive pathways are differentiated. The second, focuses on developing cell-free extracts as bioproduction platforms by analytically verifying the metabolic pathways resident in a cell free system and determining which are drawing resources towards and away from the production of specific metabolites. As a result, we have significantly expanded our ability to use cell extracts outside of their native context to solve metabolic engineering problems and provide engineers new tools that can rapidly explore the function of proteins and test novel metabolic pathways.

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.

Plant-Microbe Interfaces: Simplified community approach to investigate the dynamic host-microbiome relationship

Dana L. Carper, Dale A. Pelletier* (pelletierda@ornl.gov), David J. Weston, Leah H. Burdick, Sara S. Jawdy, Mircea Podar, and **Mitchel J. Doktycz**

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 complex *Populus* host and microbial interactions, we isolated over 3,200 bacteria and developed a synthetic community system that employs subsets of these microbes. The isolate collection contains representatives from many of the dominant and abundant community members found in *Populus* field studies. However, much of the microbial taxonomic diversity from the *Populus* rhizosphere has not been cultured. We have representatives of 6 of the 21 phyla with abundance of >0.1% of the community. Thus, we have ongoing efforts to cultivate and characterize *Populus* rhizosphere bacteria that are under-represented or absent from our existing collection using single cell sorting on plant-derived culture systems (Podar *et al*., 2019). Our prior studies utilized reduced communities (< 10 members) to evaluate questions relating to the functional genetics underlying plant – microbe interactions. Now, we have created a synthetic community approach that allows for the design and evaluation of large complex communities (>150 members) to determine how host plant genetics, nutrient and environmental interactions shape community interactions and function. To create the synthetic community, all 16S rRNA sequences from the sequenced bacterial strains within our culture collection were extracted, trimmed to the V4-V6 region, and aligned. Using our DISCo-microbe software (Carper *et al*., 2020), we have designed a community consisting of 150 members that spans 4 phyla, 9 classes, 12 orders, 32 families and 77 genera. These 150 members have been inoculated into double autoclaved soil containing *P. trichocarpa* plants and exposed to differing environmental conditions for growth over 3 weeks: control (C), warm temperatures (W), cold temperature (CT), low nitrogen (LN) and warm temperature and low nitrogen (WLN). Overall, inoculated plants were smaller than the uninoculated controls suggesting an initial negative effect from the microbial load. The negative effects between inoculated and un-inoculated were statistically significant for number of leaves, change in stem height and leaf area, although this effect depended on the environmental condition.

The W and WLN conditions had the greatest effect phenotypically on the stem height and leaf area. Initial sequencing of the bacterial community of C- and CT-conditioned plants identified 95 out of 150 bacterial members present in host tissues. Tissue type (rhizosphere, root, stem and leaves) was the main factor in structuring the community in both the weighted (45.8%, p=0.0003) and unweighted (34.0%, p=0.0001) UniFrac metrics. The environmental condition also played a role in structuring community with more variation explained from the unweighted (20.3%, p=0.0001) than weighted (9.3%, p=0.0035) UniFrac metrics. This suggests that community differences between the environmental conditions is primarily from changes in low abundance community members. The two most abundant members in both environmental conditions were *Rhodanobacter* and *Paraburkholderia* strains. Within the CT leaves, a *Pantoea* strain was the most abundant microbe. Additionally, several strains were found across both treatments and all tissues sampled, including *Rahnella aquatilis*, a strain with demonstrated nitrogen fixing ability. Further investigations are ongoing to better understand how *Populus* structures its microbiome in response to genetic and environmental change.

References

Podar M, J Turner, LH Burdick, DA Pelletier. Complete Genome Sequence of *Terriglobus albidus* Strain ORNL, an Acidobacterium Isolated from the *Populus deltoides* Rhizosphere. Microbiol Resour Announc 8:e01065-19. https://doi.org/10.1128/ MRA.01065-19

Carper DL, Lawrence TJ, Carrell AA, Pelletier DA, Weston DJ. DISCo-microbe: Design of an identifiable synthetic community of microbes. *PeerJ*; 2020. 10.7287/peerj.preprints.27898v1.

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.

Plant-Microbe Interfaces: Temporal Variation in Plant-Microbe Interactions

Nicholas C. Dove,[1] Melissa A. Cregger[1*](creggerma@ornl.gov), Toni Wahl,[1] Jake Nash,[2] Rytas Vilgalys,[2] Karen E. Mock,[3] Christopher W. Schadt,[1] and **Mitchel J. Doktycz[1]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Duke University, Durham, NC; [3]Utah State University, Logan, UT

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.

It is increasingly recognized that microorganisms living inside or in close association with plant tissues are integral for plant health and survival. Our previous work has shown that *Populus* harbors distinct microbiomes among its different tissues[1]; however, the temporal stability of these microbial communities is unclear as measurements across time have so far been limited. Here, we present work from two subprojects that aim to characterize the initial assembly as well as the intra- and inter-annual stability of the *Populus* microbiome. To assess the initial assembly of the *Populus* microbiome, we initiated a common garden study consisting of 10 *Populus* genotypes from two *Populus* species, *Populus deltoides* and *Populus trichocarpa*. Overall, we found that archaeal, bacterial, and fungal community assembly of the *Populus* microbiome is consistent among genotypes. The rhizosphere soils provide a significant proportions of taxa in the leaf and root endosphere. Additionally, using a null modeling approach, we estimate that the underlying assembly processes are at first stochastic but become more deterministic with time. We have continued to sample this common garden four times per year, and we are now in the fourth year of this experiment. These additional years will give us insights to the intra- and inter-annual variation of the *Populus* microbiome upon the first few years of planting.

To assess the inter-annual variation of *Populus* across longer times scales (*i.e.,* decades), we are taking advantage of a time-for-space substitution approach in multiple aspen (*P. tremuloides*) clones, including what is largely considered to be Earth's most massive organism: Pando. We have so far identified a gradient of ramet ages in four aspen clones, and this summer we plan to sample leaves, xylem, fine roots, and rhizospheres for microbiome analysis.

Taken together, these projects will begin to elucidate the stability of the microbiome at both large scales and fine resolution in a long-lived plant over the lifetime of the host. Such information on

the temporal dynamics of the microbiome may be useful in identifying circumstances where microbiome interventions would be most successful.

References

1. Cregger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, et al. The Populus holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome* 2018; **6**.

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

Accelerating Yield Improvement in Switchgrass through Genomic Prediction of Floral Anthesis

Neal W. Tilhou (wepking@wisc.edu)[13], **Michael.D. Casler** [23]

[1]Department of Agronomy, University of Wisconsin, Madison, WI;

[2]USDA-ARS, U.S. Dairy Forage Research Center, Madison, WI.

[3]Great Lakes Bioenergy Research Center, Madison, WI.

Project Goals:
This project aims to accelerate the breeding process for bioenergy switchgrass in the North Central USA by applying genomic selection to yield limiting traits such as cold tolerance and flowering time.

Abstract text:
The timing of the transition from vegetative to reproductive growth has a major impact on biomass accumulation in switchgrass. Late flowering switchgrass varieties produce greater biomass in both spaced and sward conditions. Genomic prediction may allow rapid identification and selection of late flowering individuals without the time and expense of phenotyping. Initial analyses were carried out using the date of anthesis for 1,532 switchgrass individuals in multiple breeding groups. Marker data from genotype-by-sequencing (~450,000 markers after filtering) was used to predict anthesis date within each group. Prediction accuracy within breeding groups indicated accuracy sufficient to reduce the time required to identify individuals with superior breeding value.

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.

Associative nitrogen fixation (ANF) in high-yielding switchgrass varieties

S. Carolina Córdova[1,2,3]* (cordov13@msu.edu), Nyduta Mbogo[4], Stephen K. Hamilton[1,2,5], **G. Philip Robertson**[1,2,3]

[1]Great Lakes Bioenergy Research Center, Michigan State University, MI;

[2]W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI;

[3]Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI;

[4]Department of Biology, University of Missouri-Kansas City, Kansas City, MO (REU student);

[5]Department of Integrative Biology, Michigan State University, East Lansing, MI.

Project Goals:

Switchgrass (*Panicum virgatum*), a model bioenergy crop due to its high adaptability to many environments, may obtain much of its nitrogen (N) through a casual association with free-living N fixing bacteria (so-called associative nitrogen fixation or ANF). Quantification of ANF is challenging as this process has not been investigated extensively, and is possibly episodic. Moreover, information on ANF rates among switchgrass varieties or ecotypes is scarce. We used $^{15}N_2$ to measure switchgrass ANF via 1) in-vitro 7-day incubations of plant and soil rhizosphere samples, and 2) 48-hour incubations of whole plants moved from the field to a greenhouse. We contrasted two high-yielding switchgrass varieties, Cave-in-Rock and Kanlow (upland and lowland ecotypes, respectively), grown for 10 years at the W.K. Kellogg Biological Station in southwest Michigan. In-vitro experiments demonstrated ANF in roots and soil samples from both varieties, but not in leaves+stems from either. Preliminary findings from the greenhouse experiment showed a significant interaction between variety and sample type (i.e., plant tissue and soil): Cave-in-Rock roots had ANF rates 25-times higher than Kanlow roots, and rhizosphere soil from Kanlow plots had ANF rates 2.8-times higher than rhizosphere soil from Cave-in-Rock plots. In follow-on experiments, we will further explore how the ANF rate varies with switchgrass variety, especially at different phenological stages.

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; the National Science Foundation Long-term Ecological Research Program (DEB 1832042) at the Kellogg Biological Station; and by Michigan State University AgBioResearch.

Crabtree-like aerobic xylose fermentation through increased metabolic flux and altered sugar signaling pathways in *Saccharomyces cerevisiae*

Sae-Byuk Lee[1], Mary Tremaine[1], Michael Place[1,2], Robert Landick[1,3], Audrey P. Gasch[1,2], David J. Krause[1,2], Chris Todd Hittinger[1,2], and **Trey K. Sato[1]* (tksato@glbrc.wisc.edu)**

[1] DOE Great Lakes Bioenergy Research Center, Univ. of Wisconsin-Madison, Madison, WI; [2]Laboratory of Genetics, Univ. of Wisconsin-Madison, Madison, WI; and [3]Department of Biochemistry, Univ. of Wisconsin-Madison, Madison, WI

https://www.glbrc.org/

Project Goals

We aimed to better understand how to enhance the conversion of xylose into fermentative end-products by yeast through directed evolution and engineering.

Abstract

Plant-based biofuel is considered as sustainable and renewable energy. Xylose, which composes up to 40% of the sugar present in plant cell walls, cannot be fermented into biofuels, such as ethanol, by native *Saccharomyces cerevisiae*, the most common biofuel-producing organism. Xylose contrasts with glucose, which *S. cerevisiae* has evolved to ferment at such high rates that it will do so aerobically (Crabtree Effect) and preferentially in the presence of other carbon substrates (glucose-repression). Despite extensive knowledge of the regulatory networks controlling carbon metabolism, little is known about how to reprogram *S. cerevisiae* to ferment xylose at rates comparable to glucose. Previously, we discovered that loss-of-function mutations in *ISU1*, *HOG1*, *GRE3* and *IRA2* enabled *S. cerevisiae* strain engineered with xylose metabolism enzymes to respire xylose aerobically and ferment xylose anaerobically. Still, however, these genetic changes do not confer xylose conversion rates similar to that for glucose. Here, we report on our approach to enhance the rate of xylose fermentation by converting our engineered, xylose-respiring yeast strain into one that ferments xylose into ethanol aerobically. First, we deleted *COX15*, which is essential for respiration and rendered the strain unable to grow aerobically on xylose. We then evolved the strain to grow aerobically on xylose, subsequently isolating two independent clones with the abilities to grow on and convert xylose into ethanol aerobically. Evolved strains expressing (or retransformed with) *COX15* metabolize xylose aerobically and anaerobically faster than the unevolved parent. Whole genome sequencing of these clones identified overlapping duplications in Chromosomes IV and XVI; overlapping regions included the site where the xylose metabolism enzymes were engineered and the pentose phosphate pathway enzyme *TKL1*. Engineered duplications of the xylose metabolism enzymes and *TKL1* in the unevolved parent strain enabled greater fermentation of xylose-to-ethanol aerobically, suggesting the greater metabolic flux of xylose is a major requirement for aerobic xylose fermentation. However, the rationally-engineered strain did not ferment xylose to the

same extent as the evolved strains, indicating that additional genetic differences lead to their Crabtree-like traits or phenotype for xylose. Together, our findings identify genetic changes that may allow for faster conversion of xylose from plant biomass into biofuels.

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.

Data-driven design and engineering of biomolecules: mRNA and DNA

Sanjan TP Gupta[1,2*](sgupta78@wisc.edu), Gina C Gordon[1,3], Xiangyang Liu[2,4], Srivatsan Raman[2,4], Brian F Pfleger[1,3], and **Jennifer L Reed**[1,2]

[1]Department of Chemical and Biological Engineering, UW Madison, WI – 53706; [2]Great Lakes Bioenergy Research Center, Madison, WI – 53726; [3]Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI - 53706 [4]Department of Biochemistry, UW Madison, WI – 53706

Project Goals:

This project aims to employ machine learning principles to understand the design principles of biomolecular function by building empirical models using sequence-activity measurements, and in turn, use these models to accelerate the design-build-test cycle.

Abstract:

The advent of high-throughput technologies coupled with dropping costs of sequencing has led to the generation of new biological datasets paving the way for data-driven engineering of biomolecules. The current work describes case studies of engineering two different classes of biomolecules (mRNA and DNA) with applications to metabolic engineering and synthetic biology.

In the first study, ML models were built for predicting the mRNA half-lives in cyanobacteria - a photosynthetic microbe that can convert CO_2 into a variety of chemicals. A set of 28 sequence and structure based features (such as GC content, predicted RBS strength, and minimum free energy based on RNA folding) were compiled for the 3,238 genes found in *Synechococcus* sp. PCC 7002. Half-live values were measured for the corresponding mRNA transcripts using a rifampicin based transcription arrest assay and used as the target variable to be predicted based on the feature values. Analyzing the importance of various features used for building the model revealed that stable transcripts have higher normalized expression levels, higher translation rates, and are less likely to be found in an operon. Later, counts of 3 to 8 lettered sequence motifs in the 5' and 3' UTRs (untranslated regions) were used as features to build half-life predictors using a variant of random forest approach. These models were able to predict the half-lives accurately (with a spearman rank coefficient of 0.88 under 10-fold cross validation) and helped reveal of set of putative sequence motifs that could be used to enhance the stability of any gene of interest.

The second study looks at building ML models to predict the inducibility of genes under the control of *de novo* promoters with potential applications to genetic circuit design and development of biosensors for detecting intra-cellular metabolites. Using one-hot encoding and support vector regression, quantitative models were built to accurately predict the fold-induction ratios for a given operator sequence corresponding to three different prokaryotic transcription factors – PmeR, TtgR,

and NalC. These models helped reveal general insights into sequence determinants of promoter activity.

Insights and recommendations generated from these quantitative biology studies will in turn be beneficial for accelerating the bioengineering pipeline as well as improving the success rate for future rounds of biomolecular design.

References

1. Liu X *et al* (2019) De novo design of programmable inducible promoters. Nucleic Acids Research doi:10.1093/nar/gkz772
2. Gordon GC *et al* Genome-wide analysis of cyanobacteria RNA decay reveals highly stable transcripts encoding photosynthesis genes. (In revision)

This work was funded by the U.S. Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409).

Designing Mixed-Solvent Environments for Acid-Catalyzed Biomass Conversion Processes

Alex K. Chew[1,2]* (akchew@wisc.edu), Theodore W. Walker[1,2], George W. Huber[1], James A. Dumesic[1,2], and **Reid C. Van Lehn**[1,2]

[1]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI, 53706, USA

[2]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, 53706, USA

Project Goals:

The objective of this project is to develop biomass deconstruction and separation strategies that optimize C yields.

Abstract:

Lignocellulosic biomass is a promising renewable resource that can be converted into transportation fuels or commodity chemicals. Biomass conversion is facilitated by liquid-phase, acid-catalyzed dehydration and hydrolysis reactions that are hindered in aqueous solution by low reactivity and poor selectivity. One approach to improve biomass conversion efficiency is to modify the solvent composition by mixing water with *organic, polar aprotic cosolvents* (e.g. dioxane, tetrahydrofuran, etc.) to form mixed-solvent environments. Mixed-solvent environments have been shown to improve the reactivity of biomass conversion reactions by *100-fold* compared to the same reactions in pure water.[1] However, identifying an optimal solvent composition experimentally by trial-and-error is cost-prohibitive and lends little physical insight into how these mixtures will perform in new processes. Instead, *in silico* techniques can understand and predict the influence of solvent composition on experimental reaction rates and selectivities to guide solvent selection.

We developed classical molecular dynamics (MD) simulation methods to understand the effects of mixed-solvent environments on seven acid-catalyzed reactions involving biomass-derived reactants.[1] We found that inclusion of polar aprotic cosolvents leads to the formation of *water-enriched local domains* around *hydrophilic* reactants, which draws the acid catalyst to these regions due to preferred catalyst-water interactions[2] and results in improved reaction rates. By quantifying the extent of water-enrichment around the reactant, we found that MD measurables can accurately predict experimental reaction rates for dioxane-water mixtures, showing that classical simulation techniques can inform reaction rates without modeling the reaction mechanism or the catalyst.[1] We improved the predictive model by analyzing MD trajectories using three-dimensional convolutional neural networks (CNNs), which can capture complex spatial features that are related to reaction rates but are difficult to quantify by human experts. MD simulations in conjunction with CNNs were used to screen solvent compositions for the same seven reactions in

three water-cosolvent mixtures of varying composition (a total of 84 mixed-solvent environments).[3] Our results show that the CNN methodology accurately predicts experimental reaction rates for these and additional reactants/solvents not present in the initial training data, enabling reaction rate predictions for a design space that is too large to study with conventional quantum mechanical methods. Finally, we used MD simulations to understand the influence of solvent composition on experimentally determined product selectivities for representative acid-catalyzed reactions.[4] We synthesized these computationally efficient methods into a workflow for the selection of mixed-solvent environments by analyzing relevant biomass conversion reactions, such as the dehydration of fructose to 5-hydroxymethylfurfural (HMF).[5] We confirmed that the solvent system identified with this approach led to high HMF production experimentally. These computational models thus enable the rational design of new liquid-phase acid-catalyzed biomass conversion processes by guiding solvent selection.

References:

1. Walker, T. W.*, Chew, A. K.* *et al.* "Universal kinetic solvent effects in acid-catalyzed reactions of biomass-derived oxygenates." *Energy Environ. Sci.* (2018). doi:10.1039/C7EE03432F
2. Chew, A. K. & Van Lehn, R. C. "Quantifying the Stability of the Hydronium Ion in Organic Solvents with Molecular Dynamics Simulations." *Front. Chem.* (2019). doi: 10.3389/fchem.2019.00439.
3. Chew, A. K. & Van Lehn, R. C. et al. "Fast Predictions of Liquid Acid-Catalyzed Reaction Rates Using Molecular Dynamics Simulations and Convolutional Neural Networks" *ChemRxiv*. Preprint. https://doi.org/10.26434/chemrxiv.11299121.v1
4. Chew, A. K.*, Walker, T. W.*, et al. "Effect of mixed-solvent environments on the selectivity of acid-catalyzed dehydration reactions". *ACS Catalysis* (2020), *10*, 1679-1691.
5. Walker, T. W.*, Chew, A. K.* *et al.* "Rational design of mixed solvent systems for acid-catalyzed biomass conversion processes using a combined experimental, molecular dynamics and machine learning approach." *Submitted.*

Engineering of the Enzymes IspG and IspH from *Zymononas mobilis* to Increase Terpenoid Production

Isabel Askenasy[1,2]* (askeansyflor@wisc.edu), Jyotsna Misra[1,2], and **Patricia Kiley**[1,2]

[1] Great Lakes Bioenergy Center, University of Wisconsin-Madison, Madison; [2] Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison.

https://www.glbrc.org

Project Goals: Our goal is to engineer *Zymomonas mobilis* strains with increased terpenoid production. Terpenoids precursors in *Z. mobilis* are synthesized by the MEP pathway. In this pathway the last two steps, catalyzed by the enzymes IspG and IspH, present a bottleneck. We hypothesize that this is due in part to the stability of the iron-sulfur cluster located in each enzyme's active site. To improve IspG and IspH activity we propose to engineer these enzymes, as well as, the ability of *Z. mobilis* to produce and deliver iron-sulfur cluster cofactors to these enzymes.

Terpenoids can substitute for petroleum in the production of compounds of economic interest ranging from vitamins and perfumes to biofuels. Bacteria, such as *Zymomonas mobilis*, produce the terpenoid precursors dimethylallyl diphosphate (DMADP) and isopentenyl diphosphate (IDP) from glucose via the methyl erythritol phosphate (MEP) pathway. It has been proposed that the activity of the MEP pathway is limited at the last two steps, which are catalyzed by the enzymes IspG and IspH. Both enzymes carry a [4Fe-4S] cluster in their active sites, which, for some orthologs, makes them prone to O_2 damage. Additionally, recent data indicate that intermediates preceding IspG and IspH accumulate in response to O_2. Thus, our objective is to determine if Fe-S cluster lability explains the accumulation of IspG and IspH precursors in the presence of O_2 in *Z. mobilis* and if engineering approaches can improve the robustness of these enzymes. To achieve this goal, we are studying several aspects of both enzymes. First, to test the effect of O_2 on *Z. mobilis* IspG and IspH activity, we took advantage of *E. coli* strains that conditionally require IspG or IspH activity for growth. Using this *E. coli* strain, we have been able to replace the *E. coli* IspG and IspH enzymes with those from *Z. mobilis*. Under anaerobic conditions, *Z. mobilis* IspH and IspG complement the *E. coli* growth requirements for these enzymes but strains carrying the *Z. mobilis* IspH show poor growth under aerobic conditions. This result suggests that *Z. mobilis* IspH is sensitive to O_2. Current experiments are focused on isolating O_2-resistant variants of IspH and purifying the [4Fe-4S] form of IspH to test whether the Fe-S cluster is O_2-labile. We also found that co-expression of *Z. mobilis* IspG and IspH improves the growth of the *E. coli* strain lacking IspH, suggesting that *Z. mobilis* IspG and IspH might form a complex that protects their Fe-S clusters from O_2-mediated damage. Another approach to increase IspG and IspH protein activity is through protein overexpression. However, since the proteins require Fe-S cofactors, it may also be necessary to increase the production of the required cofactor to drive Fe-S cluster occupancy. Thus, we are also studying the function

and expression of the *Z. mobilis* Fe-S cluster biosynthetic machinery. In *Z. mobilis* the Fe-S cluster machinery is encoded by the *suf* operon. We have showed that expression of the genes encoding the *Z. mobilis* Suf pathway can replace the function of the *E. coli* pathway. We have also identified a homolog of the [2Fe-2S]-containing transcription factor, IscR, that regulates Fe-S cluster biosynthesis in *E. coli.* The *Z. mobilis* IscR homolog carries a [4Fe-4S] cluster that is highly sensitive to O_2. This protein binds to the promoter region of the *suf* operon *in vitro*. Currently, we are developing an *in vivo* reporter system that allows us to probe the role of the IscR-homolog in *Z. mobilis.* In summary, we predict that our multipronged strategy will generate new knowledge concerning the role of IspH Fe-S cluster stability in MEP pathway function and allow us to generate a strain of *Z. mobilis* with a more robust MEP pathway and improved terpenoid 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.

Engineering *Streptomyces* to Capture Value from Lignocellulosic Biofuel Conversion Residue

Caryn S. Wadler,[1,2]* (cwadler@wisc.edu), Kurt O. Throckmorton,[1,2] and **Michael G. Thomas**[1,2]

[1]University of Wisconsin, Madison; [2]Great Lakes Bioenergy Research Center, Madison, Wisconsin

Project Goals: To increase the economic viability of biofuels by generating fatty acid and isoprenoid bioproducts from the organics remaining after biofuel synthesis.

Current methods of switchgrass fermentation to bioethanol leave behind about 60% of the organic material in the hydrolysate after ethanol distillation. This material is referred to as conversion residue (CR). To increase the economic viability of lignocellulosic biofuels, we are engineering *Streptomyces* species to maximize the conversion of CR carbon into valuable isoprenoid and fatty acid bioproducts. From a library of 120 phylogenetically distinct *Streptomyces* isolates, more than half were capable of growth on undiluted, pH-adjusted CR. From these, we generated a collection of *Streptomyces* that produce the isoprenoid lycopene or fatty acid-derived melanin from CR as reporters to assess the production potential of these isolates. The genetic elements used in constructing these reporters are mobilizable between *Streptomyces* species and constructed using a combination of traditional cloning techniques and Golden Gate assembly to allow for rapid alterations in expression levels and the generated bioproduct.

Initial screens of the engineered *Streptomyces* reporter strains showed a wide range of native production levels of lycopene and melanin. We targeted four strains for further development using a design-test-learn approach to enhance bioproduct formation. These strains showed differences in carbon utilization on both CR and synthetic CR, a defined medium containing approximately one-third the organic compounds of CR at comparable concentrations. We will leverage insight into the mechanisms underlying these differences in carbon utilization from transcriptomic analyses and molecular modeling to increase CR catabolism and bioproduct formation. Other opportunities for further engineering include increasing isoprenoid precursor pools by introducing refactored MEP and mevalonate pathways and generating new bioproducts such as isoprene, limonene, pinene, and bisabolene.

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 Award Number DE-SC0018409.

Incorporation of protocatechuaic acid (3, 4-dihydroxybenzoate) conjugates into the lignin of transgenic poplar

Faride Unda[1,2*] (farideu@mail.ubc.ca), Elizabeth Mahon[1,2], Yaseen Mottiar[1,2], Kwang Ho Kim[3], John Ralph[2,4], Aymerick Eudes[5], **Shawn Mansfield**[1,2]

[1] University of British Columbia, Vancouver, Canada; [2] DOE Great Lakes Bioenergy Research Center, The Wisconsin Energy Institute, University of Wisconsin, Madison, [3] Korean Institute of Science and Technology Biorefinery Lab, University of British Columbia, Vancouver, Canada; [4] Department of Biochemistry, University of Wisconsin, Madison, USA; [5] Joint Bioenergy Institute, Emeryville, USA.

https://www.glbrc.org/

Project Goals:

The goal of the Bioenergy Plant Design team is to enhance plant biomass deconstruction, composition, and value. As part of this team, one of our specific project goals is the modification of lignin content and composition to improve biomass utilization in poplar. The strategy employed herein was the deviation of carbon upstream of the phenylpropanoid pathway; more specifically, we expressed the bacterial gene *QsuB* targeted to the plastid in attempts to produce protocatechuate. The overproduction of protocatechuate resulted in a reduction in total lignin content and the incorporation of protocatechuate conjugates into the lignin of transgenic poplar.

Current efforts to improve processing efficiency of lignocellulosic biomass are increasingly focused on modifying the lignin of economically and ecologically important plant species. Gain of function approaches that not only reduce cell wall recalcitrance but also add value to the biomass have become a core strategy in the development of more cost-effective processing methods. Lignin is a phenolic polymer found in the secondary cell walls of xylem vessels and fibres, where it plays a key role in facilitating water transport and providing structural support to the plant. Principally, three canonical *p*-hydroxycinnamoyl alcohols (monolignols) compose lignin: *p*-coumaryl, coniferyl, and sinapyl alcohols. However, it is now known that lignin can also include monomeric units outside the traditional monolignols, such as monolignol ester conjugates: *p*-hydroxybenzoate, *p*-coumarate, or ferulate, as well as caffeyl alcohol and 5-hydroxyconiferyl alcohol, and phenolic compounds derived from outside the phenylpropanoid pathway such as flavonoids and hydroxystilbene glucosides. Recently, xylem targeted expression of a bacterial *3-dehydroshikimate dehydratase* (*QsuB*) in Arabidopsis led to the conversion of 3-dehydroshikimate, a precursor in the shikimate pathway, into protocatechuate (3, 4-dihydroxybenzoic acid). This modification specifically diverts carbon away from the phenylpropanoid pathway, resulting in a reduction in total lignin and improved saccharification

efficiency (Eudes *et al*., 2015). Based on this initial study, our research focused on examining the effects of xylem targeted expression of *QsuB* in hybrid poplar (*P. alba* x *grandidentata*). We showed that hybrid poplar expressing *QsuB* accumulated significant amounts of soluble protocatechuaic acid (PA), but largely in the glycosylated form. Additionally, transgenic trees displayed significantly decreased total lignin, and altered monomeric composition with an increased accumulation of H subunits and a concurrent alteration in S and G unit ratio. Moreover, the saccharification yields of QsuB poplar increased by as much as 40% compared to wild-type trees. Interestingly, in-depth analysis of the *QsuB* transgenic lignin showed that conjugated forms of PA were polymerized to the lignin as ester linked pendant groups, as alkaline hydrolysis (saponification) of extracted cell wall material released both phenolic glycosides and free protocatechuic acid conjugates. These protocatechuate conjugates may therefore also be incorporated in the backbone of the lignin polymer forming a new "zip lignin".

References

1. Eudes A, Sathitsuksanoh N, Baidoo EE, George A, Liang Y, Yang F, Singh S, Keasling JD, Simmons BA, Loqué D. (2015) Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. Plant Biotechnol J.13(9):1241-50.

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.

Integrated Spatially Explicit Optimization of Biofuel Supply Chains and Landscape Design Considering Biomass Supply Uncertainty

Eric O'Neill,[1,2*] (egoneill2@wisc.edu), Tyler Lark,[1,3] Bruno Basso,[1,4,5] and **Christos Maravelias**[1,2]

[1]Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, WI; [2]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI; [3]Center for Sustainability and the Global Environment, University of Wisconsin-Madison, WI; [4]Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI; [5]W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI

Project Goals: To develop a generalized spatially explicit biofuel supply chain stochastic optimization model that accounts for biomass supply uncertainty.

The overall environmental and economic performance of biofuel supply chains (SC) is highly sensitive to the spatial distribution of available land for biomass production. Because large quantities of dedicated bioenergy crops have yet to be planted, there is an opportunity to use mathematical programming techniques to optimize the biofuel supply chain network design. We modeled the supply chain network design problem as a two stage stochastic mixed integer linear program which takes into account the uncertainty in biomass supply as a set of discrete scenarios with known probability. Given (i) the spatially explicit available land, (ii) biomass yield potential at each site, (iii) potential biorefinery and preprocessing depot locations, (iv) potential preprocessing and conversion technologies, (v) cost parameters and greenhouse gas (GHG) emission parameters. The model finds the optimal (i) biorefinery and depot locations, (ii) technology, (iii) capacity, (iv) transportation, production, and inventory planning, (v) crop establishment locations, (vi) and land management; the combination of which minimizes the total annualized cost of the supply chain. Model complexity is sensitive to the number of uncertainty scenarios, the size of the study area, and the number of supply chain nodes considered. As model complexity increases, approximations that sacrifice model accuracy for computational tractability are introduced to solve larger problems. A case study is performed in southern Michigan. Our results show that by adjusting the cost of carbon ($/MgCO2e), decision makers can decide how much to value reductions in GHG emissions with respect to SC costs. Also, by considering uncertainty, the performance of optimal supply chain configurations is more robust to the realization of uncertain biomass yields. We show that when designing biofuel supply systems it is important to consider the environmental and economic tradeoffs between land management, supply chain operation, and capital decisions and how these decisions guard against system disruptions and uncertainty.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE--FC02--07ER64494)

GREAT LAKES BIOENERGY RESEARCH CENTER

Microbial conversion of chemically depolymerized lignin into valuable compounds

Miguel Perez,[1,2] Wayne Kontur,[1,2] Steven D. Karlen[1,2*] (skarlen@wisc.edu), German Umana,[1,2] Yanjun Ma,[1,2] Carson Gehl,[1,2] Yanding Li,[1,2] Canan Sener,[1,2] Jason Coplien,[1,2] John Ralph,[1,2] Daniel R. Noguera,[1,2] **Timothy J. Donohue**[1,2]

[1]DOE Great Lakes Bioenergy Research Center; [2]University of Wisconsin-Madison

https://www.glbrc.org/research/efficient-biomass-conversion

Project Goals: Develop microbes that funnel *p*-hydroxybenzoic acids, and *p*-hydroxycinnamic acids to value added products.

Abstract: Plant cell wall consists mainly of a combination of two major polymers: polysaccharides and lignin. Lignin is a heteropolymer of different types of aromatic compounds whose chemical properties make it highly insoluble, and recalcitrant, to chemical and biological degradation. This presents a major challenge to full conversion of lignocellulosic biomass into a portfolio of value added products. Chemical deconstruction of lignin produce complex mixtures of aromatic compounds with structures derived from mainly from three phenolic subunits: Syringyl (S), Guaiacyl (G), and *p*-Hydroxyphenyl (H). Some bacterial strains have been shown to utilize multiple lignin-derived aromatic compounds as their source of carbon and energy for growth. In this process they metabolize a variety of aromatic compounds by funneling them through common intermediates. This natural capability presents an attractive opportunity for upgrading aromatic compounds via metabolic engineering of suitable strains. We have shown that the α-proteobacterium *Novosphingobium aromaticivorans* DSM12444 has the ability to simultaneously catabolize multiple S, G, and H type aromatic compounds known or predicted to be present in chemically depolymerized lignin from a wide range of different biomass feedstocks. Using selected mutations, we engineered the bacterium and created a mutant strain capable of transforming aromatic compounds containing S, G, and H type substructures into a single product, 2-pyrone-4,6-dicarboxylic acid (PDC). In addition, we studied the utilization of the engineered microbe to funnel lignin-derived aromatic compounds from different biomass types into PDC in the context of an integrated lignin-to-bioproduct processing chain. We show that products of reductively depolymerized lignin from wild type poplar, sorghum, switchgrass, and maple, as well as from plants containing mutations that alter lignin biosynthesis (in collaboration with researchers in the Joint Bioenergy Institute) can be used as substrates for PDC production by this engineered *N. aromaticivorans* strain.

References

1. M. Perez, W. Kontur, M. Alherech, S. D. Karlen, S. Stahl, T. J. Donohue, D. R. Noguera, Funneling aromatic products of chemically depolymerized lignin into 2-pyrone-4-6- dicarboxylic acid with *Novosphingobium aromaticivorans*, *Green Chem* 2019, **21**, 1340-1350, doi:10.1039/C8GC03504K.
2. S. D. Karlen, P. Fasahati, M. Mazaheri, R. A. Smith, C. L. Cass, S. Sirobhushanam, M. Chen, V. Tymokhin, S. Liu, D. Padmakshan, J. Serate, D. Xie, Y. Zhang, M. Mcgee, C. E. Foster, J. D. Russell, J. J. Coon, H. Kaeppler, N. de Leon, C. T. Maravelias, T. M. Runge, S. M. Kaeppler, J. C. Sedbrook, J. Ralph, Assessing the viability of recovering hydroxycinnamic acids from lignocellulosic biorefinery alkaline pretreatment waste streams, *ChemSusChem* **2020**, *in press*, doi:10.1002/cssc.201903345R1.

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.

Microbial Valorization of Lignin: Using *Novosphingobium aromaticivorans* to Break the Bonds in Lignin and Convert Lignin Deconstruction Products into Value-added Chemicals

Wayne S. Kontur[1,2*] (wskontur@wisc.edu), J. Miguel Perez[1,2], **Daniel R. Noguera,[1,2]** and **Timothy J. Donohue[1,2]**
[1]University of Wisconsin-Madison; [2]U.S. DOE Great Lakes Bioenergy Research Center, Madison, WI

Project Goal:
To develop the bacterium *Novosphingobium aromaticivorans* DSM 12444 into a model system for understanding how bacteria break the bonds between subunits in lignin and metabolize aromatic lignin deconstruction products, and into a platform for converting lignin deconstruction products into value-added chemicals

Lignin is a heterogeneous polymer of aromatic subunits that is a major component of lignocellulosic plant biomass (comprising up to ~25% of its dry weight). In biomass processing facilities that primarily focus on converting the sugar-rich biomass fractions (cellulose and hemicellulose) into value-added commodities, lignin is currently typically burned for heat and power generation. While this utilizes the chemical energetic value of lignin, it wastes lignin's potential as a renewable resource for value-added chemicals that are currently derived from petroleum (such as fine chemicals, food additives, plastic precursors, pharmaceuticals, etc.). To help improve the overall economic potential of generating fuels and other renewable commodities from lignocellulosic plant biomass, we are therefore investigating ways to generate value-added chemicals from lignin using microbes.

The bacterium *Novosphingobium aromaticivorans* DSM 12444 is a very promising model system for valorizing lignin, due to the following characteristics:
-It efficiently metabolizes all lignin-derived aromatic compounds that it has been tested for growth on
-It can naturally produce some potential value-added chemicals from lignin, such as the flavoring agent vanillin and the epoxy/plastic precursor 2-pyrone-4,6-dicarboxylic acid (PDC)
-It can metabolize glucose, which raises the possibility that a strain could be engineered to derive growth and energy from glucose, while converting all or most of the aromatic component of plant biomass into a desired commodity

We have been working on characterizing both the microbe's ability to break the bonds between aromatic subunits in the lignin polymer and its ability to metabolize the three major classes of mono-aromatic compounds that result from lignin's deconstruction (the syringyl (S), guaiacyl (G), and p-hydroxyphenyl (H) compounds). As part of this work, we have identified and characterized enzymes involved in both processes, such as those responsible for breaking the β-

aryl ether bond, the most common inter-subunit bond in natural lignin, and those responsible for opening the aromatic ring of the mono-aromatic constituents.

Our understanding of the metabolic pathways of *N. aromaticivorans* has enabled us to develop a strain which can convert S, G, and H compounds into PDC at nearly the maximum theoretical yields. In addition, we are using what we have learned from PDC production to investigate developing strains of the microbe that can produce additional value-added compounds from lignin.

This work was supported by a U.S. Department of Energy (DOE) Great Lakes Bioenergy Research Center grant (DOE Office of Science BER DE-SC0018409).

Model-driven analysis of mutant fitness experiments improves genome-scale metabolic models of *Zymomonas mobilis* ZM4

Dylan K. Courtney[1,2*] (dcourtney2@wisc.edu), Wai Kit Ong,[1,2] Shu Pan,[1,2] Ramon Bonela Andrade, [1,2] Patricia J. Kiley,[1,3] Brian F. Pfeger,[1,2] **Jennifer L. Reed**[1,2]

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison; [2]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison; [3]Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison

http://glbrc.org

Project Goals: Our goal is to establish an updated, accurate genome-scale metabolic reconstruction of the bacteria *Zymomonas mobilis* by leveraging high-throughput experimental datasets to fill in knowledge gaps in the organisms's metabolism. This reconstruction will serve both as a knowledge base of our understanding of *Z. mobilis* metabolism and be utilized to enable the design of strains for primary biofuel synthesis as part of the Great Lakes Bioenergy Research Center's larger goals.

Zymomonas mobilis is an industrially relevant, Gram-negative, ethanologen known for high glycolytic fluxes through the Entner Doudoroff pathway, high ethanol production, and exceptionally low biomass yields. Here we present *i*ZM4_478, a genome-scale stoichiometric model of *Z. mobilis* ZM4 metabolism and apply it to analyze a published dataset from pooled mutant fitness experiments. *i*ZM4_478 contains 752 metabolic and transport reactions (of which 625 have gene-protein-reaction associations), 478 genes, and 616 unique metabolites, making it one of the most complete models of *Z. mobilis* ZM4 to date. Model predicted essential genes were compared to fitness data from the pooled mutant experiments. Several discrepancies between the model and dataset were found to be caused by polar effects, mismapped barcodes, or heterozygous mutants, highlighting potential challenges inherent to analyzing these high-throughput datasets. Functionally related modules of reactions and genes in the model were identified via flux coupling analysis. The fitness scores across all 492 experiments in the reported dataset were analyzed in the context of these modules to identify candidate genes for a reaction in histidine biosynthesis lacking an annotated gene and highlight metabolic modules where the fitness scores of mutants in the dataset are poorly correlated. Additional genes for reactions involved in biotin, ubiquinone, and pyridoxine biosynthesis in *Z. mobilis* were identified and confirmed using mutant complementation experiments. These newly identified genes improve our understanding of *Z. mobilis* metabolism, and the updated model provides a platform for future network driven studies of this organism and serves as a starting point for the development of kinetic models of *Z. mobilis* metabolism in the coming years of the Great Lakes Bioenergy Research Center

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.

MPK6-MYB46 Regulatory Module Suppresses Plant Biomass Formation During Salt Stress

Jong Hee Im[1,5], Jae-Heung Ko[3], Won-Chan Kim[1,4,5], and Kyung-Hwan Han[1,2,5]*
(hanky@msu.edu)

[1] Department of Horticulture, Michigan State University, East Lansing, MI 48824
[2] Department of Forestry, Michigan State University, East Lansing, MI 48824
[3] Department of Plant & Environmental New Resources, Kyung Hee University, Korea
[4] School of Applied Biosciences, Kyungpook National University, Korea
[5] DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan, 48824

Project Goal:

Overarching goal of this project is to develop designer bioenergy crops for sustainable production of biomass feedstock under various growth conditions.

Abstract:

Secondary cell walls, located between the plasma membrane and primary cell wall, are a defining feature of xylem fibers and vessels, providing mechanical support for plants and serving as a conduit for long-distance transport of water and solutes. They constitute the vast majority of plant biomass and are of economic importance to humans as fiber, animal feed, pulp for manufacture of paper, and as an environmentally desirable, cost-effective, renewable source of energy. The biosynthesis of secondary walls occurs in a highly-coordinated manner by successive encrustation and deposition of cellulose fibrils, hemicelluloses and lignin upon cessation of cell growth. This process requires a coordinated transcriptional activation of the biosynthetic genes for the components, suggesting the existence of one or more central transcriptional regulators. The plant specific R2R3-MYB transcription factor MYB46 functions as a master switch for secondary cell wall biosynthesis, ensuring the exquisite expression of the secondary wall biosynthetic genes in the tissues where secondary walls are critical for plant growth, such as the stem. However, suppression of MYB46 function is needed during environmental stresses that trigger nascent defense responses including impermanent cessation of vegetative growth. Little is known about how this opposing control of secondary cell wall formation is achieved with the speed and specificity of plant response to environmental changes. Post-translational modification of MYB46 may offer a such regulatory mechanism. MYB46 has two conserved mitogen-activated protein kinase (MPK) phosphorylation target sites, suggesting that MYB46 is a substrate for phosphorylation by MPKs. While phosphorylation of transcription factors is well known to modulate their levels and activities, no evidence has been shown for post-translational regulation of secondary cell wall biosynthesis. Here, we show that MYB46 is phosphorylated by abiotic stress-activated MPK6 and subsequently degraded by the proteasome pathway. This MPK6-MYB46 regulatory module provides novel insights into the tissue- and/or condition-specific activity of MYB46, and the interplay of secondary wall formation and environmental signaling.

Pathway Engineering and Re-targeting Boosts Production of High-Value Bioproducts in Plants

Jacob D. Bibik[1,2]* (bibikjac@msu.edu), Aparajita Banerjee[1,2], and **Björn Hamberger**[1,2]

[1]Department of Biochemistry and Molecular Biology and [2]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI

Leveraging synthetic biology approaches, engineered plants offer a sustainable production platform for high-value chemicals and other bioproducts. Squalene, a C30 hydrocarbon, is a biofuel candidate and the precursor to high-value triterpenoids, a diverse class of natural products with applications in the health, cosmetic, and other biotechnological industries. In this work, two strategies have been developed to increase plant production yields of squalene and triterpenoids by hijacking existing cell structures or building novel structures to sequester products within cells. First, the optimized biosynthetic pathways can be re-targeted to plastids, natural intracellular compartments, where biosynthesis can occur separate from native competing enzymes (Figure 1, right). The second strategy re-engineers lipid droplets as synthetic storage organelles with biosynthetic enzymes anchored to the surface, synthesizing and storing products in the same location (Figure 1, left). Both strategies are being implemented in poplar, a biofuel feedstock, to increase overall crop value through the addition of squalene bioproduction. These strategies demonstrate effective metabolic engineering approaches to further develop plants as platforms for production of high-value bioproducts.

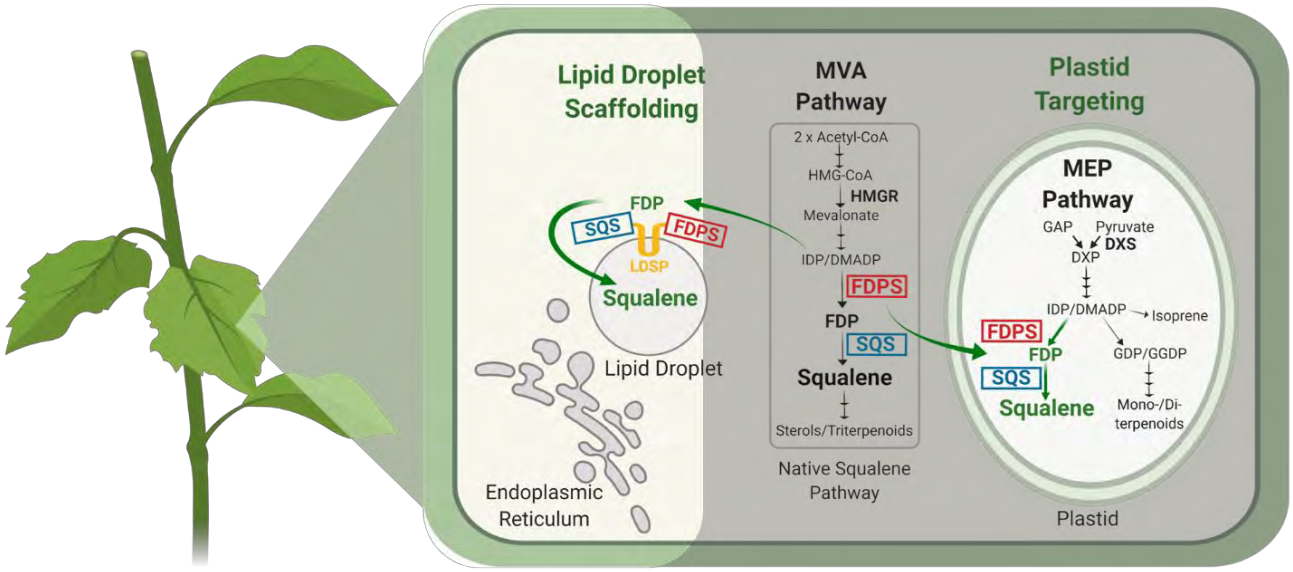


Figure 1: Strategies of re-targeting key squalene biosynthetic enzymes from the native pathway (center) to either plastids (right) or anchored to synthetic lipid droplets (left).

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.

GREAT LAKES BIOENERGY RESEARCH CENTER

Sorghum Dw2 controls stem growth by regulating PLDδ/endomembrane activity and cell proliferation

Joel Oliver[1]*(joel.aoliver@tamu.edu), Mingzhu Fan[3,4], Brian McKinley[1], Starla Zemelis-Durfee[2,4], Federica Brandizzi[2,3,4], Curtis Wilkerson[3,4,5], and **John E Mullet**[1]

[1]Department of Biochemistry & Biophysics, Texas A&M University, College Station, Texas 77843; [2]MSU-DOE Plant Research Lab, Michigan State University, East Lansing, Michigan, 48824; [3]Department of Plant Biology, Michigan State University, East Lansing, Michigan, 48824; [4]Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan, 48824; [5]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824

https://www.glbrc.org/

Project Goals: The C4 grass *Sorghum bicolor* has shown promise as a dedicated bioenergy crop due to its ability to accumulate large amounts of biomass within its stem with low inputs on marginal lands. The goal of this project is to identify the molecular pathways responsible for controlling stem growth in Sorghum to aid in the continued breeding and engineering of *Sorghum* for greater biomass yield and better biomass composition.

Abstract

Stems of the C4 grass sorghum are comprised of nodes and internodes that are produced and elongated sequentially during plant growth impacting sink strength, biomass yield and composition. Stems of high biomass bioenergy sorghum can accumulate up to 50 internodes that together span 4-5 meters and account for ~84% of harvested biomass. An AGCVIII kinase (*Dw2*) has been identified that regulates sorghum stem internode growth, but the underlying signaling network is unknown. Here we provide evidence that mutation of *Dw2* reduces cell proliferation in internode intercalary meristems, inhibits endocytosis, and alters the distribution of heteroxylan and mixed linkage glucan in cell walls. Phosphoproteomic analysis showed that Dw2 signaling modulates the phosphorylation of proteins involved in lipid signaling (PLD δ), endomembrane trafficking, hormone, light and receptor signaling, and photosynthesis. Together, our results show that Dw2 modulates endomembrane function and cell division during sorghum internode elongation providing insight into the regulation of monocot internode development.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

Stability of Switchgrass Leaf Microbiome in the Face of Natural Aerial Colonizers

Lukas Bell-Dereske[1,2]* (belldere@msu.edu), and **Sarah Evans**[1,2,3,4]

[1]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI,[2]W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI, [3]Department of Integrative Biology, Michigan State University, East Lansing, MI, [4]Department of Microbiology & Molecular Genetics, Michigan State University, East Lansing, MI

https://www.glbrc.org/research/sustainable-cropping-systems
https://rhizosphere.msu.edu

Project Goals: We are interested in exploring the role microbial communities play in sustainable growth of dedicated bioenergy crops and investigating the role of aerial dispersal on the assembly and stability of the switchgrass (*Panicum virgatum*) leaf microbiome. We will use this information to inform efforts to develop fungal inocula that can colonize and persist in the endosphere of switchgrass, with the potential to improve switchgrass disease resistance and drought tolerance.

Plant-microbial interactions are critically important for the sustainable production of cellulosic energy crops. While the majority of previous research into plant-microbial interactions has focused on the root and rhizosphere community, plants also host a diverse leaf microbial community. This leaf microbiome has the potential to be utilized to improve host disease resistance, drought tolerance, and other benefits improving the yield of cellulosic energy crops. However, past efforts to manipulate and inoculate leaves with beneficial microbial communities have had little success, perhaps because communities establish from seed microbes and are difficult to invade. For this reason, it is important to understand what are the sources of the microbes that colonize the leaf community and the stability of this community in the face of outside, natural, colonists. We established two experiments to test the effects of aerial microbial colonists on leaf microbiome development of switchgrass (*Panicum virgatum*). To determine prevalence and magnitude of natural aerial dispersal into leaf microbiomes, we germinated and grew switchgrass seedlings in pots with sterile soil and placed them near mature switchgrass field monocultures. Seedlings were left in the field for 51 days, and we collected rain and dry deposited microbes throughout the period to characterize the abundance and composition of the aerial fungal colonists. We found that the fungal community inhabiting the endosphere of seedling leaves (endophytes) showed rapid and significant shifts that were driven primarily through enrichment of species found in the aerial community. Interestingly, the fungal endophyte community of the mature switchgrass leaves did not significantly differ from the seed microbiome but did host many aerial dispersed species. In our second experiment we tested the direct effects of single rain events on the development of the leaf microbiome of germinating seedlings. We germinated and grew seedlings for ~24 days in sterilized sealed petri dishes inoculated with either live rain, autoclave sterilized rain, or nanopure water. Seedlings that germinated in live rain showed significant shifts in the fungal endophyte community compared

to the starting seed microbiome but seedlings grown in both sterile treatments showed little to no shift in the leaf endophyte community, suggesting an active interchange between the leaf microbiome and the aerial community. In further support of this, we found that the temporal turnover of the aerial fungal community strongly tracked the seasonal development of switchgrass across the year and a half of rain events under analysis. Our results suggest that the leaf microbiome of seedlings has greater sensitivity to these aerial colonists while the mature endophyte community is more difficult to invade. The invadability of seedling leaf microbiome makes it a great target for inoculation of beneficial microbes, but this flexibility also suggests that it may be necessary to reinforce inoculations with repeated applications.

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-FC02-07ER64494 in addition to Award Number DOE DE-FOA-0001207 to the MMPRNT project.

Systems Level Comparison of Medium Chain Fatty Acid Production

Nathaniel W. Fortney[1*](nfortney@wisc.edu), Kevin A. Walters[1, 2], Matthew J. Scarborough[3], Kevin S. Myers[1], Abel T. Ingle[4], Timothy J. Donohue[1,2], **Daniel R. Noguera**[1,4]

1 The Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI
2 Department of Bacteriology, University of Wisconsin-Madison, Madison, WI
3 Department of Civil and Environmental Engineering, University of Vermont, Burlington, VT
4 Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI

https://www.glbrc.org/

Project Goals: Our aims are to (1) elucidate the genetic elements that enable members of the Clostridia class of Firmicutes to produce medium chain fatty acids (MCFAs) via reverse beta oxidation and (2) identify environmental conditions that maximize MCFA production by anaerobic microbial communities grown on complex organic residues from different industrial processes.

Medium chain fatty acids (MCFAs), 6-12 carbon saturated monocarboxylic acids, are high-value compounds that can be produced from a variety of industrial residues by fermentative microbial communities. Residues from lignocellulosic biorefineries, starch ethanol plants, and the dairy industry are examples of carbon-rich residues considered low-value co-products that are typically sent to anaerobic digesters for biogas generation [1] or concentrated and sold as animal feed [2]. Following the model of the petroleum industry, diversifying product formation from the primary feedstock can help offset operating costs, reduce the selling point of the primary products (e.g. biofuel), and ultimately make these industries more economically viable [3].

MCFA-producing microbial communities contain microorganisms predicted to perform two main general functions within the communities [4]. One set of organisms hydrolyses and ferments energy-rich substrates such as carbohydrates (monomeric and oligomeric) to low-carbon fermentation products such as acetate, ethanol, and lactate, or to low-carbon intracellular metabolites such acetyl-CoA and propionyl-CoA, while the second group uses these fermentation products to produce higher-carbon products such as MCFA by “chain-elongation” via the reverse beta-oxidation pathway.

We are interested in elucidating the genetic elements necessary for MCFA production by chain elongation to enable accurate predictions of MCFA production using genome-scale metabolic models of Clostridia. We have characterized and described two MCFA-producing Clostridia, *Candidatus* Weimeria bifida, gen. nov., sp. nov., and *Ca.* Pseudoramibacter fermentans, sp. nov. [5], which were enriched in a continuously stirred tank reactor (CSTR) fed conversion residue from a lignocellulosic biorefinery and operated with a 6-day residence time, pH 5.5, and 35°C. Metatranscriptomic analyses of the microbial community predicted that these two organisms had

high expression of gene products in the reverse beta-oxidation pathway, including an electron bifurcating Acyl-CoA dehydrogenase (ACD) and associated electron transfer flavoproteins (EtfA, EtfB). In addition, both organisms were predicted to have energy conserving mechanisms via ion motive force generation using the RNF and the energy conserving hydrogenase (Ech) complexes. The metatranscriptomic data also predicted a difference in preferred organic substrates for both organisms. Whereas *Ca*. W. bifida had high transcript abundance for genes involved in carbohydrate metabolism, *Ca.* P. fermentans had high transcript abundance for genes associated with lactate and glycerol utilization. These primary metabolic features were sufficient to assemble metabolic models representing MCFA production from carbohydrates and lactate, respectively [4].

The ability to enrich for MCFA-producing Clostridia from the same inoculum source (acid digestion bioreactor from wastewater treatment plant) but fed different organic-rich substrates is currently under investigation. Identification of microorganisms using 16S rRNA gene amplicon sequencing of communities enriched on a synthetic xylose-rich medium, on ultra-filtered milk permeate (UFMP) coproducts from the dairy industry, on lignocellulosic biorefinery residues, and on thin stillage (TS) from a starch ethanol biorefinery has revealed high abundance of Clostridia in all enrichments. The synthetic xylose-rich substrate resulted in the enrichment of a *Ca.* Weimeria strain; the UFMP substrate enriched for organisms related to the recently defined *Agathobacter* genus [6] within the Lachnospiraceae and members of the Clostridiales_Incertae Sedis XIII; the lignocellulosic residues enriched for strains of *Ca.* Weimeria and *Pseudoramibacter*; the TS enriched for *Butyrivibrio, Dialister, Pseudoramibacter,* and Lachnospiraceae in the class Clostridia. Metagenomic analyses of these microbial communities is underway.

Comparative genomic analyses of these enriched members of the Clostridia class of Firmicutes, along with the available genomes for related Clostridia will help us improve metabolic models for MCFA production, which will guide future investigations on how to optimize MCFA yields with either self-assembled or synthetically-created microbial communities.

References

1. **Gerbrandt K**, **et al.** 2016. Curr Opin Biotechnol **38**:63–70. doi:10.1016/j.copbio.2015.12.021
2. **Lupitskyy R**, **et al**. 2015. Biomass and Bioenergy **72**:251–255. doi:10.1016/j.biombioe.2014.10.029
3. **Scarborough MJ et al.** 2018. Biotechnol Biofuels **11**:1–17. doi:10.1186/s13068-018-1193-x
4. **Scarborough MJ**, **et al.** 2019. bioRxiv **11**:200–36. doi:10.1101/759548
5. **Scarborough MJ**, **et al.** 2019. Appl Environ Microbiol 1–44. doi:10.1128/AEM.02242-19
6. **Rosero JA, et al**, 2016. Int J Syst Evol Micr **66**:768–773. doi: 10.1099/ijsem.0.000788

This material is based upon work supported by the DOE Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409, the National Institute of Food and Agriculture (USDA 2017-67003-25055), and the National Dairy Council (Project #AA68952).

Using Fungal Diversity to Improve Biofuel Conversion

David J. Krause (dkrause2@wisc.edu),[1,2] Ritika Punathil,[1,2] and **Chris Todd Hittinger (cthittinger@wisc.edu)[1,2]**

[1]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53726; [2]Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI 53706

Project Goals: Our major goal is to improve the conversion of sugars from lignocellulosic hydrolysates into specialty biofuels such as isobutanol. We are engineering non-traditional model species of budding yeast to develop an improved understanding of how flux can be directed from the typical ethanol fermentation into new pathways such as isobutanol production. We also aim to use budding yeast diversity to better understand adaptations to growth and production of biofuels under anaerobic conditions.

The production of specialty biofuels from dedicated bioenergy crop hydrolysates will require genetically engineered microbes capable of rapidly converting sugars into biofuels in an oxygen-free environment. The budding yeasts are a diverse group of microorganisms containing many species that have these key traits. We are currently using budding yeast diversity to address several key challenges to producing the specialty biofuel isobutanol in industrial conditions. First, the primary model yeast species, *Saccharomyces cerevisiae* is unable to grow on glucose when lacking genes encoding pyruvate decarboxylase, which is a barrier to completely transitioning this organism from ethanol production to isobutanol production. In contrast, the yeast *Kluyveromyces lactis* is capable of growth in the absence of any ethanol production machinery. We will present progress in engineering *K. lactis* for isobutanol production in the absence of ethanol production. One drawback to *K. lactis* as a biofuel organism is its inability to grow under strict anaerobic conditions. We are currently screening over 1,000 species of budding yeasts for anaerobic growth. These experiments will improve our understanding of how anaerobic growth evolved within the budding yeast subphylum Saccharomycotina, as well as shed light on how we might engineer obligate aerobic yeast species to grow anaerobically. By leveraging genetic diversity among yeasts, as well as performing genetic engineering in non-traditional model yeast species, we are employing novel strategies to overcoming barriers to specialty 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 Numbers DE-SC0018409 and DE-FC02-07ER64494. C.T.H. is a Pew Scholar in the Biomedical Sciences and Vilas Faculty Early Career Investigator, supported by the Pew Charitable Trusts and the Vilas Trust Estate, respectively.

Using the Zip-Lignin Strategy to Build the Optimal Sorghum Biofuel Crop

Rebecca A. Smith,[1*](rasmith29@wisc.edu), Emily Beebe,[1] Cynthia Cass,[2] Craig Bingman,[1] Kirk Vander Meulen,[1] John Sedbrook,[2] Steven D. Karlen,[1] Brian G. Fox,[1] and **John Ralph**[1]

[1]University of Wisconsin-Madison, Madison, WI; [2]Illinois State University, Normal, IL

https://www.glbrc.org/research/sustainable-cropping-systems

Project Goals: Increase the levels of zip-lignin in *Sorghum bicolor* plants to improve the cell wall digestibility.

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 in the form of '*clip-off*' phenolic acids. *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". Our previous work in *Brachypodium distachyon* and *Zea maize* has demonstrated that there is competition between different monolignol transferase enzymes for substrates, and accumulating pools of substrates for the enzymes is important for maximizing monolignol transferase activity. In Brachypodium, knocking out the native *PMT* gene and introducing an *FMT* gene resulted in the highest detectable levels of monolignol ferulates that we have observed to date. The level of monolignol ferulates was significantly higher than in the plants that only had increased FMT activity. In maize, accumulation of the FMT substrate feruloyl-CoA through knock-down of a lignin biosynthetic gene, *CINNAMOYL-CoA REDUCTASE* (*CCR*), also significantly increased the production of zip-lignin and improved the cell wall digestibility (Smith et al., 2017). We hypothesize that the combination of knocking out the native *PMT* gene and overexpressing the native *FMT* in *Sorghum bicolor* in the *CCR* down-regulated background will yield Sorghum lines with the highest potential as bioenergy crops.

The *Sorghum bicolor* FMT and PMT enzymes were unknown, and therefore we used phylogenetics to discover potential FMT and PMT enzymes from Sorghum based on their similarity to previously identified rice FMT and PMT enzymes. The enzymes were synthesized using the wheatgerm cell-free translation system and tested for monolignol transferase activity. Based on these results, we have identified putative FMT and PMT enzymes in Sorghum and have compared their activities to those of known monolignol transferases. These putative *FMT*

and *PMT* genes encoding the enzymes were transformed into *Arabidopsis thaliana* to test their activities and abilities to biosynthesize monolignol conjugates for lignification *in planta*. Arabidopsis does not naturally produce monolignol conjugates, which simplifies the detection of the novel compounds. The presence of monolignol ferulates and monolignol *p*-coumarates on the lignin of these transformants indicated that the targeted FMTs and PMTs are acting as functional, and efficient, feruloyl-CoA and *p*-coumaroyl-CoA monolignol transferases within plants. Constructs have been developed and transformed into *Sorghum bicolor* to overexpress the native *FMT*, knock-down the native *PMT* using CRISPR-Cas9 technology, and knock-down the native lignin *CCR* gene using CRISPR-Cas9. When the best lines for each of these transformants have been established, the lines will be crossed to generate an *FMT* overexpression/*PMT* CRISPR/*CCR* CRISPR triple-transgenic line.

References

Smith, R.A., Cass, C.L., Mazaheri, M., Sekhon, R.S., Heckwolf, M., Kaeppler, H., de Leon, N., Mansfield, S.D., Kaeppler, S.M., Sedbrook, J.C., Karlen, S.D., and Ralph, J. (2017). Suppression of CINNAMOYL-CoA REDUCTASE increases the level of monolignol-ferulates incorporated into maize lignins. Biotechnol Biofuels **10,** 109: 101-110.

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.

Influence of microbial surface litter decomposer communities on CO_2 emissions from natural soils

Sanna Sevanto[*1], Rae DeVan[2], John Heneghan[1], Brent Newman[1], M. Francesca Cotrufo[3], Michaeline Albright[2], John Dunbar[2]

[1]Earth and Environmental Science Division, Los Alamos National Laboratory, Los Alamos, New Mexico

[2]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico

[3]Department of Soil and Crop Science, Colorado State University, Fort Collins, Colorado

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

Project Goals

1) 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.

2) Establish predictive links between carbon emissions, DOC production, and carbon transport to deeper soil layers during surface litter decomposition.

Abstract

Respiration flux resulting from litter decomposition is one of the major components counteracting reduction of atmospheric CO_2 content by plant carbon uptake in natural ecosystems. Respiration fluxes can range from 10 to 50% of the ecosystem-scale gross primary production, and turn ecosystems from carbon sinks to carbon sources when plant carbon uptake is hindered by natural disturbances such as drought or unusually warm winters at high latitudes where sunlight is too low for photosynthetic production. In the litter decomposition process, as soil microbes metabolize litter, some of the carbon is immediately released into the atmosphere as CO_2, while some stays in the soil in other forms including dissolved organic carbon (DOC). DOC that travels to deeper layers of the soil is a significant contributor to soil carbon stocks. The LANL Microbial Carbon Cycling SFA has previously identified microbial decomposer communities that differ in the amounts of CO_2 and DOC produced during early phases of plant litter decomposition in controlled microcosm systems. In this experiment, our goal was to test the impact of microbial composition on carbon flow in a more complex and realistic system. To do this, we inoculated natural, arid-soil cores containing ^{13}C-labelled blue

grama grass (*Bouteloua gracilis*) litter with microbial communities previously shown to produce either high or low DOC extremes. Our preliminary results show increased CO_2 flux from soil cores inoculated with the high DOC communities compared to those inoculated with low DOC communities or controls. From a carbon balance perspective this suggests that CO_2 and DOC production during litter decomposition are not always inversely correlated, but can also be additive. We expect that further data analysis will reveal if the additive effect was due to faster decomposition, and/or presence of specific microbial communities, and if DOC production was affected by the natural soil environment.

*Funding Statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (*2018SFAF255).

Merging fungal and bacterial community profiles via an internal control

Miriam Hutchinson[1*] (miriamh@lanl.gov), Tisza Bell[2], Michaeline B.N. Albright[1], La Verne Gallegos[1], **John Dunbar**[1]

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico
[2]Division of Biological Sciences, University of Montana, Missoula, Montana

Project Goals

1) Develop a synthetic marker to serve as an "internal control" that can be used to scale bacterial and fungal profiles from next generation sequencing of environmental samples.
2) Implement this internal control to calculate the fungal to bacterial ratio, a metric which can reveal ecosystem functioning.

Abstract

In soil communities, the ratio of fungal to bacterial taxa in a community can predict responses to environmental change and potential impacts on ecosystem function, facilitating both carbon management as well as modeling. As such, these processes and the underlying microbial taxa are a relevant focus of the LANL Terrestrial Microbial Carbon Cycling SFA program. While members of a microbial community can be measured with shotgun metagenomics, the cost of this approach severely limits the number of samples that can be examined. Targeted metagenomic methods are less expensive, and therefore more practical for large-scale studies involving hundreds to thousands of samples; however, these approaches require separate measurements of domains (e.g. bacteria versus fungi) due to the lack of a universal genetic marker. In an effort to mitigate the inequities of targeted approaches, we present the addition of a synthetic internal control marker.Compatible with kingdom-specific PCR primers for both the bacterial small ribosomal subunit (SSU) V3-V4 region, the fungal large subunit (LSU) D2 region, and the fungal ITS (internal transcribed spacer) region, this marker can serve as a scaling factor to aggregate fungal and bacterial taxonomic profiles for the quantitative analysis of targeted approaches. We demonstrate the utility of this universal marker on soil communities

with known and unknown composition, using next-generation sequencing and quantitative PCR. For environmental samples of unknown composition, our internal marker predicts F:B ratios that are consistent with qPCR. Additionally, the marker outperforms qPCR in terms of producing F:B values that are closer to the actual values of a defined mock community. Furthermore, the internal marker allows for a more streamlined approach than qPCR because it is quantified in tandem with standard next-generation sequencing.

*Funding Statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (*2018SFAF255).

Microbial Community Composition Controls Carbon Flux Across Litter Types in Short-Term Litter Decomposition

Marie Kroeger[1*] (mkroeger@lanl.gov), Rae Devan[1], Jaron Thompson[2], Renee Johansen[3], La Verne Gallegos[1], Brian Munsky[2], Sanna Sevanto[4], Michaeline B.N. Albright[1], **John Dunbar**[1]

[1]Bioscience Division, Los Alamos National Lab, Los Alamos, New Mexico
[2]Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO
[3]Landcare Research, Auckland, New Zealand
[4]Earth and Environmental Science Division, Los Alamos National Lab, Los Alamos, New Mexico

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

Project Goals

1) Determine the relative importance of substrate (i.e. litter type) and functional state (DOC abundance) to explain the variance in microbial community composition.
2) Determine microbial traits that drive contrasting patterns of carbon flow in short-term litter decomposition among three litter types

Abstract

In terrestrial ecosystems, products of microbially-driven plant litter decomposition are major inputs to the soil organic carbon pool, a key carbon sink. Microbial composition plays an important role in determining the fate of plant litter carbon, which can either be respired as CO_2 or retained in the soil in other forms. However, the microbial community traits that alter carbon flow from litter decomposition are unknown. In this study we sought common microbial traits driving variation in carbon flow from three litter types (pine, oak, and a grass mix) during short-term decomposition in a common garden experiment. We measured carbon flow as carbon dioxide (CO_2) production throughout the 44-day experiment and a final dissolved organic carbon (DOC) measurement. Since we observed a greater than 3-fold microbially-driven DOC range between samples within each litter type and a significant negative correlation between DOC and CO_2 production, we did taxonomic profiling and selected a subset of high and low DOC samples for RNA-Seq analysis to determine differences in functional potential.

Hypotheses: 1) Plant litter type will select for different decomposer communities, but within a litter type, community composition will still vary with DOC abundance (high or low DOC). (2) Some microbial traits driving patterns of carbon flow will be common across the litter types.

Consistent with expectation, litter type explained more of the variance in community composition than DOC (12% and 7% respectively), but DOC constrained by litter type explained the most variance (31%). Bacteria rather than fungi appeared to drive carbon flow. Bacteria represented 82.1% and 82.5% of the differentially expressed microbial genera between low and high DOC, respectively. The limited fungal contribution to differentially expressed genes is consistent with a two-tiered decomposition process: 1) fungi dominate litter deconstruction, releasing DOC and 2) bacteria dominate consumption of DOC. The latter appears to account for

the variation in carbon flow observed in short-term decomposition. Common bacterial metabolic function gene expression signals linked to difference in carbon flow across litter types included biogenesis of cytochromes and selenoproteins, translation, protein secretion, and various carbon metabolism pathways. Observing significant differences at large functional levels that are common across litter types creates testable hypotheses for the mechanisms driving differences in carbon flow. Overall, these findings contribute to the goal of reprogramming carbon flow through microbiome engineering in terrestrial ecosystems to increase soil carbon storage.

*Funding Statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (*2018SFAF255).

Organism interactions and substrate range are the primary mechanisms linked to divergent carbon flow during litter decomposition

Michaeline B.N. Albright[*1] (malbright@lanl.gov), Jaron Thompson[2], Marie Kroeger[1], Renee Johansen[1], Danielle E.M. Ulrich[1], La Verne Gallegos-Graves[1], Brian Munsky[2], **John Dunbar**[1]

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico
[2]Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO, 80523

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

Project Goals

1) Generate hypotheses regarding mechanisms behind microbially driven differences in carbon flow from complex microbiome data
2) Use machine learning approaches to identify key microbial taxa linked to differences in carbon flow

Abstract

The LANL Microbial Carbon Cycling SFA aims to inform climate modeling and enable carbon management. The SFA uses genomics approaches to reveal microbial processes affecting biogeochemical cycling (e.g. C and N cycling) in terrestrial ecosystems. This involves discovery of fundamental principles driving the organization and interactions of soil microbes at multiple scales. Toward this end, we screened 206 soil communities decomposing plant litter in a common garden microcosm environment and examined features linked to divergent patterns of carbon flow. Carbon flow was measured as carbon dioxide (CO_2) and dissolved organic carbon (DOC) from 44-days of litter decomposition. Two large groups of microbial communities representing "high" and "low" DOC phenotypes from original soil and 44-day microcosm samples were down-selected for fungal and bacterial profiling. Metatranscriptomes were also sequenced from a smaller subset of communities in each group. Our results suggest that the high and low DOC phenotypes are a manifestation of both ontogenic effects (time dependent differences), and innate functional differences in the microbial communities. A later stage of successional development could yield low DOC, and thereby reflect an ontogenic effect. However, the difference in average respiration rate between the high and low DOC groups clearly indicates innate differences in physiology. To gain further insight, using machine learning approaches, we identified *features*—traits at the organism, pathway, or gene level—linked to the high and low DOC phenotypes. The features pointed to several potential mechanisms driving differences in carbon flow including organism interactions, substrate range and substrate affinity. Further confirmation and elucidation of innate functional constraints can inform strategies for soil carbon management.

*Funding Statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (*2018SFAF255)

Role of geographic scale in likelihood of microbial-driven functional variation during litter decomposition

M. Rae DeVan[1*] (raedevan@lanl.gov), Michaeline B.N. Albright[1], Jason Gans[1], La Verne Gallegos-Graves[1], Danielle Ulrich[2], Renee Johansen[1], Tom Yoshida[1], and **John Dunbar**[1]

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico
[2]Department of Ecology, Montana State University, Bozeman, Montana

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

Project Goal

1) Determine the magnitude of microbial driven variation in carbon flow (DOC & CO2), and the likelihood observing carbon flow variation over various geographic scales.
2) Establish how microbial communities vary in composition and carbon flow with distance.

Abstract

During litter decomposition, microbial communities release carbon as CO_2 and dissolved organic carbon (DOC). Some microbial communities produce greater DOC, while others produce more CO_2. In theory, this variation could be leveraged to increase soil carbon sequestration. A key starting point is to find natural communities that represent differences in carbon flow. Here, we examined the relationship of microbial-driven functional variation and geographic distance. Is a large geographic scale required to find substantial functional variation? We tested whether stochastic assembly, which occurs at local scales, could generate microbial communities with substantial functional variation or if larger scale, deterministic processes (climate and ecosystem gradients), were required to recover maximal functional variation. We tested this hypothesis with over 300 soil samples collected at a local (<10 m to 200 m) and regional scale (1 km to >300 km). Soil microbial communities were transferred to lab microcosms with sterile blue grama (*Bouteloua gracilis*) litter. Carbon flow (DOC and CO_2) was measured over 45 days of decomposition. The range of variation in DOC or CO_2 increased with physical distance across both collection scales However, only DOC variance was increased by including the regional scale. Variance in CO_2 reached similar levels at both the local and regional scale. Fungal and bacterial community composition were highly correlated with geographical distance at the regional, but not local scale. These findings suggest that the full range of microbially driven variation in CO_2 can be documented among samples collected at a local scale (<200m). Whereas DOC variation depends on larger differences in community composition, which tend to occur only at larger geographic scales, thus to recover the full range samples should be collected up to 150 km apart. These data guide sample collection for future projects that seek soil microbial communities driving variation in carbon cycling.

*Funding Statement: This work was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (*2018SFAF255).

High-speed spectroscopic stimulated Raman scattering microscopy for measuring biofuel synthesis

Haonan Lin[1,3]*(hnlin@bu.edu), Nathan Tague[1], Jean-Baptiste Lugagne[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

http://www.dunloplab.com/ http://sites.bu.edu/cheng-group/

Project Goal: Spectroscopic stimulated Raman scattering (SRS) produces label-free chemical maps of molecules in living systems. However, the limited speed of the conventional SRS platform makes it challenging to study highly dynamic or large-scale samples. To resolve biofuels at the single-cell level in real time, we are developing a high-speed SRS system by incorporating an ultrafast polygon scanner as a delay tuner. Using the proposed method, we achieved an acquisition speed of up to 20 μs per SRS spectrum. Direct measurement of fatty acid production levels in hundreds of individual *E. coli* cells within 4 seconds is demonstrated. Such high-throughput single-cell measurements provide insights about production levels and phenotypic variation among cells. Understanding these properties can be exploited to improve the design of biofuel production strains.

Abstract

Stimulated Raman scattering (SRS) is an emerging imaging modality that produces label-free maps of chemical vibration bonds in living biological environments. Compared to spontaneous Raman, the SRS signal is generated by overlapping two lasers, termed pump and Stokes, to generate Raman signals with up to 10^6 orders of signal enhancement. By exciting a single Raman band, SRS can reach video-rate. However, most biomolecules have overlapping Raman signatures, which reduces its chemical specificity. To overcome the limitation, spectroscopic SRS has been developed, which can generate a series of SRS images at consecutive Raman excitation bands. Yet, the imaging speed of current spectroscopic SRS is limited to minutes per image stack, which is insufficient for rapid imaging of biofuel production for a large *E. coli* production library at the single-cell level.

We use spectral focusing to generate SRS spectrum at high speeds. The scheme of spectral focusing is shown in **Fig. 1a**. Two femtosecond pulses are chirped by glass rods such that different wavelength components are temporally separated, by changing the time delay of the chirped pulses, the beating frequency is changed to excite different Raman band. The setup for the SRS system is depicted in **Fig. 1b**. Two 80 MHz femtosecond pulses generated by the same laser source (Insight DeepSee+, Spectra Physics, CA) serve as pump and Stoke lasers. The Stokes beam is directed to a polygon scanner and is subsequently scanned to a blazed grating. The grating is set to Littrow configuration such that the first-order diffraction beam is retroflected to the same optical path. The scanning of the polygon mirror results in a path difference of a few millimeters for the

retroflected beam. The delay range of the Stokes beam is freely adjustable by rotating the blazed grating such that the system can cover ~200 cm^{-1} of the Raman spectrum range. The introduced delay is used to tune the beating frequency of the two beams after chirping both beams by high-dispersive glass rods (SF-57).

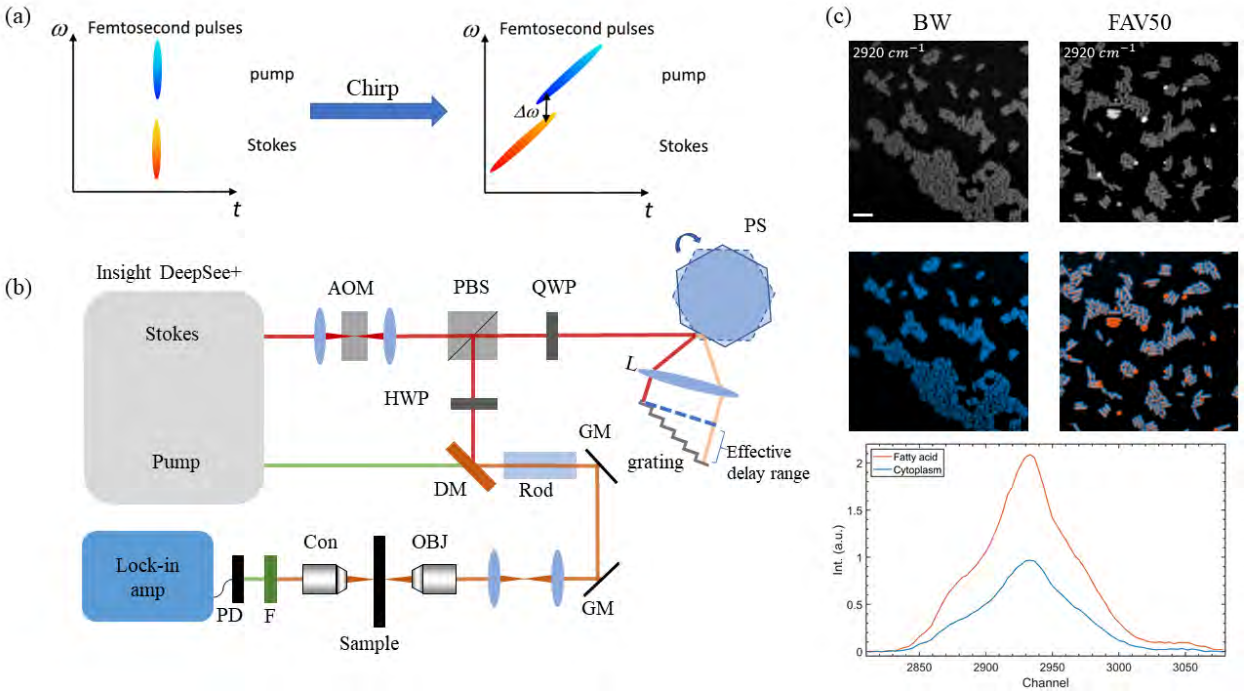


Figure 1. Experimental setup and application to imaging *E. coli* fatty acid production. (a) Concept of spectral focusing. Two femtosecond pulses are linearly chirped to separate different wavelength components in the time domain, by changing the time delay between the two pulses, the beating frequency for SRS excitation is changed. (b) Optical setup. AOM, acousto-optic modulator; C, condenser; F, filter; HWP, half-wave plate; L, lens; M, mirror; OBJ, objective; PBS, polarizing beam splitter; PD, photodiode; PS, polygon scanner; QWP, quarter-wave plate. (c) Experimental results for imaging wild-type (BW) and fatty acid producing (FAV50) strains. Single color SRS image at 2920 cm^{-1} and chemical maps for fatty acid and cytoplasm are shown, indicating a significant increase in fatty acid in the FAV50 strain. Scale bar, 10 μm.

Using the platform, we performed spectroscopic SRS imaging of *E. coli* fatty acid production strains. **Fig. 1c** shows the experimental results for wild type (BW) and fatty acid producing (FAV50) strains, where each spectroscopic image stack was acquired within 1.8 seconds. By performing linear unmixing of the spectroscopic image, we can separate fatty acid from the cytoplasm (protein) and generate chemical maps for two components. Comparing the two strains, we can observe from FAV50 an overall SRS intensity increase. In addition, the localized aggregation of fatty acid formed lipid droplets within the cell body. Cell-to-cell heterogeneity in fatty acid production levels in the FAV50 strain were also observed.

This work is supported by DOE grant BER DE-SC0019387 to MJD, WWW, and JXC.

Single Cell Chemical Imaging with Stimulated Raman Scattering for Biofuel Production Screening

Nathan Tague[1]* (ntague@bu.edu), Jean-Baptiste Lugagne,[1] Haonan Lin,[1] Jonghyeon Shin[1], Jing Zhang[1], Wilson Wong[1], Ji-Xin Cheng[1], and **Mary J. Dunlop**[1]

[1] Boston University, Boston, MA

http://www.dunloplab.com

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 engineering techniques, such a CRISPR libraries and efficient cloning methods, allow for more extensive, combinatorial strain optimization. However, accurate, high-throughput screening techniques have not kept pace with the ability to generate large strain libraries. Stimulated Raman scattering (SRS) imaging has the potential to alleviate the screening bottleneck by directly detecting chemicals' Raman spectra within single cells. Here, we demonstrate the ability to differentiate production levels of several biofuels within *E. coli* based on unique Raman spectra footprints. The results show the potential of this imaging platform to increase throughput in production strain optimization screening while maintaining single cell level information.

This work is supported by DOE grant BER DE-SC0019387.

Genomics and Phenomics to Identify Yield and Drought Tolerance Alleles for Improvement of Camelina as a Biofuel Crop

Hussein Abdel-Haleem[1*] (Hussein Abdel-Haleem@usda.gov**)**, Zinan (Lily) Luo[1], Aaron E. Szczepanek[1], Matthew Conley[1], Hao Zhang[2], Yufeng Ge[2], Daniel Schachtman[2], Megan M. Augustin[3], Jordan R. Brock[3], Noah Fahlgren[3], Toni Kutchan[3], Sheeja George[4], Russ Gesch[5], and **John M. Dyer[1]**

[1]USDA-ARS, US Arid-Land Agricultural Research Center, Maricopa, AZ; [2]University of Nebraska-Lincoln, Lincoln, NE; [3]Donald Danforth Plant Science Center, St. Louis, MO; [4]University of Florida, Quincy, FL; and [5]USDA-ARS, North Central Soil Conservation Research Laboratory, Morris, MN

Project Goals: Plant oils represent renewable sources of energy-dense hydrocarbons that can be used for biofuel, but a major challenge is to produce these oils in high-yielding non-food crops that can grow under marginal conditions. Our goal is to improve the suitability of camelina as a bioenergy crop. The major objectives are to: 1) Develop and apply automated, non-destructive high-throughput phenotyping (HTP) protocols to evaluate the phenotypic diversity of a camelina panel consisting of 250 accessions, grown under well-watered and water-limited conditions; 2) Discover alleles/genes controlling morphological, physiological, seed, and oil yield traits using genome-wide association studies; and 3) Identify, test, and validate useful germplasm under diverse environments and marginal production areas. These studies will enable the discovery of new genes associated with crop yield and stress tolerance and identify high-yielding cultivars that are suitable for certain geographical regions.

In recent years, *Camelina sativa* has received considerable attention as a potential non-food biofuel crop, but significant challenges remain to develop stable, high-yielding, geographically adapted germplasm suitable for biofuels production. Here, we are using advanced high-throughput phenotyping and genomics-based approaches to discover useful gene/alleles controlling seed yield and oil content and quality in camelina under water-limited conditions and identify high-yielding cultivars suitable for production in different geographical regions. Our experimental design includes complimentary field- and greenhouse-based HTP experiments conducted on a spring panel of 250 camelina accessions grown under well-watered and water-limited conditions.

To develop genomic tools and information for underpinning genome-wide association studies (GWAS), the camelina panel was genotyped using genome-by-sequencing technology (GBS), resulting in identification of 6,192 high-quality SNPs distributed throughout the genome. Population genetics analysis revealed two distinct populations of camelina corresponding to main geographical regions of cultivar collection. As an example of the utility of using these SNP markers to find genes associated with agronomically important traits, we conducted a GWAS of cuticular wax traits in the camelina diversity panel. Cuticular wax content and composition are important for regulating non-stomatal water loss in aerial organs of plants, and identification of markers and genes associated with these traits might be useful for developing more drought-tolerant camelina varieties. Wax content and composition were first determined for select varieties of camelina and closely related species, resulting in identification of 49 distinct lipid molecular species. These compounds were then characterized for the entire camelina diversity

panel, revealing variations in wax content and composition. Broad-sense heritabilities further revealed a strong genetic component for certain lipid classes and lipid molecular species. GWAS studies were subsequently conducted, resulting in identification of SNP markers associated with certain wax components. Several of these markers were associated with specific genes, revealing candidates (and markers) that might be used for rational manipulation of wax content using molecular breeding approaches.

To identify genes regulating camelina oil content, composition, and yield, the entire camelina panel was cultivated and analyzed using a greenhouse-based high throughput phenotyping system (LemnaTech) available at the Danforth Center, with two replicates of each accession cultivated under well-watered and water-limited conditions, from 7 to 35 days after planting. Phenotypic data for each plant were collected and included approximately 200,000 images produced during the course of the experiment. Images were processed using PlantCV software to extract trait data including plant height and plant area. Algorithms are currently being developed to detect plant flowering. Seeds were also harvested from each plant and analyzed for oil content, composition, nitrogen and glucosinolate content, as well as total seed yield. Data analysis showed significant phenotypic variations among genotypes as well stress levels. Data are being prepared for subsequent GWAS analyses.

The camelina panel was also cultivated under well-watered and water-limited conditions in fields located in Nebraska and Arizona. Field high-throughput phenotypic data including plant height, canopy temperature, and spectral reflectances (e.g., NDVI) were collected using ground-based phenotyping vehicles equipped with various sensors and data logging equipment. More than 20 phenotypic traits related to yield components and fatty acids were collected. These data from well-watered and water-limited trails were analyzed and indicated significant phenotypic diversity in the camelina population. Genome-wide association analysis using GAPIT was conducted to identify putative single nucleotide polymorphism (SNP) markers that are significantly associated with these traits. Together, seven and nine significant SNP markers on different chromosomes were associated with plant height, seed weight and seed yield under water-limited and well-watered conditions, respectively, meaning that some genes might be actively triggered under either of the two conditions. No significant markers overlapped between environments. BLASTx annotation identified gene functions for 8 out of 16 genes associated with significant SNP markers. These functions were mainly related to general maintenance of normal cell function and development. These SNP could be good candidates for marker assisted selection. Data from more environments, years and locations will be combined for a more comprehensive GWAS analysis. This GWAS study will lay a foundation for future molecular breeding programs to help accelerate the selection of *C. sativa* varieties with superior traits of interests. Field data will be compared to data obtained from the greenhouse-based studies to help identify select cultivars that consistently perform better under water-limited conditions, and also provide a means to determine how well greenhouse-based studies mimic conditions observed in the field.

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2016-67009-25639.

Genome remodeling to control the persistence of engineered functions in soil microbes

Joshua Elmore,[1] Ryan McClure,[1] Ritu Shrestha,[1] Ryan Francis,[1] Henri Baldino,[1] Bill Nelson,[1] Ernesto Nakayasu,[1] Yuliya Farris,[1] Shara Balakrishnan,[2] Valentine Trotter,[3] Janet Jansson,[1] Enoch Yeung,[2] Adam Deutschbauer,[3] and **Robert Egbert**[1]* (Robert.Egbert@pnnl.gov)

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA; [2]University of California Santa Barbara; [3]Lawrence Berkeley National Laboratory, Berkeley, CA

Project Goals: Our overarching vision is to develop fundamental understanding of what controls the persistence of engineered microbial functions in complex environments. Complex traits are often governed by multiple genes, but no methods exist to predict how remodeling a genome will impact cellular function or govern the conditions under which a microbe will persist. We are developing a data-driven genome reduction strategy to control persistence across environmental variations such as temperature or pH, focusing on two closely related bacteria: *Pseudomonas putida*, which is of high value to industrial biotechnology; and *Pseudomonas fluorescens*, a plant growth promoting bacterium.

Prediction and control of the persistence of recombinant microorganisms in the environment are key to threat assessment across the biodefense enterprise. Effective biocontainment strategies based on this fundamental understanding will enable DOE Biosystems Design research to advance from the laboratory to the field. Key knowledge gaps in controlling microbial persistence in soil are how genes function in those conditions and how gene networks contribute to complex phenotypes[1]. Our team's recent research is beginning to address these gaps through high-throughput genetics[2], synthetic biology chassis development[3,4], multi-omics characterization of soil communities[5], and machine learning applied to biological systems design. We have also recently elucidated the impacts of genome reduction on the persistence of violacein bioproduction in *E. coli* and are poised to extend this research to new microbes and new environments. This work

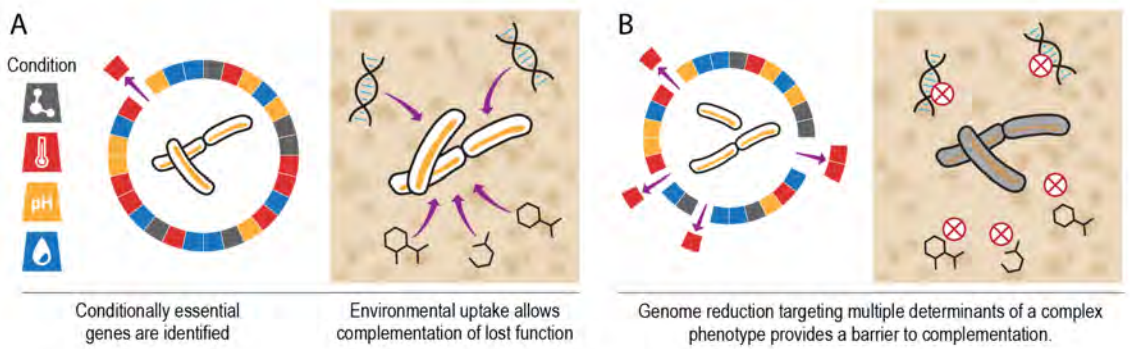


Figure 1. Developing fundamental understanding of how disparate genes contribute to complex phenotypes will enable effective control of persistence. (A) Existing biocontainment strategies rely on loss of function phenotypes that can bypassed by environmental uptake of complementing DNA or metabolites. (B) A fundamental understanding of the genetic determinants of persistence will enable effective, condition-specific biocontainment through genome reduction. Removing multiple genes of environmental importance are less likely to be overcome by complementation.

leverages expertise from the PNNL Soil Microbiome Scientific Focus Area (SFA), which is funded to model and predict microbial dynamics in grassland soils.

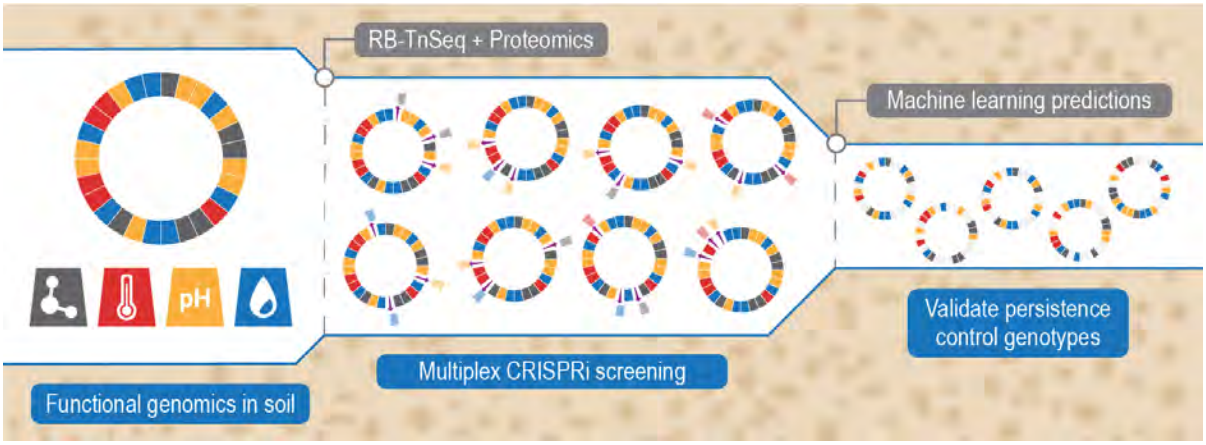


Figure 2. This work applies expertise in functional genomics, global proteomics, synthetic biology, and machine learning to enable validation of genome reduction as an effective strategy to control persistence. Gene-level conditional essentiality scores based on random-barcode transposon sequencing drive the design of pooled combinatorial multi-gene CRISPR-Cas repression assays. These assays and data-driven machine learning inferences will guide the engineering of conditional persistence phenotypes in soil environments.

This project integrates tools from functional genomics, global proteomics, and synthetic biology with machine learning to identify and validate the genetic determinants of microbial persistence in soil. The primary outcomes of this project will be (1) to elucidate gene function across environmental perturbations in soil via high-throughput genetics and proteomics, (2) to develop high-throughput genome editing tools for a plant growth-promoting bacterium, and (3) to create machine learning approaches that synthesize genome remodeling designs from multi-omics inputs. Ultimately, these efforts will advance science critical to engineering enhanced stability, resilience, and controlled performance in DOE-relevant biological systems.

This work is funded as a pilot project through the Secure Biosystems Design program within the Office of Biological and Environmental Research of the Department of Energy (proposal #0000246830).

References

1. Venturelli, O. S., Egbert, R. G., & Arkin, A. P. (2016). Towards engineering biological systems in a broader context. *Journal of molecular biology*, *428*(5), 928-944.
2. Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, M., Ray, J., Liu, H., ... & Carlson, H. K. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*, 557(7706), 503.
3. Egbert, R. G., Rishi, H. S., Adler, B. A., McCormick, D. M., Toro, E., Gill, R. T., & Arkin, A. P. (2019). A versatile platform strain for high-fidelity multiplex genome editing. *Nucleic acids research*, *47*(6), 3244-3256.
4. Elmore, J. R., Furches, A., Wolff, G. N., Gorday, K., & Guss, A. M. (2017). Development of a high efficiency integration system and promoter library for rapid modification of Pseudomonas putida KT2440. *Metabolic engineering communications*, *5*, 1-8.
5. Hultman, J., Waldrop, M. P., Mackelprang, R., David, M. M., McFarland, J., Blazewicz, S. J., Harden, M.R., Turetsky, A.D., McGuire, M.B., Shah, N.C., VerBerkmoes, N. C., Ho Lee, L., Mavrommatis, K., Jansson J.K. (2015). Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature*, *521*(7551), 208.

Transgenic Poplar Lines to Probe Host Genes Involved in Defense Against Rust

Liang Hu[1], Gen Li[1], James F. Parsons[1], Gary Coleman[1,2], Shunyuan Xiao[1,2] and **Edward Eisenstein**[1,3,*] (eisenstein@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: *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. A screen has been developed in tobacco leaves and poplar protoplasts for CSEPs that affect host immunity. Two subsets of effectors have been targeted for analysis: one with significant homology to CSEPs in other fungal pathogens of plants, and another, large family, containing unique members with no similarity to any proteins in sequence databases. Although results from these screens show that many CSEPs suppress the host immune response, unexpectedly, several promote an immune reaction, triggering a hypersensitive response. Genes for these immune-reactive effectors have been used to construct stable transgenic poplar lines with constitutive and inducible promoters to identify host genes involved in defense against rust, for the evaluation of susceptibility mechanisms in compromised hosts, and to provide reagents to the community.

Abstract: Fungal pathogens of plants, such as *Melampsora larici-populina* which promotes rust disease in *Populus* spp., contain genes for small, cysteine-rich secreted proteins that are specifically up-regulated for expression during infection, signifying a key role in host colonization by the pathogen. Computational analysis of the *Melampsora larici-populina* genome suggests that its complement of 1,524 candidate secreted effector proteins (CSEPs) belong to 807 structural families, including 600 single-member families, another 206 families that contain 2-36 members, and one family that has undergone a remarkable expansion to 117 members. 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. The two subsets of CSEPs were expressed transiently in tobacco along with well-known 'autoactive' domains of *R* genes or an *R* gene-*AvrP* pair that promotes a hypersensitive response (HR) or HR-like cell death. 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. Although roughly half of the effectors screened affect host immunity, interestingly, a smaller subset stimulates an immune response in the screen, and when examined alone, promote a hypersensitive response in tobacco. Transient expression of the reactive CSEPs in poplar protoplasts is challenging, and low throughput, but initial

results support the tobacco-based large-scale screens. A high-efficiency transformation protocol has been developed and used to construct stable transgenic poplar lines that express *Melampsora* CSEPs from strong, constitutive or inducible promoters. These lines are being used to identify host genes involved in defense against rust, for the evaluation of susceptibility mechanisms in compromised hosts, and to provide reagents to the community.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0017886

Free-living Nitrogen Fixation in the Switchgrass Rhizosphere

Darian Smercina [1*] (marinisd@msu.edu), Lisa Tiemann [1], Alan Bowsher [3], Maren Friesen [4,5], James R. Cole [1], Elizabeth Eder [6], David Hoyt[6], William Chrisler [6], John Cliff [6], Kirsten Hofmockel [7], **Sarah Evans [2]**

[1] Michigan State University, Dept. of Plant, Soil and Microbial Sciences
[2] Michigan State University, Dept. of Integrative Biology
[3] Michigan State University, Dept. of Microbiology and Molecular Genetics
[4] Washington State University, Dept. of Plant Pathology
[5] Washington State University, Dept. of Crop and Soil Sciences
[6] Environmental Molecular Sciences Lab, PNNL
[7] Earth and Biological Sciences Directorate, PNNL

http://rhizosphere.msu.edu/

This work aims to improve our understanding of an important nitrogen source, free-living nitrogen fixation, (FLNF) and its potential to support the nitrogen demands of bioenergy cropping systems like switchgrass (*Panicum virgatum*). Our work systematically characterizes the association between switchgrass and free-living N-fixing organisms (diazotrophs) including evaluation of the conditions which promote FLNF, characterization the rhizosphere diazotroph community, and testing the "carbon-for-nitrogen exchange" hypothesis.

FLNF is an important terrestrial N source that occurs under diverse environmental conditions, which are distinct from symbiotic N-fixation.[1] FLNF likely occurs predominately in the rhizosphere, where bioavailable carbon (C) is readily accessible. Switchgrass, an important bioenergy crop, harbors a diverse rhizosphere community of diazotrophs which it may rely on as a significant N source when grown on marginal lands.[2,3] It is becoming increasingly clear that diazotrophs are present and actively fixing N in the switchgrass rhizosphere, yet it is not known if or how fixed N is exchanged between diazotrophs and switchgrass and if this is coupled with root C exudation.

In order to better characterize the switchgrass-diazotroph association and its potential to contribute N to switchgrass cropping systems, we first optimized the methods to mimic rhizosphere conditions.[4] Using these optimized methods, we measured potential FLNF rates across two growing seasons of switchgrass growth. We evaluated the impact of legacy and short-term fertilizer N additions on FLNF rates and the switchgrass-associated diazotroph community (*nifH* gene sequencing) through a greenhouse study of switchgrass grown in Michigan marginal land soils. 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] Building from these previous studies, we are now using a novel, sterile growth system to assess interactions between switchgrass and known diazotrophs and test the "C for N exchange" hypothesis by pairing multiple techniques including fluorescent

in situ hybridization (FISH), NanoSIMS via ^{13}C and ^{15}N labelling, and assessment of the rhizosphere metabolome via NMR and LC-MS.

Our findings indicate that FLNF in the rhizosphere is promoted by low, but not anaerobic oxygen conditions and availability of diverse C sources. We consistently detected FLNF in the switchgrass rhizosphere across two growing seasons. We also found that switchgrass cultivates a consistent diazotroph community regardless of legacy or short-term N fertilization or initial diazotroph community composition. Despite a lack of response to N by the diazotroph community, rhizosphere metabolite chemistry is driven more strongly by N availability than by diazotroph presence. To examine switchgrass-diazotroph interactions further, we grew switchgrass, inoculated with known diazotrophs, in a sterile system. We used NanoSIMS images of switchgrass roots labeled via introduction of 99 atom% ^{13}C-CO_2 to show direct uptake of subsequent ^{13}C labeled root exudate by diazotrophs. Addition of 98 atom% $^{15}N_2$ will allow us to simultaneously visualize if FLNF of diazotrophs is coupled to C uptake. Development of diazotroph species-specific FISH probes allowed us to visualize establishment of diazotroph populations on switchgrass roots, confirming potential for the "C for N exchange" hypothesis.

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. (2019). Optimization of the 15 N 2 incorporation and acetylene reduction methods for free-living nitrogen fixation. Plant and Soil, 445(1), 595-611.
5. Smercina, D. N., Bowsher, A. W., Evans, S. E., Friesen, M. L., Eder, E. K., Hoyt, D. W., & Tiemann, L. K. (2019). Switchgrass Rhizosphere Metabolite Chemistry Driven by Nitrogen Availability. bioRxiv, 802926.

This work was conducted under the MMPRNT project, funded by the DOE BER Office of Science award DE-SC0014108. Support for this research was provided in part by the Great Lakes Bioenergy Research Center, DOE BER Office of Science Awards DE-SC0018409 and DE-FC02-07ER64494. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) program and used resources at DOE BER Office of Science user facilities, the Joint Genome Institute (DE-AC02-05CH11231) and the Environmental Molecular Sciences Laboratory (DE-AC05-76RL01830; grid.436923.9). In particular, metabolomics profiling, FISH, and NanoSIMS imaging were carried out using EMSL instrumentation (Proposal IDs: 49977 and 50682) and was supported by the DOE, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE - SC0014664.

Relic DNA dynamics mask the resilience of switchgrass bacterial communities to extreme drying rewetting

Heather Kittredge[1,2,3]* (kittred8@msu.edu), Kevin Dougherty, Kate Glanville[2,4], Maren Friesen[5], Lisa Tiemann[1,4], **Sarah Evans**[1,2,3,6]

[1]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI;[2]W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI; [3]Department of Integrative Biology, Michigan State University, East Lansing, MI; [4]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI; [5]Crop and Soil Sciences, Washington State University, Pullman, WA; [6]Department of Microbiology & Molecular Genetics, Michigan State University, East Lansing, MI;

Project Goals: We characterized the active and 'relic' (inactive or dead cells) microbiome of switchgrass and corn soils over a short-term disturbance. We measured the abundance and composition of living and relic bacterial communities before and after an extreme rewetting event (80mm of rain after 28 days of drought). Measuring both changes in viable taxa and relic DNA provides valuable information about the bacterial taxa most resilient to drying rewetting and the sensitivity of different soil bacterial communities to drying rewetting.

The soil microbiome is likely to be an important contributor to the sustainability of cellulosic bioenergy feedstocks. Future climates are expected to have more variable rainfall, including longer droughts and intensified rainfall events. Plant-microbe interactions can increase the resilience of cellulosic energy crops to extreme drying rewetting events, but bacterial responses to short-term disturbances like drying rewetting can be context-dependent and difficult to characterize. Specifically, DNA-based characterizations include dead or dormant cells, obscuring responses of the live community. Here, we asked 1) how do viable bacteria respond to drying rewetting and does this differ between corn and switchgrass soil? 2) Does the inclusion of relic DNA mask responses and 3) how does the composition of viable communities relate to relic DNA pools across a disturbance? To address these questions, we quantified changes in the size and composition of the viable bacterial community and relic DNA pools using the chemical treatment propidium monoazide (PMA) on soil samples collected at three time points; 6 hours before, 1 hour after, and 18 hours after an extreme drying rewetting event (80mm of rain every 28 days). We found that viable bacterial taxa and relic DNA in corn and switchgrass soil responded differently to extreme drying rewetting. In corn soil, we observed an increase in relic DNA 1-hr after extreme rewetting, and a simultaneous decrease in the relative abundance of viable bacteria. In switchgrass soil, we observed the opposite trend, an increase in the relative abundance of living bacteria, associated with a rapid decrease in relic DNA after rewetting. When we completed the analyses without controlling for relic DNA, we found that the increase in relic DNA in corn soil masked the sensitivity of corn bacterial communities to extreme rewetting, while the decrease in relic DNA in switchgrass soil masked the resiliency of switchgrass bacterial communities to extreme rewetting. To improve our understanding of the sources of relic DNA, we also looked at the number of taxa that appeared in both the viable community and the relic DNA community and found that 72% of taxa were present in both the living and relic communities. Furthermore, among the top 25% most abundant taxa, 100% were present in both the living and relic communities. However, only 55% of living taxa were present before and after extreme drying rewetting. Together, our results indicate

that relic DNA dynamics differ by soil cropping system and can mask important responses to disturbance that are critical for the sustainable production of cellulosic bioenergy crops in future climates.

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 and DE-FC02-07ER64494 in addition to Award Number DOE DE-FOA-0001207 to the MMPRNT project.

Title Soil Microbes Affect Switchgrass Germination More than Seedling Growth Under Drought

Tayler C. Ulbrich[1,2*] (ulbrichtayler@gmail.com), Lukas Bell-Dereske[1,2], Harry Ervin[2], Shanna Hilborn[3], **Sarah E. Evans**[1,2]

[1]Department of Integrative Biology, Michigan State University, East Lansing, MI

[2]W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI

[3] Lyman Briggs College, Michigan State University, East Lansing, MI

Project Goals: Switchgrass, a bioenergy candidate, can grow in marginal, drought-prone soils, but is limited by drought-sensitive early life stages. We investigated how soil microbial communities affect switchgrass germination and seedling growth during drought. These findings can be used to inform how switchgrass establishment could be improved by selecting or manipulating soil microbial communities that improve switchgrass germination and growth under drought.

A limitation to the bioenergy potential of switchgrass (*Panicum virgatum*), is its drought-sensitive early life stages (e.g. germination and seedling establishment). Single microbial isolates have been shown to improve switchgrass germination and growth under drought,[1,2] but less is known about how whole soil communities influence switchgrass drought tolerance. Additionally, little is known about how the degree of host-association in the soil communities affects switchgrass drought responses. For instance, soils with a high degree of microbial-host association, collected from mature switchgrass fields, have negative effects on early switchgrass growth[3], but no studies to our knowledge have investigated if host-specific, mature switchgrass soil microbes affect germination and, further, if these interactions shift under drought.

We used a greenhouse experiment to investigate how switchgrass germination and seedling growth under drought is affected by microbial presence (sterile bulk vs. live bulk soils) and the degree of microbial-host association (bulk vs. mature switchgrass rhizosphere soils). We found that drought negatively affected switchgrass germination and growth, and that microbially-mediated drought tolerance is stronger during germination than seedling growth. Seeds sown into live soils had 40% higher germination and 63% higher survival than sterile soils under drought. The degree of host-microbial association also had greater effects on germination than seedling growth, but the effects differed under drought and ambient conditions. Under ambient conditions, more seeds germinated in host-specific, rhizosphere soils, but this was reversed under drought. The initially similar bacterial community composition of the bulk and rhizosphere soils also changed during germination but did not diverge during seedling-growth, suggesting that germination-specific signaling may recruit microbes which may aid in germination and survival under drought. In summary, this study shows that microbially-mediated drought tolerance differs for switchgrass germination and seedling life-stages, where germination is more sensitive to drought and more responsive to differences in microbial communities. Future efforts to improve switchgrass establishment during drought should focus on germination-specific microbial interactions.

References

1) Ghimire, Sita R., and Kelly D. Craven. “Enhancement of Switchgrass (Panicum Virgatum L.) Biomass Production under Drought Conditions by the Ectomycorrhizal Fungus Sebacina Vermifera.” *Applied and Environmental Microbiology* 77, no. 19 (2011): 7063–67. https://doi.org/10.1128/AEM.05225-11.
2) Wang, Bingxue, John R. Seiler, and Chuansheng Mei. “A Microbial Endophyte Enhanced Growth of Switchgrass under Two Drought Cycles Improving Leaf Level Physiology and Leaf Development.” *Environmental and Experimental Botany* 122 (2016): 100–108. https://doi.org/10.1016/j.envexpbot.2015.09.004.
3) Hawkes, Christine V., Stephanie N. Kivlin, Jennifer Du, and Valerie T. Eviner. “The Temporal Development and Additivity of Plant-Soil Feedback in Perennial Grasses.” *Plant and Soil* 369, no. 1–2 (2013): 141–50. https://doi.org/10.1007/s11104-012-1557-0.

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 and DE-FC02-07ER64494 in addition to Award Number DOE DE-FOA-0001207 to the MMPRNT project.

Elucidating the Molecular Mechanisms Underlying Drought Resilience in Sorghum

Maxwell Braud,[1] Philip Ozersky,[1] Yuguo Xiao,[1] Brian Dilkes,[2] Duke Pauli,[3] 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; and [3]University of Arizona, Tucson, AZ

Project Goals:

- **Overall project objective: To define and functionally characterize genes and pathways related to drought stress tolerance in sorghum and the molecular mechanisms by which these factors drive phenotypic diversity.**
- **Establish a foundation for deep explorations of gene regulatory networks in sorghum through integrative genomics analyses.**
- **Enhance understanding of how genotype drives phenotype and environmental adaptation using high-resolution, field-based phenotyping of sorghum mutant collections.**
- **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. A sorghum mutant population will be 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 other DOE-funded initiatives and will be extended here 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.

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.

Cultivation-independent expansion of the Nucleocytoplasmic Large DNA Viruses

Frederik Schulz[1*] (fschulz@lbl.gov), Simon Roux[1], David Paez-Espino[1], Sean Jungbluth[1], David Walsh[2], Vincent J. Denef[3], Katherine D. McMahon[4], Konstantinos T. Konstantinidis[5], Emiley A. Eloe-Fadrosh[1], Nikos Kyrpides[1], **Tanja Woyke[1]**

[1] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; [2] Concordia University, Montréal, Québec, Canada; [3] University of Michigan, Ann Arbor; [4] University of Wisconsin-Madison, Madison; [5] Georgia Institute of Technology, Atlanta

Project Goals: Employ genome-resolved metagenomics on public environmental sequencing data to assess the diversity and coding potential of giant viruses and predict interactions of these viruses with their eukaryotic hosts.

The discovery of large and giant Nucleocytoplasmic Large DNA Viruses (NCLDV) with genomes in the megabase-range and equipped with a wide variety of features typically associated with cellular organisms was one of the most unexpected, intriguing and spectacular breakthroughs in virology. Current knowledge about the NCLDV is largely derived from viral isolates co-cultivated with protists and algae. Building on the rapidly increasing wealth of publicly available metagenome data, we reconstructed 2,074 NCLDV genomes from 8,535 metagenomes from sampling sites spanning the globe. This led to an 11-fold increase in phylogenetic diversity and a parallel 10-fold expansion in functional diversity. Analyzing 58,023 metagenomic major capsid proteins of large and giant viruses revealed global distribution patterns and underlined their cosmopolitan nature. The discovered viral genomes encoded a wide range of proteins with putative roles in photosynthesis and diverse substrate transport processes, revealing host reprogramming as likely common strategy in the NCLDV. Furthermore, horizontal gene transfer inferences connected viral lineages to diverse eukaryotic hosts. We anticipate that the vast diversity of NCLDV revealed here on a global scale will establish giant viruses as key ecosystem players across Earth's biomes, associated with most major eukaryotic lineages.

References

1. Schulz et al., (2019). Giant virus diversity and host interactions through global metagenomics. *Nature* https://www.nature.com/articles/s41586-020-1957-x

This work was conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, under Contract No. DE-AC02–05CH11231 and made use of resources of the National Energy Research Scientific Computing Center, also supported by the DOE Office of Science under Contract No. DE-AC02–05CH11231

The National Microbiome Data Collaborative: Empowering the Research Community to More Effectively Harness Microbiome Data

Pajau Vangay[1]* (pvangay@lbl.gov), Lee Ann McCue[2], Chris Mungall[1], Stanton Martin[3], Shane Canon[1], Patrick Chain[4], Kjiersten Fagnan[1], Elisha Wood-Charlson[1], Nigel Mouncey[1], and **Emiley Eloe-Fadrosh**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Pacific Northwest National Laboratory, Richland, WA; [3]Oak Ridge National Laboratory, Oak Ridge, TN; and [4]Los Alamos National Laboratory, Los Alamos, NM

https://microbiomedata.org/

Project Goals: The vision of the National Microbiome Data Collaborative (NMDC) is to empower the research community to harness microbiome data exploration and discovery through a collaborative and integrative data science ecosystem. The NMDC will address fundamental roadblocks in microbiome data science through implementation of guiding principles to make data findable, accessible, interoperable, and reusable (FAIR). To realize this vision, our multi-Lab collaborative partnership will pilot an integrated, community-centric framework to fully leverage existing microbiome data science resources and high-performance computing systems available within the DOE complex for data access, integration, and advanced analyses.

Abstract

The cross-cutting nature of microbiome research in environmental sciences, health, agriculture, energy, and natural and built environments requires the development of new solutions and community coordination to tackle grand challenges that will accelerate basic discovery and lead to transformative advances. The velocity at which microbiome data are generated has far outpaced current infrastructure resources for collection, processing, and distribution of these data in an effective, uniform, and reproducible manner, even at the largest data centers. The Interagency Strategic Plan for Microbiome Research outlined three areas of focus for strategic investments over the next five years, importantly highlighting the development of platform technologies and specifically support for open and transparent data through the development of a user-friendly, robust, integrated system with expert curation. The NMDC will tackle these infrastructure challenges in microbiome data science through developing a community-centric framework based on large-scale, collaborative partnerships leveraging unique capabilities, expertise, and resources available across four DOE National Laboratories (Lawrence Berkeley National Laboratory (LBNL), Los Alamos National Laboratory (LANL), Pacific Northwest National Laboratory (PNNL) and Oak Ridge National Laboratory (ORNL)).

During Phase I (the first 27 months), we aim to deliver a set of unique microbiome data science capabilities through leveraging existing microbiome resources hosted across the DOE complex,

as well as taking advantage of DOE's HPC systems at LBNL. The activities that will enable a fully functional NMDC are organized into four aims, which include: leveraging existing ontology mapping software and curation resources to enable automated annotation of standardized metadata; developing microbiome workflows for metagenome, metatranscriptome, metaproteome, and metabolomics data processing leveraging HPC systems, and integrating the execution of these pipelines to produce NMDC-compliant data products; developing data registration, indexing, and access services to link data through a suite of publicly available APIs; and, developing communication and sustainability strategies to assess current and future needs and capabilities to empower users, collaborate on web-based interfaces for search functionality, and promote the NMDC to the larger scientific community.

Funding statement

The work conducted by the National Microbiome Data Collaborative is supported by the Office of Science of the U.S. Department of Energy.

Arbuscular Mycorrhizal Fungi Transport Water to Host Plants

Anne Kakouridis[1*] (annekakouridis@berkeley.edu), Jack A. Hagen[1], Megan P. Kan[1], Stefania Mambelli[1,2], Lewis J. Felman[1] and **Mary K. Firestone**[1,3]

[1]University of California Berkeley, Berkeley, CA; [2]Center for Stable Isotope Biogeochemistry, Berkeley, CA; [3]Lawrence Berkeley National Laboratory, Berkeley, CA.

Project Goals: Our project explores the potential of arbuscular mycorrhizal fungi (AMF) to enhance the tolerance of an annual grass to soil desiccation. We specifically ask whether AMF can access water in soil unavailable to plants and transport that water across air gaps to host plants, thus improving plant-water relations. We used stable isotopes to track and quantify water flow from the soil, through fungal hyphae, to plants. In addition, we used fluorescent dyes to elucidate the location and mechanisms of water transport in fungal hyphae. This research substantially deepens our understanding of AMF functioning and AMF-plant interactions. Our findings have possible ramifications for the management of plant drought tolerance in the context of climate change.

Arbuscular mycorrhizal fungi (AMF) form symbiotic associations with 80% of surveyed land plant species and are well-recognized for mobilizing and transferring nutrients to plants. Yet AMF also perform other essential functions, notably improving plant-water relations. Some research attributes this solely to improved plant nutrition and better ability to osmoregulate when plants are partnered with AMF. However, we hypothesize that AMF can also directly transport a significant amount of water to their plant hosts. We used stable isotopes and fluorescent dyes to track and measure water transport by AMF to plants in a greenhouse experiment. *Avena barbata* inoculated with *Rhizophagus intraradices* was planted in one compartment (the 'plant compartment') of two-compartment microcosms. The second compartment (the 'no-plant' or 'soil' compartment) was separated from the first by an air-gap and either an 18 μm mesh, thus excluding roots but allowing AMF hyphae in, or a 0.45 μm mesh, thus excluding both roots and hyphae. ^{18}O-labeled water and fluorescent dyes (one membrane permeable, the other membrane impermeable) were injected into the soil in the no-plant compartment. Transpired water from *A. barbata* shoots was collected, measured, and tested for ^{18}O. Hyphae and *A. barbata* roots were collected and imaged with fluorescence microscopy to identify the presence and location of the dyes.

Plants with AMF able to access the soil compartment transpired twice as much water, and the transpired water had a δ^{18}O value three times higher, compared to plants with AMF excluded from the soil compartment. We detected the presence of both fluorescent dyes in the roots of plants with AMF accessing the soil compartment, but not in the roots of plants with AMF excluded from the soil compartment. The membrane permeable dye was broadly visible throughout the roots and nearby soil, suggesting hyphal transport and diffusion out. The membrane impermeable dye was visible in/on hyphae within roots only, indicating that it travelled between the hyphal cell membrane and cell wall, since it cannot cross cell membranes. Based on these findings, we conclude that *R. intraradices* transported a significant amount of water to *A. barbata* and likely carried the water more efficiently through hyphae apoplast than symplast. We believe our results are the first to show that AMF can directly transport a

significant amount of water to host plants. We suggest that AMF can play an important role in improving plant-water relations, not only indirectly, but also by acting as direct extensions of root systems, bridging air-gaps, penetrating micropores, and accessing water in soil not easily accessible to roots, and thereby increasing the amount of soil water available to plants.

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-SC0010570 and DE-SC0016247 at UC Berkeley.

Connecting switchgrass-microbe-soil interfaces for sustainable bioenergy crop production on marginal soils: stable-isotope labeling, genomics and exometabolomics

Nameer Baker,[3] Kateryna Zhalnina,[1,2] Jacob Jordan,[3] Yuan Wang,[4] Erin Nuccio,[5] Ella Sieradzki[3], Don Herman[3], Benjamin P. Bowen,[1,2] **Jizhong Zhou,**[6] **Kelly Craven**,[4] **Malay Saha**,[4] **Javier Ceja Navarro**,[1] **Jennifer Pett-Ridge**,[5] **Eoin Brodie**,[1,3] **Trent Northen**[1,2] and **Mary Firestone**,[1,3]* (mkfstone@berkeley.edu)

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]Joint Genome Institute, Berkeley, California; [3]University of California, Berkeley; [4]Noble Research Institute, Ardmore, Oklahoma; [5]Lawrence Livermore National Laboratory, Livermore, California; [6]The University of Oklahoma, Norman

Project Goals: Our project studies molecular mechanisms driving beneficial plant–microbial interactions in switchgrass cv. Alamo (*Panicum virgatum*) growing in marginal soils under a range of resource limitations. Genomics and chemistry of plant-microbe interactions are examined during switchgrass establishment to gain insight into how symbiotic and associative microbes improve plant performance and soil carbon persistence in marginal soils. We are integrating focused (single plant-microbe pairing) and 'community' systems biology approaches to examine the complex interplay among plants, microbes, and their physio-chemical environment.

Switchgrass (SG) is a perennial grass and one of the most promising bioenergy crops in the United States. SG is well-adapted to resource-limited environments, such as low-nutrient or droughty soils. We hypothesize that SG releases exudates to recruit beneficial rhizosphere microbial communities. These microbial communities are associated with improved plant performance during environmental stresses and increased C stocks in marginal soils. However, the molecular mechanisms that underlie SG-microbe relationships in the rhizosphere remain poorly understood. In this study, we analyzed complex exudate chemistry of SG grown under nutrient stresses, and link SG exudates to the abundance, activity and substrate preferences of bacteria found in the SG rhizosphere.

We used field, greenhouse and controlled reductionist approaches to dissect SG-microbe interactions. We performed a greenhouse study to investigate plant-associated microbes connected to the improved plant performance and C and N transformations in marginal soils, and used $^{13}CO_2$ stable-isotope labeling, genomics, and exometabolomics to identify the nature and dynamics of C in the SG rhizosphere. For this study, we reconstituted three nutrient-deplete soil horizons with different physio-chemical properties in one-meter deep soil columns in greenhouse mesocosms. We grew SG clonal plants in these reconstructed soil horizons and subjected these mesocosms to N and/or P amendments and two watering regimes. We collected rhizosphere and bulk soil from $^{12}CO_2$- and $^{13}CO_2$-labeled SG and extracted DNA and exometabolites.

Analysis of switchgrass exudate chemistry. We used liquid chromatography mass spectrometry-based exometabolomics (LC-MS) to analyze small organic molecules (metabolites) released by SG grown in soil mesocosms and hydroponic systems. We identified specific molecules exuded by SG during phosphorus (P), nitrogen (N) stresses and water limited conditions. We found that aromatics such as shikimic acids, salicylic acids, were more abundant in the exudates of N-stressed SG plants. In contrast, N-containing compounds such as amino acids and nucleosides were depleted in this treatment. During P stress, SG increased production

of carboxylic organic acids (e.g. succinic, malic acids), whereas osmolytes were more abundant in water-limited treatments.

Linking plant chemistry to microbial community composition. To connect rhizosphere metabolites and SG rhizosphere communities, we also identified dynamics of different microbial groups under N, P, N/P and reduced water treatments in the mesocosms. We found that N addition decreased the relative abundances of Actinobacteria and Bacteroidetes. Phosphorus treatment had lower relative abundances of Alphaproteobacteria and Verrucomicrobia. Planctomycetes, Chloroflexi and TM6 significantly decreased when any type of nutrients were added. A 50% reduction in water resulted in decreased relative abundances of Enterobacteriales and Planctomycetales.

Using a reductionist approach, we sought to disentangle the complexity of interactions between exudates and SG rhizosphere microbial communities, and to identify specific linkages between exudates and microbial substrate preferences. First, we isolated and characterized 300 bacteria from the SG rhizosphere. We then analyzed the abundances of these isolates during SG development in soil and identified bacteria associated with SG development. Finally, we used exometabolomics to connect stress molecules released by SG under N and P stresses to the metabolite uptake preferences of the rhizosphere isolates and related plant growth phenotypes. Our results indicate that during N and P nutrient stress SG mediates the release of stress-specific exudates to regulate the abundance of specific microbial taxa in the rhizosphere.

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-SC0014079 to the UC Berkeley, Noble Research Institute, the University of Oklahoma, the Lawrence Livermore National Laboratory and the Lawrence Berkeley National Laboratory. Part of this work was performed at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231 and at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Cross-Kingdom Interactions: the Foundation for Nutrient Cycling in Grassland Soils

Joanne B. Emerson[1]* (jbemerson@ucdavis.edu), Mengting Yuan,[2] Kateryna Zhalnina,[3] Graeme Nicol,[4] Erin Nuccio,[5] Trent Northen,[3] Jizhong Zhou,[6] Nhu Nguyen,[7] Javier A. Ceja-Navarro,[3] Jennifer Pett-Ridge,[5] and **Mary K. Firestone**[2,3]

[1]University of California, Davis, California; [2]University of California, Berkeley, California; [3]Lawrence Berkeley National Laboratory, Berkeley, California; [4]Universite Lyon, Lyon, France; [5]Lawrence Livermore National Laboratory, Livermore, California; [6]University of Oklahoma, Norman, Oklahoma; and [7]University of Hawaii, Manoa, Hawaii

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 (*e.g.*, by fungal hyphae and fauna) determine the occurrences and rates of N-cycling processes.

Decades of research have revealed key microbial mediators of terrestrial nutrient cycling, their edaphic sensitivities, and the functional genes and enzymes involved. While some aspects of bacterial, fungal, and microfaunal mediation of nutrient cycling are reasonably well understood, these and other organisms interact in a complex biotic milieu, and we know little about how such interactions shape nutrient cycling in soil. Considering interactions among viruses, bacteria, archaea, arbuscular mycorrhizal fungi (AMF), saprotrophic fungi, microfauna, and plant roots, we will explore the effects of predation, competition, and cooperative/antagonistic interactions on terrestrial N- and C-cycling. As an introduction to our new project, here we highlight relevant results from our previous work and describe several of our planned experiments.

Building on our large-scale field sampling and manipulative field experiments at the University of California's Hopland Research and Extension Center (HREC) over the last ~20 years, we will continue our work on HREC's Mediterranean grassland soils, which are dominated by the annual grass, *Avena barbata*. In our most recent precipitation manipulation experiment at HREC, *A. barbata* was labeled with $^{13}CO_2$ to trace C flow from the plant into active soil biota and to determine the impacts of reduced precipitation on soil food web dynamics and the C cycle. In Spring 2020 we will leverage the existing infrastructure and conduct a $^{13}CO_2$ field labeling and ^{15}N pool dilution experiment to establish cross-kingdom interactions and connect them to their gross N-cycle effects in the field. To identify how different N-cycling microbial groups, fauna, and viruses respond to N inputs, we will add a urea treatment. We will follow changes in microbial community composition and activity through SIP-enabled metagenomics, amplicon sequencing, and quantification of marker genes for N-cycling processes. Faunal communities

will be tracked via 18S rRNA gene sequencing, and viral dynamics will be measured via viral metagenomics (viromics). We will use liquid chromatography mass spectrometry (LC-MS) to identify changes in metabolites in response to urea addition and to link the dynamics of these molecules to changes in the relative abundances of microbial, viral, and faunal populations.

In a recent greenhouse experiment, we studied both rhizosphere and bulk soils associated with multiple growth stages of the annual grass, *Avena fatua*, labeled with $^{13}CO_2$. Analyses of metagenomes, ITS regions, and 16S rRNA genes revealed a strong influence of root development on the composition and associations of microorganisms and highlighted AMF as extensions of plant roots in mediating soil microbial interactions and associated nutrient cycling. Our upcoming greenhouse experiment will build on these results to measure microbial community dynamics and activity across trophic scales in four different soil compartments (rhizosphere, detritusphere, hyphosphere, and bulk soil), along with differences in N-cycling processes between compartments and how microbiota interact and/or travel between compartments to impact C- and N-cycling.

We expect that viral lysis will impact N-cycling processes, both indirectly through mortality of host populations responsible for specific N metabolisms and directly through the release of organic N-rich cellular contents. Some viruses of bacteria and archaea are capable of switching replication strategies between lysogeny (integration in microbial host genomes as prophages) and lysis, depending on host and environmental cues. We will perform laboratory experiments to link viral populations to their replication strategies, and then we will follow these populations in our field and greenhouse omic data to identify the conditions under which prophage integration or lysis is favored for a particular viral population and the conditions under which specific hosts can lyse. We will focus particularly on changes in N-cycling processes that could drive switches in viral replication strategies or that could result from host lysis.

We have previously shown a strong influence of root development on the composition and associations of microorganisms, including specific bacterial-fungal-protozoan co-occurrence patterns revealed through network analysis. Based on these previously identified patterns, we will prepare simple rhizosphere communities and iteratively add trophic complexity to elucidate changes in C- and N-cycling processes across trophic scales. Specifically, plants inoculated with AMF in EcoFab systems will be iteratively exposed to different bacterial, microfaunal, and viral populations, all isolated from HREC field soil. We will follow changes in community composition, dynamics of genes involved in N-cycling processes, and CO_2 and N_2O emissions.

Combining multiple novel stable isotope techniques with current molecular methods and modeling will allow us to explore, map, and quantify the complex web of biotic interactions that mediate and control N-cycling in soil.

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, UC Davis, University of Hawaii, and University of Oklahoma. 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 U.S. Department of Energy Contract DE-AC52-07NA27344.

Effects of Switchgrass Cultivation on Deep Soil Carbon Stock and Long-term Carbon Dynamics in Marginal Lands

Jialiang Kuang[1]* (kjialiang@ou.edu), Colin Bates[1], Gangsheng Wang[1], Eric Slessarev[2], Karis McFarlane[2], Christina Ramon[2], Erin Nuccio[2], Amrita Bhattacharyya[2,3], Arthur Escalas[4], Lauren Hale[5], Yuan Wang[6], Don Herman[3], Liyou Wu[1], **Peter S. Nico**[3], **Malay Saha**[6], **Kelly Craven**[6], **Jennifer Pett-Ridge**[2], **Jizhong Zhou**[1], and **Mary Firestone**[7]

[1]The University of Oklahoma, Norman; [2]Lawrence Livermore National Laboratory, Livermore, CA; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]University of Montpellier, Montpellier, France; [5]USDA-ARS, Parlier, CA; [6]Noble Research Institute, Ardmore, OK; [7]University of California, Berkeley, CA.

Project Goals: Switchgrass (*Panicum virgatum* L.) is a perennial C_4 grass native to the tallgrass prairies and a most promising feedstock in the U.S. for bioenergy production. Capable of abundant biomass yield with minimal fertilizer or water, switchgrass can survive on marginal soils, and even thrive once established. We hypothesize that successful establishment and sustainable cultivation of switchgrass in marginal soils is in part enabled by beneficial plant-microbial interactions. We are investigating the succession of rhizosphere microbial communities, and ecosystem-scale effects of high- and low-performing switchgrass plants grown in nutrient-limited soils in southern Oklahoma. The outcome of this research will provide a better genomic basis for switchgrass cultivation in marginal soils, expand our knowledge of the interactions between soil microbiomes, plants and ecosystems, and ultimately guide efforts for translation into agronomic row crops.

Accrual of soil organic carbon (SOC) acts as a potential carbon sink, helping to mitigate anthropogenic C emissions. We hypothesize that cultivation of deep-rooted perennial crops, such as switchgrass, could provide long-lasting carbon sequestration, that enhances deep soil C stocks via root biomass inputs. But there may be trade-offs associated with this strategy: the preexisting soil organic matter (OM) pool may be displaced if rates of microbially-stimulated OM degradation increase in response to deep C inputs. In surface soils, many studies have shown 'priming', i.e. that inputs of fresh root biomass and exudates increases OM accessibility and stimulate microbial activity. However, it is unknown whether deep soil mineral-associated organic C (MOC) may be more vulnerable to microbial mineralization under switchgrass cultivation. Our study aims to address this fundamental uncertainty regarding the effects of establishment and sustainable cultivation of switchgrass on deep soil C stocks and long-term C dynamics in marginal lands.

We collected deep core soils (0 to ~3 meters) from two N and P poor sites (silt-loam and clay-loam) in Oklahoma, USA with long-term (>10 years) cultivation of perennial switchgrass and adjacent fields managed with annuals crops. Soil cores were collected from three replicate locations within each site and cropping system. We measured soil C stocks, ^{14}C, ^{13}C and ^{15}N, soil texture and chemistry, root biomass, and microbial community composition. At the silt-loam site with low native SOC stocks, switchgrass cultivation caused a 60-80% increase in the median SOC stock and a shift towards more modern ^{14}C values. At the clay-loam site, where initial SOC stocks where higher, switchgrass root biomass inputs had no significant effect on the overall SOC stock. At both sites, $\delta^{13}C$ of all soil depths increased with switchgrass cultivation, reflecting additional inputs of C_4 organic matter into the soil. Microbial community composition was significantly different with depth at both sites and responded to switchgrass root inputs in the

low-carbon site. These results partially support our hypothesis that deep-rooted perennial switchgrass can enhance deep soil C stocks in a relatively short time frame.

In addition to these bulk soil measurements, we set up the long-term monitoring plots in both the silt-loam and clay-loam sites and measured soil respiration monthly between perennial switchgrass plots and adjacent fields managed with annuals crops. We estimated continuous root biomass inputs between topsoil and subsoil layers based on satellite-measured gross primary production (GPP) and measured root biomass distribution with depth. We also measured heterotrophic respiration from a 60-day incubation in the lab using the deep core soils incubated with additional ^{13}C-labeled plant materials. During this incubation, we found that C decomposition was different across sites, root systems, and depths, suggesting that the different responses of soil microbial communities to the root biomass input can result in differential community level C degradation dynamics.

We are currently applying ecosystem modeling to estimate whether root biomass input into the subsoil layer may impact long-term C dynamics and accelerate the old-C (MOC) loss. For this purpose, we improved our MEND (Microbial-ENzyme Decomposition) model to represent multiple soil layers and used it to simulate the C processes along the soil profile. We are now constraining and optimizing the parameters of MEND to fit our observed CO_2 fluxes from the lab and the field. The ongoing model analyses will further project the changes of different C pools for a longer time-period (such as 100 years) to assess the potential release of the relatively stable old-C in deep soils. Our research will provide new insights into the prediction of deep soil C accrual after switchgrass cultivation by comprehensively considering microbially mediated soil C dynamics along the soil profile.

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-SC0014079 to the UC Berkeley, Noble Research Institute, the University of Oklahoma, Lawrence Livermore National Laboratory and Lawrence Berkeley National Laboratory. Part of this work was performed at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231 and at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Microbial and Viral Niche-Differentiation in Time-Resolved Metatranscriptomes from Rhizosphere and Detritusphere Soil

Ella Sieradzki*[1] (ellasiera@berkeley.edu), Erin Nuccio[2], Evan Starr[1], Jill Banfield[1], **Mary K. Firestone**[1], and **Jennifer Pett-Ridge**[2]

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

Project goals: Our primary research goals are to identify and quantify the importance of key interactions including predation, competition, and cooperation as controllers of (i) N and C mineralization/immobilization, (ii) depolymerization of macromolecular organic N compounds, (iii) nitrification and denitrification, and (iv) N and C transfers by fungi and fauna that enhance or constrain N-cycling processes. Our primary hypotheses are: (1) Viral infection of bacteria, fungi, and fauna initiates cell lysis and plays a significant role in macromolecular depolymerization, and N-mineralization and culls specific communities mediating N-transformations. (2) Faunal predation mobilizes biomass nutrients present in microbiota, differentially affects specific prey populations, and potentially transports viruses, bacteria and fungi across soil habitats. (3) Fungi mediate C transfer, transport bacteria, shape bacterial N-transformation, and themselves are key to depolymerization, mineralization, immobilization, and denitrification. (4) Dominant interactions differ by soil habitats, affecting N process rates.

The rhizosphere is a hotspot for microbial activity and biomass and the entry point for organic carbon. Plants deposit a significant proportion of their photosynthates into soil as root biomass or exudates, and plant-derived polymeric carbohydrates such as cellulose and hemicellulose are the most abundant polysaccharides in soil. Additionally, breakdown and utilization of macromolecular nitrogen may be mediated in soil near roots. Rhizodeposits create a high resource, high activity environment, and stimulate microbial succession as roots grow and senesce, potentially selecting for organisms that benefit mineral nutrition and plant health. This bloom of microbial growth may trigger a succession of viral populations in response. Rhizodeposits may also stimulate depolymerization by cellulases, chitinases, proteases, and carbohydrate degradation genes. Using *Avena fatua*, a common annual grass, we analyzed time-resolved metatranscriptomes to compare microbial functional dynamics related to C and N processing in rhizosphere, detritusphere, and combined rhizosphere-detritusphere habitats. We also used this dataset to delve into the largely unknown realm of soil RNA viruses and explore their diversity and spatiotemporal dynamics.

During three weeks of root growth, microbial community composition shifted only slightly, whereas gene expression profiles changed significantly, indicating that mRNA is more sensitive to changes in environmental conditions and can reflect community shifts before they are detectable by diversity markers. With carbohydrate active enzymes (CAZy) identified in the metatranscriptomes, we used hierarchical clustering and mapping to a site-specific metagenome to identify functional guilds. These guilds revealed taxa specializing in rhizosphere carbohydrate degradation vs. taxa that specialize in detritus degradation and, surprisingly, a guild that specialized in the breakdown of aging roots. Rhizosphere and detritusphere guilds expressed enzymes for cellulose and xylan degradation and their byproducts. This indicates that complex

cross-feedings networks could promote coexistence within highly interconnected rhizosphere communities (Nuccio et al., 2020).

Nitrogen cycling gene expression varied over time and in the presence/absence of detritus, with distinct rhizosphere, detritusphere and aging root functional guilds. The taxonomy of these guilds often mirrored those previously identified using CAZymes. Expression of extracellular proteases was significantly lower in the rhizosphere compared to bulk soil. Ammonia oxidation (AO) transcripts were dominantly archaeal and more highly expressed in bulk soil, which may reflect competition for ammonium with plant roots. While archaeal ammonia oxidation genes were several-fold more abundant than their AO bacterial counterparts, expression of archaeal ammonia monooxygenase was higher by orders of magnitude, implying that the overexpression of AO genes cannot be attributed simply to a higher abundance of ammonia oxidizing archaea.

Our large metatranscriptomics dataset also allowed us to identify RNA viruses—which are understudied in most environments and have been largely ignored in soil (with the exception of plant pathogens). There is almost no knowledge of the diversity and host range of these viruses, which could have a significant impact on microbial community structure and soil carbon cycling. We targeted the RNA-dependent RNA-polymerase (RdRp) gene, which is universal to RNA viruses. We improved the set of hidden Markov models (HMMs) used to identify this gene and significantly increased the known diversity of RNA viruses in existing databases. Most of these viruses likely infect abundant taxa such as fungi and Proteobacteria, indicating that they may significantly affect microbial dynamics in soil. The temporal dynamics of RNA-viruses we identified also indicate that they are replicating, and therefore infecting their hosts, potentially leading to a release of OC into soil. The diversity of RNA viruses as well as of potential hosts was structured by the presence or absence of detritus (Starr et al., 2019).

Our analyses are a step towards understanding multitrophic interactions in soil and emphasize the need to think beyond community structure. Gene expression studies may be valuable in elucidating the complex to-and-fro between plants and their associated microbial and viral communities, as well as identifying the guilds of organisms that orchestrate C and N processing in soil.

References

Nuccio, E. E., Starr, E., Karaoz, U., Brodie, E. L., Zhou, J., Tringe, S. G., Malmstrom, R. R., Woyke, T., Banfield, J. F., Firestone, M. K., & Pett-Ridge, J. (2020). Niche differentiation is spatially and temporally regulated in the rhizosphere. *The ISME Journal*. https://doi.org/10.1038/s41396-019-0582-x

Starr, E. P., Nuccio, E. E., Pett-Ridge, J., Banfield, J. F., & Firestone, M. K. (2019). Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(51), 25900–25908.

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-SC0010570. Part of this work was performed at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231 and at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Multitrophic and Metabolite Responses to Drought in Grassland Soils

Javier A. Ceja-Navarro[1*] (JCNavarro@lbl.gov), **Kateryna Zhalnina**[1], Abelardo Arellano[1], Leila Ramanculova[1], Katerina Estera[2], **Eoin L. Brodie**[1,2], **Trent R. Northen**[1], **Jennifer Pett-Ridge**[3], **Mary K. Firestone**[1,2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of California, Berkeley, CA; [3]Lawrence Livermore National Laboratory, Livermore, CA

Project Goals: Our project addresses both a fundamental understanding of carbon (C) cycling as mediated by multi-trophic interactions in the rhizosphere and the potential impacts of altered precipitation regimes on these interactions. Specifically, our work employs the use of stable isotopes to identify and quantify pathways of C-flow through multi-trophic interactions in the rhizosphere. Primary goals are to broaden knowledge of the roles played by multi-trophic interactions in terrestrial C cycling and to discover if drought alters the interactions and/or C cycling. This research will substantially expand our knowledge of soil microbial ecology, belowground food webs, and terrestrial C cycling under a changing climate.

In all terrestrial ecosystems, populations of organisms such as protozoa and nematodes interact with the soil's free-living and plant-associated bacterial and fungal populations. Together these multi-trophic interactions are key regulators of soil carbon and nutrient cycling processes. To predict the responses of these processes to environmental change, we have optimized multiple approaches to characterize the belowground populations from each trophic level, while also incorporating metabolomics analyses to associate changes in community composition with biological activity. Our primary objective is to illuminate the contribution of bacterial, archaeal, fungal, and other eukaryotic communities to carbon and nutrient cycling in soil by applying multi-omics approaches, stable isotope tracing, and field manipulations to the development of mechanistic network theory of material flows in the soil.

At a Mediterranean annual grassland undergoing simulated drought (-50% average precipitation), soil was collected, and nematode and protozoan populations extracted using gradient centrifugation. Specimens were concentrated by filtration and their DNA extracted. Bacterial and fungal populations were characterized by amplicon sequencing from whole soil DNA extracts, and both diversity and community composition of the analyzed populations showed changes in response to drought. After five months of drought, while the diversity of nematodes remained unchanged, their community composition changed significantly, with 10-15% of variance explained by the water treatment. Protozoan and bacterial communities showed higher diversity in soils under ambient conditions, and protozoan community structure was significantly altered. Fungi, on the other hand increased diversity in soils under drought along with significant changes in community composition. Differential abundance analyses were identifid specific trophic groups that responded to drought. Bacterial groups from the genus *Arthrobacter*, *Gemmatimonas*, *Clostridium*, *Hydrogenispora*, and *Anaeromyxobacter* together with protists such as *Lagenidium*, *Eustigmatos*, *Phytophtora*, and *Theratromyxa* declined under drought conditions. Fungal populations including *Podospora*, *Rozellomycota*, and *Olpidium* increased under drought

Co-occurrence analyses using Random Matrix Theory was used to define network structure for all analyzed trophic groups. Network complexity was found to be higher in soils under ambient water conditions than under drought. Predicted modules in the interaction network will be experimentally tested using populations of organisms isolated from the field.

We used liquid chromatography mass spectrometry-based metabolomics (LC-MS) to identify differences in metabolites profiles from our field drought experiment. Our results show that the drought treatment significantly decreased the relative abundance of primary metabolites such as salicylic acid, threonine, guanine and disaccharides, and increased the abundance of metabolites known as organic osmolytes (e.g. aspartic acid, asparagine, sorbitol). Metabolite exchange impacts the abundance and physiology of organisms at multiple trophic levels and could help define multi-trophic interactions within rhizosphere communities. Further analyses in metabolite profiles will aim to identify molecules associated with signaling for microbial predation, exploratory microbial growth, and biofilm production.

In field $^{13}CO_2$ labeling of the *Avena fatua* plants was carried out and rhizosphere samples with attached soil were collected for DNA extraction and SIP-fractionation. Currently, labeled DNA fractions are being used to identify protozoan groups that had access to labeled C through predation of microbial cells.

The approaches developed here provide the foundation for quantitative molecular understanding of how soil trophic networks and cross-kingdom interactions influence C and N cycling and ecosystem resilience to environmental change.

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 U.S. Department of Energy Contract DE-AC52-07NA27344.

Spectroscopic Diagnosis of Plant Phosphorus Availability and Relationship to Tissue Chemistry and Productivity of a Bioenergy Feedstock

Zhao Hao[1], Yuan Wang[2], Na Ding[2], **Wolf Scheible**[2], **Kelly Craven**[2], **Michael Udvardi**[2], **Mary Firestone**[3], **Eoin Brodie**[1,3*]

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]Noble Research Institute, Ardmore, Oklahoma; [3]University of California, Berkeley, California

Project Goals: Switchgrass (*Panicum virgatum* L.) is a perennial C_4 grass native to the tallgrass prairies and a promising feedstock for U.S. bioenergy production. Capable of abundant biomass yield with minimal fertilizer or water, switchgrass can survive on marginal soils, and even thrive once established. We hypothesize that successful establishment and sustainable cultivation of switchgrass in marginal soils is in part enabled by beneficial plant-microbial interactions. We are investigating the succession of rhizosphere microbial communities, and ecosystem-scale effects of high- and low-performing switchgrass plants grown in nutrient-limited soils in southern Oklahoma. The outcome of this research will provide a better genomic basis for switchgrass cultivation in marginal soils, expand our knowledge of the interactions between soil microbiomes, plants and ecosystems, and ultimately guide efforts for translation into agronomic row crops.

Economic bioenergy/bioproduct feedstock production will likely require the utilization of marginal soils. Although crops such as switchgrass (*Panicum virgatum L.*) may obtain a fraction of their nitrogen requirement through associate N-fixation, phosphorus (P) is exclusively derived from soil or added fertilizer. Soil P availability is chemically complex and varies significantly based on pH, mineralogy, hydrology and the chemistry of inputs - resulting in heterogeneous P availability across soils of production systems. Plants and microbes have evolved mechanisms to liberate mineral, and organic associated P – these mechanisms frequently require allocation of resources in the form of carbon cost. Better understanding of these mechanisms calls for high-throughput, non-invasive yet accurate prediction tools we have developed, and will lead to improved biomass yield and feedstock quality for bioenergy feedstock development program.

In this work, we used infrared spectroscopy to monitor plant P dynamics and tissue chemistry during switchgrass growth, establishing a linkage between P availability and spectral signatures using plants grown in a controlled laboratory environment. We used these data as a benchmark, assisted by machine learning algorithms, to evaluate P availability in the switchgrass plants cultivated in two marginal soils in Oklahoma. This represents a relatively noninvasive approach to quantifying inorganic and organic phosphorus content and other chemical constituents of leaf tissue during plant growth. We used these data to observe relationships between P availability, P uptake and re-allocation dynamics and biomass productivity in subsequent growing years.

We report three important findings: (1) In response to P stress (but not N stress), switchgrass cell wall composition had increased lignin concentration, leading to lower cellulose:lignin ratios—an important index of biofuel feedstock quality. P-limitation also appeared to lead to lipid accumulation and a relative decrease in amide (protein) content. (2) In our field plots of mixed genotypes, plants growth in soils with higher P availability accumulated more organic P and re-allocated more P during pre-harvest senescence. Within each site, we observed significant variability in P pools, reflecting either heterogeneity in switchgrass cultivar genetics, subsurface

chemical properties or variability in soil-plant-microbial interactions that mobilize P. (3) Greater re-allocation of organic P during senescence was associated with increased biomass productivity of individual plants during the subsequent growing season.

Overall this work demonstrates the potential of high-throughput spectroscopic approaches to diagnose plant nutrient availability, its consequences for feedstock compositional quality and productivity, and the relationships with chemical and biological heterogeneity in soils of production systems.

This work was funded by the US Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number DE-SC0014079 to the UC Berkeley, Noble Research Institute, the University of Oklahoma, the Lawrence Livermore National Laboratory and the Lawrence Berkeley National Laboratory. Work performed at Lawrence Berkeley National Lab was under contract DE-AC02-05CH11231.

Succession of Rhizosphere Biotic Communities During Switchgrass Establishment in Marginal Soils

Yuan Wang[1]* (ywang@noble.org), Javier Ceja Navarro,[2] Erin Nuccio,[3] Daliang Ning,[4] Renmao Tian,[4] Travis Simmons,[1] Katerina Estera-Molina,[5] Christina Fossum,[5] Jialiang Kuang,[4] Colin Bates,[4] Lauren Hale,[4] Na Ding,[1] Josh Barbour,[1] Nameer Baker,[5] Abelardo Arellano,[2] Kateryna Zhalnina,[6] Eoin Brodie,[2] Trent Northen,[6] Wolf Scheible,[1] Michael Udvardi,[1] Jennifer Pett-Ridge,[3] Malay Saha,[1] Liyou Wu,[4] Jizhong Zhou,[4] **Mary Firestone[5]** and Kelly Craven[1]

[1]Noble Research Institute, Ardmore, Oklahoma; [2]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, California; [3]Lawrence Livermore National Laboratory, Livermore, California; [4]University of Oklahoma, Norman, Oklahoma; [5]University of California, Berkeley, California; [6]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California.

Project Goals: Switchgrass (SG; *Panicum virgatum* L.) is a perennial C_4 grass native to the tallgrass prairies and a most promising feedstock in the U.S. for bioenergy production. Capable of abundant biomass yield with minimal fertilizer or water, SG can survive on marginal soils, and even thrive once established. We hypothesize that successful establishment and sustainable cultivation of SG in marginal soils are, in part, enabled by beneficial plant-microbial interactions. We are investigating the succession of rhizosphere microbial communities, and ecosystem-scale effects of high- and low-performing SG plants grown in nutrient-limited soils in southern Oklahoma. The outcome of this research will provide a better genomic basis for SG cultivation in marginal soils, expand our knowledge of the interactions between soil microbes, plants, and ecosystems.

In the soils surrounding roots (rhizosphere), biotic, chemical and physical drivers enrich for specific biotic communities. To generate a better understanding of the plant-microbe interactions that facilitate switchgrass cultivation in marginal soils, we investigated the succession of rhizosphere communities at multiple trophic levels during switchgrass establishment in marginal soils. Two research farms (Red River Farm and Third Street Farm) featuring 'marginal' nutrient soils were selected from Southern Oklahoma, and genetically diverse switchgrass seedlings (lowland ecotype Alamo) were planted at both sites in the spring of 2016. Rhizosphere soils were sampled from sixty plants (thirty from each site) at five time points, corresponding to different plant developmental stages. Plant growth was monitored at the same time. We then surveyed the rhizosphere communities by amplicon sequencing of phylogenetic marker genes specific to bacteria (16S), fungi (ITS), and soil protozoa (18S).

The two marginal soils surveyed in this study have distinct background communities. Regardless of the site effect, the bacterial, fungal, and protist communities all showed a significantly lower diversity in the rhizosphere than those found in the background soils at both sites. This indicates that switchgrass plants grown in the marginal soils exert a selective effect on the soil bacterial, fungal and protist populations. The composition of rhizosphere communities also differed from that of the background soils. Betaproteobacteria and Gammaproteobacteria were enriched, and Firmicutes, Verrucomicrobia and Planctomycetes were depleted in the rhizosphere at both sites. The enrichment of beta-proteobacteria in the rhizosphere was particularly dramatic at Red River site (silt loam). Rhizosphere fungi that were enriched included Eurotiomycetes (Red River, silt loam) and Leotiomycetes (Third Street, clay loam). Specific groups of protozoa were

also recruited to the rhizosphere, including those feeding on bacteria and fungi from the genera Rhogostoma, Cermononas, and Allas.

Rhizosphere bacterial, fungal, and protozoan communities all dynamically changed over the growing season. In the early growing season (T1-T3), the rhizosphere bacterial communities strongly differed from their background populations, while in the late plant developmental stages (T4 and T5), they became more similar to background communities. Moreover, specific bacterial OTUs were significantly correlated with plant biomass production at the first three time points. Analysis of ITS amplicons showed fungal communities in the background soils changed over time, and rhizosphere communities became more dissimilar from background communities at late plant vegetative growth stages (T2) and late in the growing season. The rhizosphere protist communities differed from those found in the background soils starting from two months after planting (T2); the largest difference occurred at the end of the growing season (T5).

We assessed if these differing community-level responses are due to different mechanisms of rhizosphere community assembly for bacteria, fungi, and protists. Both decreased stochasticity and increased selection were observed in the switchgrass rhizosphere for bacterial communities (Red River site) and fungal communities (Third Street site). This site-dependent rhizosphere selection might be related to the baseline of the soil communities: Red River soils exhibited generally high bacterial diversity, whereas Third Street soils have relatively high fungal diversity. Rhizosphere-enhanced selection of bacteria and fungi was most evident in later plant developmental stages (T3 and T4). Switchgrass roots also increased the importance of dispersal limitation in microbial community assembly, which was more obvious in Red River. In general, homogeneous selection mechanisms were more influential for bacteria rather than fungi or protists; heterogeneous selection was more critical for protists; and dispersal limitation showed a more noticeable impact on fungi.

In conclusion, during establishment in marginal soils, switchgrass plants recruited specific groups of bacteria, fungi, and protists into their rhizosphere, leading to less stochastic and more selected rhizosphere communities that are distinct from background soil communities in diversity, composition and structure. These rhizosphere communities were strongly influenced by plant developmental stages, and different microorganisms (bacteria, fungi and protists) had distinct successional patterns over the growing season. The specific assembly mechanisms of rhizosphere communities and their dynamic changes with the plant development may play an essential role in switchgrass establishment in marginal soils.

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-SC0014079 to the UC Berkeley, Noble Research Institute, the University of Oklahoma, the Lawrence Livermore National Laboratory and the Lawrence Berkeley National Laboratory. Part of this work was performed at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231 and at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Unravelling Rhizosphere-Microbial Interactions in the Rhizosphere of Alamo Switchgrass (*Panicum virgatum*) under Abiotic Stresses

Na Ding[1*](nding@noble.org), Nameer Baker[2], Yonatan Sher[2], Don Herman[2, 3], **Wolf Scheible**[1], **Michael Udvardi**[1], **Jennifer Pett-Ridge**[4], and **Mary Firestone**[2,3]

[1]Noble Research Institute, Ardmore, OK; [2]University of California, Berkeley, CA; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Lawrence Livermore National Laboratory, Livermore, CA

Project Goals: Switchgrass (*Panicum virgatum* L.) is a perennial C_4 grass native to the tallgrass prairies and a promising feedstock for U.S. bioenergy production. Capable of abundant biomass yield with minimal fertilizer or water, switchgrass can survive on marginal soils, and even thrive once established. We hypothesize that successful establishment and sustainable cultivation of switchgrass in marginal soils is in part enabled by beneficial plant-microbial interactions. We are investigating the succession of rhizosphere microbial communities, and ecosystem-scale effects of high- and low-performing switchgrass plants grown in nutrient-limited soils in southern Oklahoma. The outcome of this research will provide a better genomic basis for switchgrass cultivation in marginal soils, expand our knowledge of the interactions between soil microbiomes, plants and ecosystems, and ultimately guide efforts for translation into agronomic row crops.

In the rhizosphere, root exudation is a key process for C transfer into the soil, influencing the role of soil microbial communities in organic matter decomposition and in nutrient cycling. Root exudates increase the number and activity of soil microbes and fauna found in the rhizosphere through a myriad of complex interactions. Soil microorganisms depend upon plant C and, in turn, provide plants with nitrogen (N), phosphorus (P) and other mineral nutrients in part through decomposition of soil organic matter. We grew Alamo switchgrass (SG) in two types of greenhouse experiments to investigate how SG growth development and exudates shape rhizosphere microbial community composition, nutrient fractions (i.e. P species) and soil characteristics, as well as how these interactions are affected by abiotic stresses. In the first experiment, we tested the hypothesis that SG rhizosphere microbes enhance its access to organic P (Po) pools by stimulating root growth and/or liberating and solubilizing P from Po. In a second experiment, we assessed the response of soil extracellular polymeric substances (EPS) stocks to SG cultivation under varying N, P, and water stress, testing the hypothesis that soil microbes produce EPS (particularly polysaccharides) in response to nutrient and moisture limitation.

We collected marginal soils for the first experiment from three locations in Oklahoma (Ardmore 3rd Street, Red River, Anadarko). We amended soils with 50 mg/kg inorganic-P (KH_2PO_4, phosphate) or organic-P (phytate), and used non-amended soils as a control. Phosphate significantly enhanced plant biomass, total root length and surface area in all three soils after 8-weeks growth. Interestingly, phytate was an equally good source of P for plant growth promotion in 3rd Street soil. At all sites, phosphate increased resin-P in bulk soils, but substantially decreased resin-P in rhizosphere soils from 3rd Street and Anadarko soils, consistent with substantial uptake by plants. Phytate addition increased resin-P in the rhizosphere relative to bulk soil and decreased NaOH-P_o species in the rhizosphere relative to bulk soil in 3rd Street soil, consistent with mobilization by the microbiome and plant uptake of P from phytate in this soil.

SG accumulated more Pi at Red River in both shoots and roots than soils at other sites; phytate addition decreased Pi and increased Na^+ in shoots and roots at 3rd street. We also analyzed the bacterial and fungal microbiome from the bulk, rhizosphere and root associated compartment (endosphere + rhizoplane), respectively. Rhizosphere and bulk soils had a similar number of OTUs (α-diversity), and significantly more than were observed in the root associated compartment for both bacteria and fungi. Bacterial and fungal communities differed at the genus level across compartment and soil types. Soil type (rhizosphere or bulk), compartment and P treatment strongly influenced bacterial and fungal community composition. HCl-P and total P strongly influenced bacterial community composition at Red River. Organic P pools ($NaHCO_3$-Po and NaOH-Po) markedly impacted bacterial community composition in soils with added phytate. Interestingly, Tichomeriaceae were positively correlated with NaOH-Po in both rhizosphere and root-associated soils.

For a second greenhouse experiment, we collected marginal soils from Anadarko, OK. We grew ramets of a single switchgrass genotype for 18 weeks in three reconstituted field soil horizons, subjecting them to five treatments: added N, P, N plus P, 50% water, and controls. We then labeled plants either with $^{12}CO_2$ or $^{13}CO_2$ for 12 days before destructively harvesting plants and soil horizons. To determine how abiotic stresses altered the size and nature of EPS stocks and soil characteristics, we assessed root biomass, soil chemistry, EPS content, the monosaccharide composition of EPS, and the amount of water-stable aggregates.

Soils with both added N and P had the highest EPS content, root biomass, and percentage of water-stable soil aggregates. Multiple linear regression analysis showed root biomass was the most important determinant for soil EPS production, potentially by controlling carbon supply and diurnal changes in soil water stress. Root biomass and soil water potential were also correlated with water-stable aggregates, indicating that EPS concentration and soil aggregation have similar drivers in this soil. High mannose content confirmed the microbial origin of EPS. $^{13}CO_2$ labeling indicated that 0.18% of newly fixed plant carbon was incorporated into EPS. Analysis of field soils obtained via deep coring indicates that EPS concentrations are significantly enhanced under long-term switchgrass cultivation, suggesting a mechanism by which deep-rooted perennial grass cultivation may positively affect soil aggregation in soils with low organic material.

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-SC0014079 to UC Berkeley, Noble Research Institute, the University of Oklahoma, the Lawrence Livermore National Laboratory and the Lawrence Berkeley National Laboratory. Part of this work was performed at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231 and at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Creation of an Acyltransferase Toolbox for Plant Biomass Engineering

Craig A. Bingman,[1] **Brian G. Fox**[2*] (bgfox@wisc.edu),[1] Steven D. Karlen,[1] Shawn D. Mansfield,[2] Kirk Vander Meulen,[1] John Ralph,[1] Rebecca A. Smith[1]

[1]University of Wisconsin-Madison, Madison, Wisconsin USA; [2]University of British Columbia, Vancouver, British Columbia CANADA

Project Goals and Abstract:

The goal of this project is to expand the understanding and utility of the **acyl-CoA ligases** and **BAHD acyltransferases** in plant engineering. The impressive diversity of bioproducts created by their coordinated action includes simple aromatic and aliphatic esters, derivatives of flavanols, anthocyanidins, alkaloids, terpenoids, cutin, suberin, and many others. Our work has shown that monolignol esters can be polymerized into plant lignins with essentially no impact on growth and viability, but with profound positive impacts on the processing and energy content of the engineered plant. The ester-enriched lignin created by engineering efforts using monolignol ferulates, biosynthesized via hydroxycinnamoyl-CoA monolignol transferases, to produce so-called 'zip-lignin' is particularly encouraging.

Overall there is little experimental evidence for the breadth of substrate specificities, rates of catalysis, structural properties, stabilities, tissue localization, or *in planta* synergistic capabilities of the acyl-CoA ligases and BAHD acyltransferases, particularly for important bioenergy crops such as poplar, sorghum, and switchgrass. As such, these limited understandings represent major knowledge gaps, and potential bottlenecks to the successful engineering of new properties into plant cell walls and optimization of the biosynthesis of many other useful metabolic products.

We are proceeding with a systematic evaluation of all members of these important gene families in DOE's priority bioenergy plants. The proposed research adopts an integrative pipeline approach, including genomics, bioinformatics, cell-free translation, *in vitro* biochemistry, structural biology, optical and fluorescence microscopy, chemical synthesis, state-of-the-art methods for cell wall analysis including NMR spectroscopy, and cell-wall analytics. We will leverage in-depth knowledge of plant cell wall structure, lignin biosynthesis, plant transformation and cultivation, and practical aspects of processing plant material. Results will show how we have used this experimental pipeline to identify new acyltransferases with a breadth of new activities and have begun to introduce them into bioenergy plants and assess their impacts.

The major deliverable of this project will be experimentally validated lists of acyltransferase and ligase genes from DOE priority bioenergy plants with an accounting of their *in planta* abilities to modify the properties of bioenergy crop plants. Ultimately, our acyltransferase toolbox will provide an experimentally validated *'knock-out/knock-in'* list for performing precision engineering of plant metabolism to obtain new and improved properties.

This work is funded by the US Department of Energy, Department of Science, Office of Biological and Environmental Research, Genome-Enabled Plant Biology for Determination of Gene Function Funding Opportunity Number DE-FOA-00002060, project number 248370.

Using Systems Biology to Untangle the Complex Physiology of Bacterial Xylan Utilization

Andrew D. Blake, Nina R. Beri, Hadassa S. Guttman, Raymond Cheng, and **Jeffrey G. Gardner*** (jgardner@umbc.edu)

University of Maryland - Baltimore County, Baltimore, MD

Project Goals: Completion of the project will identify and characterize the physiologically relevant carbohydrate active enzymes required to consume the polysaccharides found in lignocellulose by the saprophytic soil bacterium *Cellvibrio japonicus*. Additionally, over the course of the project the utility of these enzymes, including assessment of novel functions, will be evaluated for biotechnology applications.

Lignocellulose degradation by microbes plays a central role in global carbon cycling, human gut metabolism, and renewable energy technologies. While considerable effort has been put into understanding the biochemical aspects of lignocellulose degradation, much less work has been done to understand how these enzymes work in an *in vivo* context. Here, we report a systems level study of xylan degradation in the saprophytic bacterium *Cellvibrio japonicus*. Transcriptome analysis indicated seven genes that encode carbohydrate active enzymes were up-regulated during growth with xylan containing media. In-frame deletion analysis of these genes found that only *gly43F* is critical for utilization of xylo-oligosaccharides, xylan, and arabinoxylan. Heterologous expression of *gly43F* was sufficient for the utilization of xylo-oligosaccharides in *Escherichia coli*. Additional analysis found that the *xyn11A*, *xyn11B*, *abf43L*, *abf43K*, and *abf51A* gene products were critical for utilization of arabinoxylan. Furthermore, a predicted transporter (CJA_1315) was required for effective utilization of xylan substrates, and we propose this unannotated gene be called *xntA* (xylan transporter A). Our major findings are (i) *C. japonicus* employs both secreted and surface associated enzymes for xylan degradation, which differs from the strategy used for cellulose degradation, and (ii) a single cytoplasmic ß-xylosidase is essential for the utilization of xylo-oligosaccharides.

Publications

1. Garcia CA, Narrett JA, Gardner JG. 2019. Complete Genome Sequences of *Cellvibrio japonicus* Strains with Improved Growth When Using α-Diglucosides. *Microbiology Resource Announcements*. 8(44).

2. Blake AD, Beri NR, Guttman HS, Cheng R, and Gardner JG**.** 2018. The complex physiology of *Cellvibrio japonicus* xylan degradation relies on a single cytoplasmic β-xylosidase for xylo-oligosaccharide utilization. *Molecular Microbiology*. 107:610-22.

3. Nelson CE, Attia MA, Rogowski A, Morland C, Brumer H, and Gardner JG**.** 2017. Comprehensive functional characterization of the Glycoside Hydrolase Family 3 enzymes from *Cellvibrio japonicus* reveals unique metabolic roles in biomass saccharification. *Environmental Microbiology*. 19:5025-39.

4. Nelson CE, Rogowski A, Morland C, Wilhide JA, Gilbert HJ, and Gardner JG**.** 2017. Systems analysis in *Cellvibrio japonicus* resolves predicted redundancy of β-glucosidases and determines essential physiological functions. *Molecular Microbiology*. 104:294-05.

This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0014183.

PROTIST PREDATION MEDIATES THE TEMPERATURE RESPONSE OF MICROBIAL COMMUNITIES

Jennifer Rocca[1], Andrea Yammine[1*], **Jean P. Gibert**[*1]
[1]Department of Biology, Duke University, Durham, NC, USA

Goals: To understand whether predation by microbial consumers mediates the temperature response of microbial communities.

Microbial communities are one of the largest biotic controls on the global carbon cycle. Understanding the top-down control that bacteria consumers, such as protists, exert on these microbial communities is thus essential to understand feedbacks on global climate. Theoretical and empirical studies have shown that temperature responses can be mediated by ecological interactions. But how microbial communities may respond to temperature in the presence of predators is poorly understood. Here, we examine the effects of protist predation on the structure and function of a freshwater microbial community at different temperatures. Microbial communities were collected from a seasonal pond at the Duke Forest, and cultured in the presence and absence of two generalist bacterial predators, the protists *Tetrahymena pyriformis*, and *Colpidium* sp, at two different temperatures (22°C and 25°C). Changes in community function were assessed over time through optical density and respirometry. Changes in structure were quantified through 16S rRNA amplicon sequencing. The larger protist (*Colpidium sp*) significantly suppressed bacterial biomass and mediated the temperature response of microbial respiration rates: predator presence decreased microbial respiration at high temperatures compared to lower temperatures. Temperature and time both had noticeable impacts on the structure and function of microbial communities, but predation did not affect structure. This suggests that predation impacts on microbial community function may occur through reductions in biomass, as observed, or other predator induced physiological responses (i.e., the landscape of fear). While bacterial impacts on the global carbon cycling are now routinely accounted for in global earth system models, biotic controls on microbial temperature responses are not. These results highlight the potential for important but largely overlooked impacts of consumption on the temperature response of microbial function, but not structure, which ought to be taken into account in future iterations of warming forecasting models.

A "Marionette" *S. cerevisiae* Strain to Control Metabolic Pathways

Marcelo Bassalo[1]* (bassalo@mit.edu), Chen Ye[1], Joep Schmitz[2], Hans Roubos[2], Ryan Gill[3], **Christopher Voigt[1]**

[1]Massachusetts Institute of Technology, Cambridge, MA; [2]DSM Biotechnology Center, Delft, The Netherlands. [3]Technical University of Denmark, Copenhagen, Denmark.

Project Goals: Optimizing metabolic networks often requires fine-tuning of gene expression levels to minimize buildup of toxic intermediates while maximizing productivity. Inducible promoters are a straight-forward strategy to systematically test different expression levels, providing levers to independently control targeted genes. However, the limited availability of orthogonal transcriptional sensors in the yeast, *Saccharomyces cerevisiae*, hinders their use to optimize an engineered biosynthetic pathway. Our objective is to expand the set of inducible promoters and develop a "Marionette" yeast strain, containing a genome integrated array of optimized sensors.

We have taken steps towards this "Marionette" strain by constructing and testing an initial set of 4 orthogonal sensors, engineered by placing bacterial operator elements into yeast core promoters. We then demonstrate "Marionette" in yeast by tuning a toxic metabolic pathway to produce the monoterpene Linalool, a valuable fragrance and fuel additive. Initially, a two-level factorial experiment was performed to uncover expression rules of the targeted genes. By incorporating these rules, we performed a second optimization round. Overall, this pilot test of expression profiles allowed us to explore the equivalent of ~300 kb of pathway variant constructs with a single genetic design. Finally, we also demonstrate staging order of operations on the controlled genes.

The ability to establish a synthetic metabolic pathway control to independently tune component genes will accelerate metabolic engineering cycles in yeast, enabling rapid testing of multiple expression levels that ultimately can be used to train learning algorithms and uncover rules for optimal pathway flux.

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science, and from Royal DSM.

Design and engineering of native regulatory networks in non-model microbes

Margaret Habib[1*](maha0174@colorado.edu), Emily Freed[1], Evan Johnson[1], Jeff Cameron[1], Min Zhang[2], Christopher Voigt[3], Adam Arkin[4], Carrie Eckert[1,2], and **Ryan Gill**[1,5]

[1]Renewable and Sustainable Energy Institute, University of Colorado, Boulder CO; [2]National Renewable Energy Laboratory, Golden, CO; [3]Massachussetts Institute of Technology, Boston, MA; [4]Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Danish Technical Institute, Copenhagen, Denmark.

Project Goals: Our objective is to develop a new standard for the engineering of microbial systems based on rational design, engineering, and optimization of hybrid regulatory networks. We will create both molecular tools and computational infrastructure to meet this overall goal in both *E.coli* and DOE relevant non-model organisms. Through increasing our understanding of native regulatory networks and then using heterologous synthetic circuits to recode them, we 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). Here we focus on expanding our work into non-model microbes to engineer designer organisms that have exquisite and predictable control architectures governing the expression of a range of valuable traits.

The realization of a sustainable bioeconomy requires our ability to understand and engineer complex design principles for the development of platform organisms capable of efficient conversion of cheap and sustainable feedstocks (e.g. sunlight, CO_2, non-food biomass) to biofuels and bioproducts at sufficient titers and costs. Despite recent advances in DNA synthesis allowing for the construction of small, synthetic genomes and the development of high-throughput genome editing and metabolic engineering tools in model microbes, our ability to design genomes and predict design principles for intricate functions such as tolerance are still limiting. In the proposed work, we will leverage our knowledge and expertise in cutting-edge synthetic biology techniques currently only available for model microbes by partnering with expert collaborators in the adaptation of these methods for DOE-relevant microbial systems, accelerating and expanding genome editing capabilities for metabolic engineering in these organisms. Through doing so, we will be able to more quickly uncover genotype to phenotype relationships to better engineer these microbes for optimal production of chemicals and fuels from renewable feedstocks.

Zymomonas mobilis ZM4 is a facultative anaerobe that is of interest for the fermentation of biomass sugars to biofuels and bioproducts, especially those branching from pyruvate. Collaborator Min Zhang demonstrated that carbon flux can be deviated from ethanol production at the pyruvate node into 2,3-butanediol as well as to farnesene (Yang et. al. 2016). They have generated a large volume of –omics data, providing insights into hydrolysate tolerance and feedstock utilization. Transformation efficiencies with a variety of vectors have been optimized, homologous recombination is performed routinely for genome engineering, and a number of functional promoters, both constitutive and inducible, have been characterized. In addition, heterologous expression of Cas9 and targeting gRNA was demonstrated to cure native plasmids in *Z. mobilis* (Cao et. al., 2017), representing initial proof of concept for CRISPR-based gene editing and regulation.

We are currently working to adapt genome engineering tools for *Zymomonas* to enable technologies such as CRISPR interference (CRISPRi) for gene knockdown and CRISPR-Enabled Trackable Geneome

Engineering (CREATE) for multiplexed editing. The investigators on this project have developed an array of vectors for Cas9/dCas9/Cas12a and gRNA expression in a range of organisms, and a large number of additional vectors are available through Addgene (http://www.addgene.org/). These vectors have been adapted, as necessary, for stable replication and predictable expression. Targeting, cutting, and recombination efficiencies are being evaluated and optimized for Cas9/Cas12a/gRNA utilizing an appropriate screen/selection method (antibiotic resistance, auxotrophy, counterselective, colorimetric, etc.). We will evaluate and optimize knockdown efficiencies using dCas9 in a similar fashion. Although we found that Cas9 expression was toxic in *Z. mobilis,*we have achieved high editing efficiency with the Cpf1/Cas12a system. We are additionally working on developing a dCas9 system for CRISPRi.

Once CRISPR-based genome editing is validated, we will adapt the CREATE method for targeting global regulators to mirror the regulator libraries utilized by other teams on this project (Ryan Gill and Chris Voigt) to expand the search space for regulatory control switches. Utilizing existing pathway maps for the organism, we will build pathway prediction models to identify global regulators to target. Information gleaned from the selection experiments (improved growth, tolerance, etc.) will guide further improvements to the models for development of synthetic regulatory networks by another team member (Adam Arkin). Together, this work will provide a blueprint for the development of systems to accelerate the engineering of non-model microbial systems as bioproduction chassis organisms.

References

1. Shihui Yang, Ali Mohagheghi, Mary Ann Franden, Yat-Chen Chou, Xiaowen Chen, Nancy Dowe, Michael E. Himmel & Min Zhang (2019) Metabolic engineering of Zymomonas mobilis for 2,3-butanediol production from lignocellulosic biomass sugars, Biotechnology for Biofuels, 9:189

2. Qing-Hua Cao, Huan-Huan Shao, Hui Qiu, Tao Li, Yi-Zheng Zhang & Xue-Mei Tan (2017) Using the CRISPR/Cas9 system to eliminate native plasmids of Zymomonas mobilis ZM4, Bioscience, Biotechnology, and Biochemistry, 81:3, 453-459

This research is supported by the Office of Biological and Environmental Research in the DOE office of Science, grant DE-SC0018368.

Engineering of Regulatory Networks for Improved C3-C4 Alcohol Tolerance and Production in *E. coli* and *S. cerevisiae*

Rongming Liu[1], Liya Liang[1], Emily Freed[1*] (emily.freed@colorado.edu), Alaksh Choudhoury[1] Adam Arkin[2], Christopher Voigt[3], Carrie Eckert[1,4], and **Ryan Gill**[1,5]

[1]Renewable and Sustainable Energy Institute, University of Colorado, Boulder CO; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Massachusetts Institute of Technology, Cambridge, MA; [4]National Renewable Energy Laboratory, Golden, CO; [5]Danish Technical Institute, Copenhagen, Denmark.

http://www.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. 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 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 fuel-grade compounds (e.g. C3–C5 alcohols). However, the titer and productivity of engineered strains that produce C3-C5 alcohols 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*/yeast 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* and yeast global regulatory networks. We performed growth competition experiments for library mutants

conferring increased C3-C4 alcohol tolerance and also screened for increased alcohol production. The titer of isopropanol and isobutanol in the best producing *E. coli* mutants were improved 3-fold and 2-fold compared to the parent strain, and the tolerance for isopropanol and isobutanol were increased to 30 g/L and 10 g/L, respectively. The tolerant yeast mutants tolerated 60 g/L isopropanol and 15 g/L isobutanol while the parent strain did not grow in these high concentrations of alcohol.

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Towards Integration of "Cello", the Computer-Aided Design Platform for Genetic Circuits, into KBase.

[2] Omree Gal-Oz (ogaloz@lbl.gov), [2,4] Mika Tei, [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; [4] Japan Agency for Marine-Earth Science and Technology

http://kbase.us/ , **https://narrative.kbase.us/#appcatalog/module//cello** , **http://cidarlab.org/cello/**

Project Goals: Cello is a suite of tools for biological circuit design and functional prediction created by the Voigt and Densmore Labs. Our objective is to allow biologists to easily design and analyze custom genetic circuits using this framework while also leveraging the interlinked resources within the DOE Systems Biology Knowledgebase to support advanced information and modeling capabilities for gene and organismal function. KBase will also ultimately enable sharing of the design, build, test and learn processes that are transferable and generalizable across applications. We seek to increase the efficacy of biological design procedures to reduce the number of experimental implementation iterations necessary for product optimization.

Advances in DNA reading and writing technologies are driving the adoption of new paradigms for engineering of biology. One of the current challenges is to be able to decide which genetic parts must be rewired to achieve higher production of desired bioproducts. To do this, there must be standards for rapid design and testing of custom gene regulatory systems. Therefore, this project aims to develop a smart platform in which engineers can synthesize and test their circuits *in silico* within one virtual location and test the circuits in an environment that allows leverage of diverse information and models of diverse biological function in variable environments through multiple programs. In addition, engineers will avoid having to deal with the laborious process of managing the configuration of the program on their personal computers. Toward this goal, we used the KBase Software Development Kit to add the Cello biocircuit design software into the system and established the framework for design and analysis of synthetic genetic circuits.

Cello is a computer-aided-design program created by CIDAR labs (Boston University), the Synthetic Biology Center at MIT, and the Biosystems and Biomaterials Division at the NIST. The program applies the concept of Circuit-Design Automation from computer circuit design to genetic parts in living organisms. It uses the modularity of transcription factors such as repressors and promoters to control the gene regulatory system. One of the inputs to the program is a set of instructions which lists relationships between certain promoters and their related outputs (e.g. the presence of compound A implies Gene B is being expressed, the presence of compound B means Gene C is not being expressed); the other inputs are the

promoters and outputs themselves. This information is used to generate the instructions for a plasmid which would realize these relationships and predict its operation in the target organism.

Now that Cello is packaged in KBase, users can design genetic circuits that are predicted to express target pathways under different environmental conditions, and produce and visualize a plasmid design to implement this circuit. While still under development and being optimized for integration with the rest of KBase capabilities, this brings a powerful design tool into a powerful environment for everyone to use and build upon.

Ultimately, this research will enable predictable and dynamic control of multiple economically important traits such as growth on plant-derived feedstocks, tolerance to toxic byproducts and other stresses, and production of multiple and complex target molecules. By deploying these new biological engineering platforms for non-model microorganisms with potential industrial relevance, this project will advance the DOE's mission to develop sustainable biofuel production from renewable sources.

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Controls on the Composition of Microbial Derived Necromass in Soil

Kirsten S Hofmockel[1]* (kirsten.hofmockel@pnnl.gov), Sheryl Bell[1], Chris Kasanke[1]

[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) characterizing microbial necromass, (2) identifying the microbial pathways that contribute to necromass accumulation under controlled conditions, (3) characterizing microbial necromass accumulation in soil incubation experiments and (4) comparing long-term, cross-site microbial responses to cropping system inputs and edaphic factors.

Abstract text. Crop selection and soil texture influence the physicochemical attributes of the soil, which structure microbial communities and influence soil organic matter formation, cycling and long-term storage. At the molecular scale, microbial metabolites and necromass alter the soil environment, which creates feedbacks that influence ecosystem functions, including soil organic matter accumulation. Yet the generalizable mechanisms regulating the accrual and long-term stabilization of soil organic matter are still unclear. Using a long-term soil incubation and ^{13}C-labeling, we are testing the hypotheses that (1) microbial derived necromass is a significant component of soil organic matter and (2) soil texture and mineralogy significantly influence microbial derived organic matter composition and accumulation. By integrating lab to field studies, we aim to identify the molecules, organisms and metabolic pathways that control the formation of compounds that contribute to long-term organic matter stabilization in bioenergy soils.

As plant-derived inputs undergo microbial decomposition, some of the resulting organic residues are incorporated into microbial biomass, and a significant proportion of soil organic matter is attributed to the resulting microbial derived residues, or necromass. This includes biomass residues (lipids, proteins, amino sugars) and microbial exudates (enzymes, exopolysaccharides, lipids, glycoproteins). Yet little empirical evidence is available to support this conceptual model and inform management decisions that aim to amplify biological processes that enhance soil organic matter formation and persistence. To address this knowledge gap, we are conducting a long-term lab incubation and characterizing microbial residues using mirco-IRMS, GC-MS and ssNMR. Soil samples were collected from switchgrass (*Panicum virgatum*) plots from the DOE the Great Lakes Bioenergy Research Center (GLBRC) Intensive Biofuel Cropping System Experiments, including sandy loams of the Kellogg Biological Station (KBS) in MI, and silty loams from the Arlington Agricultural Research Station (AARS) in WI. These soils were amended with ^{13}C-labeled glucose, which was rapidly incorporated into microbial biomass. The sandy soils had a microbial community that was more active over the course of the two-month incubation, where 18.9 % of the added carbon (C) was respired in the first 24 hours for sandy

soils (MI) and 17.2% for silty (WI) soils. By tracking the remaining ^{13}C over time, we are determining the persistence and form of microbial derived molecules. To demonstrate the transformation of microbial necromass into soil organic pools, we waited for the initial ^{13}C labeled microbial community to turnover (2 months lab incubation) before sampling. We measured C pools and ^{13}C incorporation into microbially derived metabolite, protein and lipid pools as well as residuals that were unextractable and strongly associated with the soil. Despite respiring roughly 20% of the added glucose as CO_2 in the first 24 hours, approximately half of the added ^{13}C remained in the soil after 2 months. After extraction ~1/3 of the total soil C remained in the soil and contained ~20% of the soil ^{13}C. Although the metabolite, lipid and protein pools contained <0.2% of the samples on a mass basis, they contained 20-25% of the remaining label. Relative to proteins and lipids, the metabolites were the most highly enriched (>5000δ‰ $\delta^{13}C$), despite containing the lowest C concentration. Together these data reveal how C assimilated by microbes is transformed into pools of soil organic C. By using a cross-site approach this study demonstrates the influence of soil texture on the persistence of necromass pools. Incubations are on-going, and the final harvest will occur after the turnover of multiple generations of the soil microbiome to reveal the form and fate of microbial necromass in soil.

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.

Deconstructing the Soil Microbiome into Reduced-Complexity Functional Modules

Dan Naylor[1], (Dan.Naylor@pnnl.gov), Sarah Fansler[1], Colin Brislawn, William Nelson[1], Ryan McClure,[1], **Kirsten S. Hofmockel[1] and Janet K. Jansson[1]***

[1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

Project Goals:

PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We focus on a multi-scale examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments are designed to confront spatial challenges and inter-kingdom interactions that regulate networks of biochemical reactions. The exchange among bacteria, fungi, viruses and plants are being characterized in the context of microbial metabolism and community function. These experimental data have been 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. Our cross-scale experiments are coordinated together to investigate the influence of moisture on the interkingdom-interactions. 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: The soil microbiome is an invaluable component of ecosystem function. Improving our understanding of the soil microbiome will aid predictions of how microbially-mediated processes are affected by external perturbations. However, the soil microbiome's extensive taxonomic and functional diversity hinders comprehensive analysis of this system. One option for studying the soil microbiome in detail is to deconstruct it into discrete functional units for individual analysis. Here, we aimed to dissect the complex soil microbiome through targeted enrichments based on metabolic capacity, thereby obtaining reduced-complexity consortia we term 'functional modules'. We hypothesized that, through targeted enrichments of a starting soil inoculum, we can obtain functional module communities that are low-diversity, reproducible and predictable, and encapsulate a significant extent of soil phylogenetic diversity while also enriching for underrepresented soil taxa. Furthermore, we hypothesized that functional module communities are also distinct from one another with respect to gene expression patterns.

To generate functional module communities, we cultured a soil inoculum in minimal media with specific modifications for the module in question. Module categories included addition of simple carbon substrates (e.g. sugars, organic acids) or complex plant polysaccharides, supplementation with antibiotics, anaerobic modules with alternative redox acceptors, or imposition of alternative growing conditions reflective of common field stresses. In total, 324 communities were obtained across 66 distinct functional modules. Analyzing community composition via 16s rRNA amplicon sequencing revealed that all modules were significantly reduced in diversity and richness relative to control soil communities, with polysaccharide modules significantly more diverse than all remaining categories. With respect to reproducibility, anaerobic modules were the least predictable (highest variability between module replicates for community composition), followed by polysaccharide and stress modules.

Our approach isolated not only known soil microbiomes but also several that are not found in amplicon analysis of soil. We found that of the 241 unique taxa in the soil core microbiome, 90 were found in at least one functional module core, collectively encapsulating approximately 37.3% of soil phylogenetic diversity. While most major soil phyla were represented in module cores, there were several that were underrepresented (Verrucomicrobia) or not found (Planctomycetes) and likely require alternative strategies for enrichment. In addition, we were also able to obtain 481 taxa that were uniquely found in module cores, showing there is a significant extent of soil diversity that becomes measurable through our strategy. We also investigated functional trends across a subset of modules using a metatranscriptomics approach, to confirm that modules are significantly different at the functional as well as taxonomic level. Functional patterns varied by module: for example, pectin and xylan modules were elevated for transcripts involved in 'Glycan biosynthesis and metabolism' relative to the other three modules. Comparing patterns of enriched transcripts showed that each of these five modules has its own distinct pattern on the collective microbial metabolic map. These results highlight the potential for combining separate modules into functional patterns and ultimately reconstructing the full biochemical capacity of the soil microbiome.

The functional module approach used here has significant applications for microbiome analysis: module communities are more tractable for omics analysis while retaining interactions of the native parent community. When studying a particular metabolic niche or biochemical process, a reduced-complexity consortium of microbes implicated in this process can be obtained through this strategy. Furthermore, through investigation of functional trends, one can piece together the full functional potential of a microbiome system with higher detail than would be achievable through holistic analyses. Establishing this methodology will not only be beneficial for improving our understanding of the soil microbiome but could conceivably be applied to other complex microbiomes as well.

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.

Fungal hyphal networks play a key role in soil microbiome micronutrient acquisition and transport during drought

Christopher R. Anderton*[1] (christopher.anderton@pnnl.gov), Arunima Bhattacharjee[1], Jocelyn Richardson[2], Odeta Qafoku[1], Dušan Veličković[1], Kaitlyn Schwarz[1], Zihua Zhu[1], Mark Engelhard[1], Mark Bowden[1], Sheryl L. Bell[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

Project Goals: PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We focus on a multi-scale examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments are designed to confront spatial challenges and inter-kingdom interactions that regulate networks of biochemical reactions. The exchange among bacteria, fungi, viruses and plants are being characterized in the context of microbial metabolism and community function. These experimental data have been 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. Our cross-scale experiments are coordinated together to investigate the influence of moisture on the interkingdom-interactions. 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: Increasing evidence suggest that inter-kingdom interactions are critical to microbial resiliency under water and nutrient stress. Within soil microbiomes, fungal constituents are believed to be among the most successful under water and nutrient limited conditions. This is in part due to the filamentous hyphal networks that most soilborne fungi can develop, which permits them to link and exploit discrete substrate pools. Furthermore, widespread fungal interactions with plants and bacteria make fungi integral to all aspects of C, N, P, and S cycling within soil. Nonetheless, the precise mechanisms of how soil fungal communities biosense and access specific nutrient sources under drought remains vastly understudied. Here, we explored the presence and role of fungal hyphal networks in native soil experiencing drought-like conditions, and then we aimed to gain deeper insight into the processes that govern mycelial bridging of nutrient sources within the soil microenvironment.

We built a controlled soil environment (a so-called 'SoilBox')[1], which enables determination of spatial organization and interaction of soil microbial communities using optical and mass spectrometry imaging techniques.[2] Within the SoilBox, we were able to visualize fungal hyphal networks bridging different chitin islands over distances of 27 mm in native soil under limited soil moisture conditions.

Moreover, optical and molecular imaging results showed the rate of chitin island decomposition by the native soil microbial community under different moisture regimes, where degradation of the islands was not significantly altered under dry soil conditions in comparison to soil moisture at field capacity under the timeframe we measured. We then inoculated *Fusarium chlamydosporum* into micromodels that concurrently simulate the porosity and mineralogy of soil. We observed increased hyphal density and fungal thigmotropism around obstacles and through soil-like pore spaces within mineral-doped soil micromodels in comparison to micromodels without minerals. Secondary ion mass spectrometry analysis showed K^+ and Na^+ enrichment and translocation in fungal hyphae grown in mineral-doped channels. The translocation from minerals by fungal hyphae resulted in K^+ speciation, as was observed using X-ray near edge absorption analysis. In comparison, fungal mycelia grown in micromodels without minerals did not exhibit thigmotropic behavior, micronutrient cation translocation, or K^+ speciation. These results provide the first direct evidence of hyphal translocation of micronutrients from a mineral surface under nutrient limiting conditions. Taken together, these studies demonstrate that mycelial acquisition and transport of mineral-derived inorganic nutrients provides fungal communities a survival advantage, where they can access scarce and discrete organic nutrient pools within soil microenvironments and act as nutrient highways during drought.

References

1. A. Bhattacharjee, D. Velickovic, T.W. Wietsma, S.L. Bell, J.K. Jansson, K.S. Hofmockel, C.R. Anderton, "Visualizing microbial community dynamics via a controllable soil environment," *mSystems* **2020**, *in revision*
2. D. Veličković, R. K. Chu, A. A. Carrell, M. Thomas, L. Paša-Tolić, D. J. Weston, C. R. Anderton, "Multimodal MSI in conjunction with broad coverage spatially resolved MS^2 increases confidence in both molecular identification and localization," *Anal. Chem.* **2018**, *90*, 702

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.*

Generation and Analysis of Reduced Complexity Model Soil Consortia

Ryan McClure,[1]* (Ryan.McClure@pnnl.gov), Dan Naylor[1], Yuliya Farris[1], Sarah Fansler[1], Michelle Davison[1], **Kirsten S. Hofmockel[1] and Janet K. Jansson[1]**

[1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

Project Goals:

PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We focus on a multi-scale examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments are designed to confront spatial challenges and inter-kingdom interactions that regulate networks of biochemical reactions. The exchange among bacteria, fungi, viruses and plants are being characterized in the context of microbial metabolism and community function. These experimental data have been 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. Our cross-scale experiments are coordinated together to investigate the influence of moisture on the interkingdom-interactions. 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 microbial communities are critical to the overall carbon cycle and to the decomposition of complex biopolymers such as chitin and cellulose. Despite the critical nature of these microbiomes, a detailed understanding of how the interactions between members lead to emergence of community functions is lacking. This is due, in part, to the complex nature of the soil microbiome with thousands of species across several kingdoms contributing to the overall response of soil. In order to gain a more detailed view of the soil environment, we took an approach based on developing and analyzing reduced complexity microbial consortia that contain fewer species than the native soil but are still representative of this site and are more experimentally tractable. We hypothesize that analysis of these model consortia can identify reaction modules (sets of interacting species or enriched metabolic functions), and other fundamental aspects of soil microbiome interactions. These microbial relationships can serve as functional predictions for future tests to be carried out at our native field site, greatly expanding our knowledge of the functioning of soil microbiomes. We expect that such consortia will serve as valuable tools for the research community at large to identify metaphenomes and microbial traits.

To generate reduced complexity soil microbial consortia, we collected samples from our native field site containing a grassland silt loam soil. This native soil was then diluted to various levels, ranging from 10^{-2} to 10^{-3} and cultured on agar plates containing soil extract (water soluble soil nutrients) as well as chitin (100 ppm). Cultivation was carried out for several months with replating performed once per week to maintain the community. We found that species richness rapidly fell when soil microbial communities were cultured on plates, reaching levels of between 20-70 OTUs in only 2-3 weeks. In addition, several of the communities examined via 16s amplicon analysis still represented a fairly diverse community with

several different phyla, orders and genera represented. Very few of the communities were completed dominated by only a few bacterial or fungal species.

We also found that consortia could be stored as glycerol stocks or lyophilized stocks and reconstituted with a high degree of reproducibility. Most of the reconstituted consortia tested were very similar to the parent consortia on plates and there was consistent similarity across replicates of reconstituted stocks showing that regrowth from a stock on a plate is a deterministic process. We were also able to isolate many of the constituent species from these consortia and carry out pair wise incubations to begin to explore their interspecies interactions. These data, combined with co-abundance networks of these consortia during growth in soil, show that Rhodococcus, Streptomyces and Rhizobiales species found in our consortia are likely critical to the functioning of this community. These species occupy high central positions in our species co-abundance network and co-cultivation experiments show that, especially in the case of Rhodococcus, there are likely beneficial interactions with other species of these consortia.

This series of representative and reduced complexity are powerful tools that can be used by the soil community at large to interrogate the response of soil microbiomes to a number of perturbations and to confirm critical interactions between microbial species, particularly inter-kingdom interactions that characterize the emergent behavior of soil microbiomes. Constructed reduced complexity consortia provide a means to more powerfully leverage high-throughput, multi-omic techniques to better characterize these interactions and the major constituent players that are a part of them. Further knowledge of these interactions will help us better understand the overall metaphenome of soil systems, especially as they respond to critical perturbations including drought.

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,[1] Sneha Couvillion,[1] Kent J. Bloodsworth,[1] Jennifer E. Kyle,[1] Vanessa Paurus,[1] **Janet Jansson**[1]**, Kirsten Hofmockel**[1]

Project Goals:

PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We focus on a multi-scale examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments are designed to confront spatial challenges and inter-kingdom interactions that regulate networks of biochemical reactions. The exchange among bacteria, fungi, viruses and plants are being characterized in the context of microbial metabolism and community function. These experimental data have been 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. Our cross-scale experiments are coordinated together to investigate the influence of moisture on the interkingdom-interactions. 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 from drought is increasing in frequency with unknown outcomes for soil microbiomes. The mechanisms in which the microbial communities respond to dehydration is very important, particularly in arid and marginal lands. As soils dry, the water potential across the cell wall decreases leading to osmotic stress which is compounded by limited diffusion. To address environmental changes caused by differences in moisture level, microbes must shift physiology and metabolic interactions to survive. One mechanism microbes employ to offset decreases in moisture levels is the production of osmolytes to reduce their internal water potential and maintain fluid balance. Mostly, microbes use simple organic molecules with good solubility such as amino compounds (amino acids and glycine betaine) and simple sugars (trehalose). However, the synthesis and storage of these compounds is energetically costly to the microbe and little is understood about the molecular response to the changes in soil moisture, and how that affects the phenotype.

Here, we aimed to understand how the physiology, metabolism, and interactions of soil microbes change in response to moisture, and to use this understanding as a basis for predicting the soil metaphenome. While microbes respond to drought by different mechanisms, a common phenomenon is the rapid mineralization of soil organic matter and increased rate of CO_2 release upon rewetting dry soil, termed the 'Birch Effect'. We are testing the hypothesis that during desiccation in an arid, marginal soil, microbes will initially accumulate osmolytes only to rapidly metabolize them upon rewetting. One mechanism for the disposal of the osmolytes is the rapid release of CO_2, DOC and nutrients, arising from intracellular material and varies in magnitude based on the biomass. Most analyses measure respiration on the hour to daytime frame, thus missing the immediate rewetting events. To provide information about the initial microbial response to wetting, we developed and an atmospheric monitoring instrument (RTMS) that measures the levels of CO_2, O_2, N_2, H_2O and other relevant gases, simultaneously in real time. In these

experiments, desiccated soil was rewet, and gas levels were measured in seconds, minutes and hours after rewetting.

After rewetting desiccated soil, we observed a rapid production of CO_2 in the first 90 minutes after which the rate stabilized. In order to determine the origin of the carbon, we amended the water used in the rewetting step with glucose to act as a surrogate for extracellular carbon. In these experiments, we observed a biphasic CO_2 response; a rapid release of CO_2 that leveled off at 90 minutes, and a larger production of CO_2 over the next 4 hours. We hypothesized that this initial burst was due to the metabolism of intracellular compounds and the secondary burst was due to the metabolism of extracellular compounds. To test this hypothesis, we used ^{13}C labeled glucose in the experiment to differentiate between $^{12}CO_2$ and $^{13}CO_2$ production. The $^{12}CO_2$ portion of the glucose addition resulted in a respiration response that was identical in rate and amplitude to that observed in a water-only control. The $^{13}CO_2$ production was identical to 4-hour response, leading to the supposition that internal metabolites are respired before external ones after rewetting. In the same experiment, we were also able to detect O_2, N_2 and H_2O over the same time scale. By taking measurements of all the gasses in real time, we were able to deconvolve the order and rate in which the microbial communities were producing or consuming the different gasses in the first 90 seconds after the addition of water.

Evaluation of the soil metabolomes at specified time points within 3 hours after wetting were also examined. GC-MS based analysis revealed 119 metabolites, including trehalose, glycine and other sugars, and amino acids. Of these metabolites, 58 were confidently identified, and 61 spectral features represent opportunities to discover new metabolite identifications. GC-MS based lipidomic analysis revealed changes in fatty acid profiles and related molecules in response to rewetting dry soil. LC-MS/MS based lipidomic analysis revealed significant changes across multiple lipid classes including glycerophospholipids, glycerolipids and sphingolipids. For example, triglycerides, which can potentially function as storage molecules and regulate cellular fluidity, were found to increase in rewet soil and also showed a higher degree of saturation. Together our multi-omic analysis is contributing towards a deeper understanding of the soil microbiome metabolic and metaphenomic responses to drought.

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.

Viral diversity: decoding hidden potential for metabolic functions in soils

Ruonan Wu[1], Michelle R. Davison[1], Emily B. Graham[1*] (emily.graham@pnnl.gov), William C. Nelson[1], Sarah J. Fansler[1], Sheryl L. Bell[1], Iobani Godinez[1], David Paez-Espino[2], Russell Y. Neches[2], Nikos C. Kyrpides[2], Jason E. Mcdermott[1], **Kirsten S. Hofmockel[1]**, and **Janet K. Jansson[1]**

[1]Earth and Biological Sciences Directorate Pacific Northwest National Laboratory, Richland, WA
[2]Joint Genome Institute, Walnut Creek, CA

Project goals:
PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We focus on a multi-scale examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments are designed to confront spatial challenges and inter-kingdom interactions that regulate networks of biochemical reactions. The exchange among bacteria, fungi, viruses and plants are being characterized in the context of microbial metabolism and community function. These experimental data have been 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. Our cross-scale experiments are coordinated together to investigate the influence of moisture on the interkingdom-interactions. 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.

Although viruses are the most abundant biological entities on the planet, they have been challenging to study in soil systems. Based largely on work in other systems, soil viruses are hypothesized to be vital to ecosystem functions through transferring auxiliary metabolic genes, lysing microbial cells, and inducing microbial taxonomic turnover. Auxiliary metabolic genes (AMGs) in marine viruses, for instance, encode numerous metabolic functions involved in carbon, nitrogen, and phosphorous cycling. Yet, we have limited understanding of the identity, distribution, and ecological function of soil viruses due in part to the vast range of soil ecosystems, and until recently the lack of appropriate molecular screening tools.

Due to the lack of a universal marker gene, such as 16S rRNA that is used in prokaryote phylogeny, soil viruses have mainly been identified by screening for viral sequences from metagenomes. Recent advances in high-throughput sequencing and computational approaches make it feasible to uncover virome from the soil microbiome with high complexity. Current bioinformatic tools each have limitations on their abilities to identify viral sequences from environmental metagenomes, and viral reference databases are insufficient for soil viromes. We therefore developed the 'VirFunnel' workflow to increase confidence in soil viral assignments by leveraging multiple computational tools and viral reference databases, thereby mitigating the weaknesses of each approach with another publicly available tool. The VirFunnel workflow includes the VirSorter and VirFinder tools as well as self-curated and public-available viral

databases (JGI IMG_VR and NCBI RefseqVirus). It contains four distinct modules to (1) provide increased confidence in assignment of viral sequences from soil metagenomes (*viral mining module*), (2) assess viral diversity with a standardized and step-wise workflow (*viral clustering module*), (3) determine host composition (*host assignment module*), and (4) extract potential auxiliary metabolic genes (*AMG classification module*).

We challenged our resulting computational workflow to uncover viruses from three highly complex, deeply sequenced soil metagenomes (>1 Tbp each) that were obtained from native grassland soils in Washington, Kansas and Iowa. These soils span a range of physicochemical conditions and harbor diverse microbiomes. To date, there have been few studies of soil viruses using a metagenome screening approach, and ours represents the first for grassland soils containing incredibly high microbial diversity and complexity. By application of VirFunnel, we were able to detect 2,631 soil viral sequences representing 28% and 230% more putative viruses than solely VirSorter or Virfinder, respectively. These include many previously undescribed soil viruses.

We were particularly interested in possible ecological roles for viral clusters found in all three grassland soils. Commonly detected viral clusters contained significantly more viral sequences with more average CRISPR hits than site-specific clusters, indicating more frequent virus-host interactions. Some of these viral clusters were linked to multiple hosts across different microbial phyla, suggesting a potential for viral generalists with a wide host range in soils. Additionally, we identified potential viral AMGs previously undescribed in soils and thought to be involved in mannose degradation pathways in all soils. Auxiliary metabolic genes involved in xylan degradation were detected in WA and Kansas soils. Since xylan and mannan/mannose are common plant cell well components, multiple related AMGs detected across our three disparate grassland soils supports recent suggestions that viruses have the ability to cycle plant-derived carbon compounds in soil. Finally, we located many AMGs involved in complex carbohydrate metabolism, energy acquisition and fatty acid biosynthesis that collectively point to a central role for viruses in soil carbon cycles more broadly.

To expand our understanding of the soil virome to a global scale, we also collaborated with the JGI to collect over 3200 samples with metagenomic shotgun sequences for use in the most extensive characterization of soil viruses to date. The 3200 samples are from numerous sources including IMG, MG-RAST, EMP, NEON, and other collaborators. We plan use this dataset to define the global soil virome in terms of its size, diversity, and function. We will highlight biogeographical and eco-evolutionary patterns of viral sequences and their potential hosts, as well as identify sequences attributed to giant viruses, virophages, and AMGs. We predict that viruses are involved in multi-domain interactions in soils and play central roles in soil metaphenomes by serving as a genetic reservoir of metabolic genes.

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.*

Effects of Warming on Bacterial Growth and Element Fluxes in Soil

Bruce Hungate[1,2] (Bruce.Hungate@nau.edu), Alicia Purcell[1,2], Benjamin Koch[1,2], Bram Stone[1], Chao Wang[3], Egbert Schwartz,[1,2], Ember Morrissey[3], Jeff Propster[1,2], Jennifer Pett-Ridge[5] Kathryn Nantz[1], Kirsten Hofmockel[4], Megan Foley[1,2], Michaela Hayer[1], Michelle Mack[1,2], Paul Dijkstra[1,2], Samantha Miller[1], Steve Blazewicz[5]

[1]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, 86011, USA; [2]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, 86011, USA; [3]Department of Biology, West Virginia University, Morgantown, WV, 26506, USA; [4]Environmental and Molecular Sciences Division, Earth Scientist Pacific Northwest National Laboratory, Richland, WA 99352; [5]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, 94550, USA

Project Goals: This project aims develop and apply new methods to understand the ecology of soil microorganisms using stable isotope tracers and genomics. This new suite of techniques will investigate and describe the microbial ecology of nutrient cycling in soil environments as microorganisms grow and die. The work will focus on particular soil microorganisms, bacteria and fungi, that make up the majority of life in soil, and which are responsible for most of the nutrient transformations in soil that are vital to ecosystems, and to people. This project will also evaluate how soil microorganisms and the nutrient cycling processes they catalyze are sensitive to shifts in temperature, a major driver of biological processes.

Abstract

We are combining isotopes and genomics to understand the ecology of soil microorganisms, focusing on their responses to warming across biomes. The work relies on four long-term field experiments where temperature-treated and control plots occur in arctic, boreal, temperate, and tropical biomes. Our first step in evaluating nutrient uptake is to measure growth rates of microorganisms, *in situ*, because growth is an excellent integrator of resource assimilation. We conducted experiments to measure rates of growth of soil microorganisms in response to a large temperature gradient in the laboratory (5 - 45 degrees C) and in the field (using in situ warming experiments), and have found that soil bacterial growth rates are strongly sensitive to temperature in both contexts. We also find considerable taxonomic variation in the temperature responses of soil bacteria. Phylogenetic signals of these responses, while often significant, in general do not support phylum-level generalizations about bacterial traits. Rather, when signals are statistically significant, they appear to reflect finer-scale phylogenetic organization. In the field, warming in temperate and arctic ecosystems caused changes in bacterial growth that depended on the duration of exposure to the warming treatments, and responses were concomitant not only with changes in temperature but also with shifts in the plant community and with effects on soil processes. In general, we find that bacterial growth rates — and their responses to warming — scale to measured rates of biogeochemical fluxes, making it possible to estimate taxon-specific rates. In general, this work aims to ascribe element fluxes with taxonomic resolution, and to test the sensitivities of these

processes to temperature. The approach used here interrogates community and taxon-specific microbial controls over key biogeochemical processes in terrestrial environments, and test quantitative ecological and biogeochemical principles using genomics and isotope data, including theories of element limitation, growth efficiency, and nutrient use efficiency.

Funding statement. This work was supported by the US Department of Energy, Program in Genomic Sciences, Award Number: DE-SC0020172. Work at LLNL was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. A portion of the work was contributed by PNNL, a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830

Measurement of Isotope Assimilation Rates into Microbial DNA Through Quantitative Stable Isotope Probing with Internal Standards.

Egbert Schwartz,[1,2*] (Egbert.Schwartz@nau.edu), Michaela Hayer[1], Kathryn Nantz[1], Ember Morrissey[3], Steve Blazewicz[4], Jennifer Pett-Ridge[4], Rex Malmstrom[5], and **Bruce Hungate**[1,2]

[1]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, 86011, USA; [2]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, 86011, USA; [3]Department of Biology, West Virginia University, Morgantown, WV, 26506, USA; [4]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, 94550, USA; [5]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Project Goals: This project aims to develop and apply new methods to understand the ecology of soil microorganisms using stable isotope tracers and genomics. This new suite of techniques will investigate and describe the microbial ecology of nutrient cycling in soil environments as microorganisms grow and die. The work will focus on particular soil microorganisms, bacteria and fungi, that make up the majority of life in soil, and which are responsible for most of the nutrient transformations in soil that are vital to ecosystems, and to people. This project will also evaluate how soil microorganisms and the nutrient cycling processes they catalyze are sensitive to shifts in temperature, a major driver of biological processes.

In DNA-based quantitative Stable Isotope Probing (qSIP) a tracer enriched in heavy stable isotopes is added to a microbial community. Microbes that assimilate the tracer and replicate their genome incorporate the heavy isotopes into their DNA which can be separated from DNA with natural abundance levels of isotopes along a cesium chloride gradient generated in an ultracentrifuge. By analyzing the DNA in each of the 20-30 fractions taken from the gradient it is feasible to calculate the atom percent enrichment of each bacterial taxon's DNA. However, it has proven challenging to compare qSIP results among different experiments and laboratories. This may be because different types of tracers are used or because the equipment used for qSIP is not the same in all labs. We propose that using internal standards in qSIP will improve comparisons, decrease variance within a qSIP experiment and the amount of labor required for qSIP analysis. We constructed 2 different plasmids, approximately 9 kb in size, which contain bacterial 16S V4 primer sites commonly used to characterize a fragment of the bacterial 16S rRNA gene via Illumina based amplicon sequencing. The sequence in between the primer sites was taken upstream from the E. coli 16S rRNA gene and is not highly homologous to known 16S rRNA gene sequences, so that it is possible to distinguish between the internal standards and the 16S rRNA gene sequences derived from the microbial community. We used one of the plasmids as a template in PCR with nucleotides that contained natural abundance levels of ^{13}C and ^{15}N while the other plasmid served as a template in PCR with nucleotides that contained 98 atom% ^{13}C and ^{15}N. As a result the PCR product from one plasmid was approximately 26.5 neutrons heavier than the other PCR product. This resulted in a 0.05 g/mL difference in density as measured through isopycnic centrifugation on a cesium chloride gradient. We will present a strategy for

standardizing the difference in density between the two standards so that DNA-qSIP results from different laboratories and experiments can be compared.

This work was supported by the US Department of Energy, Program in Genomic Sciences, Award Numbers: DE-SC0016207 and DE-SC0020172. Work at LLNL was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Exascale Networks for *Arabidopsis* in KBase

Michael Garvin[1], David Kainer[1], Jared Streich[1], Angelica M. Walker[1], **Daniel A. Jacobson**[1]* (jacobsonda@ornl.gov)

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

Project Goals: To create large interaction networks using SNPs, transcriptomic, metabolomic, microbiome taxa, climatypes and other types of phenotype data as well as tissue-specific epigenetic and expression data. The creation of such network-based models involves computing ensembles of custom correlation metrics, mixed linear models and explainable-AI methods at extreme scales on the Oak Ridge Leadership Computing Facility supercomputer Summit. The results of these workflows are modeled as networks (and hyper-networks) in order to provide an integrated systems biology view of an organism. This includes using a number of previously developed analysis & modeling methods that, in combination with explainable-AI approaches, predict (high combinatorial order) epistatic architectures for all available traits. This is being developed as a community resource in KBase for the model plant *Arabidopsis thaliana*.

The rapid increase in biological assays, high-throughput phenotyping studies, and computational prediction capabilities has resulted in an enormous wealth of biological data for many model species. These data layers (e.g., genomic, transcriptomic, metabolomic, protein-protein interactions, climatype, phenomic) are developed with the goal of understanding the operation of overarching biological systems and discovering the basis for emergent phenotypes. Each data layer is often interpreted within the context of that specific dataset, which provides useful, but limited, insights. This is because biological elements rarely operate in isolation within and between the cellular environment; data from a single layer reveals only part of the story, and can possibly be misleading. As the primary model plant species, there is a great deal of publicly available data derived from *Arabidopsis thaliana*. We are creating data products composed of *A. thaliana* data layers that will be integrated into KBase. Furthermore, we will develop novel algorithms that can be ported into KBase in order to provide a flexible, open-source product for plant biologists interested in *A. thaliana*. We will also develop a systems-biology resource enabling KBase users to rank candidate genes and to predict the function of unknown genes. While each layer of data intrinsically has a level of ability to predict gene function based on the connectivity and topology of the nodes and edges of its network, some layers are more informative than others. Various methods of analysis, including explainable machine learning, can be applied to evaluate the predictive ability of each layer by cross-validation using known genes from well characterized pathways or interaction networks. This resource will also be integrated into the KBase knowledge engine, providing a powerful new tool to plant biologists.

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 ERKP972.

Peta- and Exa-scale for Arabidopsis in KBase

Michael R. Garvin[1]* (garvinmr@ornl.gov), David Kainer, Jared Streich, Angelica M. Walker, Daniel A. Jacobson

[1]Oak Ridge National Laboratory, Oak Ridge, TN

Project Goals: To create large interaction networks using SNPs, transcriptomic, metabolomic, microbiome taxa, climatypes and other types of phenotype data as well as tissue-specific epigenetic and expression data. The creation of such network-based models involves the running of ensembles of custom correlation metrics, mixed linear models and explainable-AI methods at extreme scales on the Oak Ridge Leadership Computing Facility supercomputer Summit. The results of these workflows are modeled as networks (and hyper-networks) in order to provide an integrated systems biology view of an organism. This includes using a number of previously developed analysis & modeling methods that, in combination with explainable-AI approaches, predict (high combinatorial order) epistatic architectures for all available traits. This is being developed as a community resource in KBase for the model plant *Arabidopsis thaliana*.

The rapid increase in biological assays, high-throughput phenotyping studies, and computational prediction capabilities has resulted in an enormous wealth of biological data for many model species. These data layers (e.g., genomic, transcriptomic, metabolomic, protein-protein interactions, climatype, phenomic) are developed with the goal of understanding the operation of overarching biological systems and discovering the basis for emergent phenotypes. Each data layer is often interpreted within the context of that specific dataset, which provides useful, but limited, insights. This is because biological elements rarely operate in isolation within and between the cellular environment; data from a single layer reveals only part of the story, and can possibly be misleading. As the primary model plant species, there is a great deal of publicly available data derived from *Arabidopsis thaliana*. We are creating data products composed of *A. thaliana* data layers that will be integrated into KBase. Furthermore, we will develop novel algorithms that can be ported into KBase in order to provide a flexible, open-source product for plant biologists interested in *A. thaliana*. We will also develop a systems-biology resource enabling KBase users to rank candidate genes and to predict the function of unknown genes. While each layer of data intrinsically has a level of ability to predict gene function based on the connectivity and topology of the nodes and edges of its network, some layers are more informative than others. Various methods of analysis, including explainable machine learning, can be applied to evaluate the predictive ability of each layer by cross-validation using known genes from well characterized pathways or interaction networks. This resource will also be integrated into the KBase knowledge engine, providing a powerful new tool to plant biologists.

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 ERKP972.

Conserved Genetic Mechanisms for Biotic Stress in Sorghum

Tiffany Jamann[1*](tjamann@illinois.edu), Bill Rooney, [2] and Santiago Mideros,[1]

[1]University of Illinois, Urbana; [2]Texas A&M University, College Station

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

Developing durable disease resistance for biofuel crops is crucial, particularly as the range of biofuel crop production expands and pathogens of other plant species evolve to cause diseases of bioenergy feedstocks. *Setosphaeria* species are significant fungal pathogens of the Andropogonae, and *S. turcica* can infect both maize and sorghum, making it a strong candidate to be an increasing problem for biofuel sorghum. Our overall objective is to gain a systems-level understanding of the pathosystem by leveraging natural genetic variation, host specificity of the pathogen, and transcriptome analysis to improve biotic stress resistance in sorghum. The proactive strategy of paired identification of fungal effectors and plant resistance genes in a pathosystem with a high likelihood of producing a host jump is a paradigm shift in disease management through host resistance.

This work is funded by DOE award number DE-SC0019189 (Plant Feedstocks Genomics).

Cell Wall *O*-Acetyl and Methyl Esterification Patterns of Leaves Reflected in Atmospheric Emission Signatures of Acetic Acid and Methanol

Rebecca A. Dewhirst[*1](radewhirst@lbl.gov), Cassandra A. Afseth[1], Cristina Castanha[1], Jenny C. Mortimer[2,3], and **Kolby J. Jardine[1]**

[1]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [2]Joint BioEnergy Institute, Emeryville, CA, USA; [3]Environmental Genomics and Systems Biology, Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Project Goals: The main goal of the poplar esterified cell wall transformations and metabolic integration study (**PECTIN**) is to quantify *O*-acetyl and methyl ester cell wall metabolism in poplar trees and its interactions with central metabolism and volatile metabolite exchange with the atmosphere. We aim to 1) **Characterize** leaf and stem cell wall esterification reservoirs and sources/sink dynamics during plant development, 2) **Quantify** integration of cell wall esters with central metabolism and the atmosphere during abiotic stress, 3) **Modify** *O*-acetylation and methyl esterification of cell walls through the identification of natural genetic variants, and 4) **Evaluate** impacts of cell wall ester metabolism on plant physiology and stress responses.

Abstract

Polysaccharides are major components of plant cell walls that can be converted into fuels by microbial fermentation, making plant biomass an important bioenergy resource (Mortimer 2019). However, a substantial fraction of plant cell wall polysaccharides are chemically modified with methyl and acetyl groups that impact the yield of microbial fermentation. Although little is known about the biochemical and physiological functions of those cell wall modifications in trees, evidence suggests that they may be highly dynamic and play central roles in the control of cell growth, tissue development, and function (e.g. proper development and function of xylem vessels and leaf stomata), facilitate within and between plant signaling in response to abiotic and biotic stress, and integrate into primary C_1 and C_2 metabolism. While deesterification reactions result in the formation of volatile intermediates (methanol: meOH and acetic acid: AA) these central metabolites are not captured by traditional metabolomics analysis, representing an important gap in our knowledge of cell wall ester metabolism. Plants emit high rates of meOH, generally assumed to derive from pectin demethylation, and this increases during growth and abiotic stress. In contrast, less is known about the emission and source of AA. In this study (Dewhirst *et al*., 2019), we connect leaf volatile emissions of meOH and AA to patterns of plant cell wall *O*-acetyl- and methyl-esters for the first time. We present a new concept of leaf cell wall *O*-acetyl/methyl ester ratios and demonstrate that they are quantitatively reflected in the AA/meOH emission ratios.

Populus trichocarpa leaves in different developmental stages were desiccated and quantified for total meOH and AA emissions together with bulk cell wall acetylation and methylation content. While young leaves showed high emissions of meOH (140 µmol m^{-2}) and AA (42 µmol m^{-2}), emissions were reduced in mature (meOH: 69%, AA: 60%) and old (meOH: 83%, AA: 76%) leaves. In contrast, the ratio of AA/meOH emissions increased with leaf development (young: 35%, mature: 43%, old: 82%), mimicking the pattern of *O*-acetyl/methyl ester ratios of leaf bulk cell walls (young: 35%, mature: 38%, old: 51%), which is driven by an increase in *O*-acetyl and decrease in methyl ester content with age. The results are consistent with meOH and AA emission sources from cell wall de-esterification, with young expanding tissues producing highly methylated pectin that is progressively demethyl-esterified. In the near future, it will be necessary to grow dedicated bioenergy crops as a feedstock for the production of liquid transport fuels and bioproducts. The presented methods will help advance rapid non-invasive phenotype screening and genetic manipulation of the cell wall ester content, with the goal of increasing biofuel yields. Moreover, the methods could be used in future studies at a wide range of spatial and temporal scales to help understand the impacts of cell wall esterification on cell wall structure and function, and numerous physiological and biochemical process including growth, stress responses, signaling, plant hydraulics, and central carbon metabolism.

References

1. Cell wall *O*-acetyl and methyl esterification patterns of leaves reflected in atmospheric emission signatures of acetic acid and methanol (2019) Dewhirst R, Afseth C, Castanha C, Mortimer J, and Jardine K, *Plant Cell and Environment* (in review).
2. Plant synthetic biology could drive a revolution in biofuels and medicine. Mortimer J, *Experimental Biology and Medicine* 244.4 (2019): 323-331.

This research is supported by the Office of Science Early Career Research Program (FY18 DOE National Laboratory Announcement Number: LAB 17-1761), Topic: Plant Systems for the Production of Biofuels and Bioproducts (Program Manager: Pablo Rabinowicz Pablo.Rabinowicz@science.doe.gov)

Accelerating Pathway Engineering of Non-Model Organisms Through Novel Cell-Free to In Vivo Workflows

Fungmin (Eric) Liew,[1] Shivani Garg,[1] Blake Rasor,[2] Aislinn Davis,[1] Alex Juminaga,[1] Ching Leang,[1] Steven D. Brown,[1] Sean D. Simpson,[1] Ashty Karim,[2] Michael Köpke (michael.koepke@lanzatech.com)[1]* and **Michael C. Jewett[2]**

[1]LanzaTech Inc, Skokie, IL; [2]Northwestern University, Evanston, IL

http://www.lanzatech.com; http://jewettlab.northwestern.edu

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).

Climate crisis and rapid population growth are posing some of the most urgent challenges to mankind. The accelerating rate of extraction and combustion of fossil resources for fuel, energy and chemicals over the past 100 years has resulted in carbon dioxide (CO_2) accumulation in the atmosphere to levels unprecedented since the Pliocene Epoch (5.3 - 2.6 million years ago). Although the effect that elevated atmospheric CO_2 will have on the Earth's climate has been predicted by scientists for several decades, it was only in 2016 through the Paris Agreement that nations formally laid plans to abate atmospheric CO_2 release. In each case, these plans necessitate that "above ground" carbon resources increasingly displace fossil resources as feedstocks for fuel and chemical production.

Gas fermentation offers a solution using carbon-fixing chemolithoautotrophic microorganisms. After a decade of scale up, the technology has recently been commercialized by LanzaTech with the first 48k MTA plant turning emissions from the steel industry into fuel ethanol operating successfully and additional units under construction. The process has been demonstrated to accept a broad range of feedstocks including waste gases from various industrial sources (e.g., processing

plants or refineries) or syngas generated from any biomass resource (e.g., agricultural waste, unsorted and non-recyclable municipal solid waste, or organic industrial waste) (1).

In order to maximize the value that can be added to the array of gas resources that the process can use as an input, LanzaTech has pioneered the development of a genetic toolbox for acetogenic clostridia, considered genetically inaccessible not even a decade ago (1). While automated high-throughput engineering is possible for these anaerobic organisms today, cell-free systems offer a path to further reducing cycle times and maximizing throughput, accelerating pathway engineering by more than an order of magnitude beyond what is feasible today.

We have demonstrated the application of cell-free systems to guide various aspects of strain engineering, including selection of best pathway variants and optimal expression levels (2) or prioritizing gene knock-outs for competing reactions. Furthermore, we developed a new framework that allows to seamlessly go from cell-free to cell designs and feed into ensemble and machine learning models.

First, we established an *in vitro* Prototyping and Rapid Optimization of Biosynthetic Enzymes (iPROBE) platform where cell lysates are enriched with biosynthetic enzymes by cell-free protein synthesis and then metabolic pathways are assembled in a mix-and-match fashion to assess pathway performance. Through this approach, we demonstrated optimization of two multistep pathways, leading to 20-fold improvement in cellular production (2). Often times, introduction of a new pathway into a cellular host leads to formation of unwanted byproducts through native reactions. Identification and iterative knock-outs of the responsible enzyme(s) can be cumbersome and time-consuming. We have demonstrated that the iPROBE platform can rapidly identify candidate enzymes to guide cellular engineering.

To enable a seamless transition from iPROBE to cellular engineering, we have established a GoldenGate based vector system that work in context of cell-free protein synthesis and allow identified gene variants to be directly assembled into cellular expression constructs in a single step without re-synthesis or complex cloning and we showed automation of the whole workflow.

Using this platform, we have optimized pathways for acetone, butanol and 3-HB and have demonstrated completely new biosynthesis routes to 1,3-BDO and MEG.

References

1. Liew et al. (2019) Gas Fermentation – A Flexible Platform for Commercial Scale Production of Low Carbon Fuels and Chemicals from Waste and Renewable Feedstocks. *Front Microbiol* 2016, **7**:694.
2. Karim et al. (2019) *In vitro* prototyping and rapid optimization of biosynthetic enzymes for cellular design. *BioRxiv* doi: https://doi.org/10.1101/685768

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.

Determining Protospacer Adjacent Motif Preferences of Industrially Relevant Clostridial Type I-B CRISPR-Cas Systems

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.

The recent discovery and in-depth characterization of CRISPR-Cas9 has led to a variety of technologies, including genome editing, genome modification, nucleic acid sensing, and next generation antimicrobials. Although CRISPR-Cas9 is a powerful tool, it is representative of only a small fraction all of CRISPR-Cas systems and is toxic when heterologously expressed in many bacteria. Fortunately, about half of all bacteria that have been sequenced encode at least one CRISPR-Cas system in their genome which provides an alternative to CRISPR-Cas9 for genome manipulation. The most prevalent subtype of CRISPR-Cas system is type I-B, which comprises 20% of all naturally occurring CRISPR-Cas systems. This subtype is predominant in Firmicutes, including solventogenic Clostridia which have immediate relevance to industrial bioprocessing. Endogenous type I-B CRISPR-Cas systems have been used to successfully edit the genomes of *Clostridium pasteurianum*, *Clostridium tyrobutyricum*, and *Clostridium thermocellum*, but expansion to other organisms is hindered by the unique protospacer adjacent motif (PAM) sequence of each CRISPR-Cas system required to target a DNA sequence. Past efforts address this by generating nucleotide alignments between CRISPR array spacers and sequences within a variable database, manually curating alignments to hypothesize a few potential PAMs, and then testing these hypotheses individually by plasmid interference assays. Here we present a standardized *in silico* pipeline to predict PAM sequences for a given CRISPR-Cas system from annotated CRISPR spacers and use crude cell-free extracts to characterize all possible 5-

nucleotide PAM sequences for the *Clostridium autoethanogenum* type I-B CRISPR system. The *in silico* PAM prediction pipeline is able to recapitulate experimentally determined PAMs from a variety of CRISPR-Cas systems. The prediction strength and accuracy is heavily dependent on the number of nucleotide alignments generated in the first step which makes it difficult to predict PAMs for CRISPR-Cas systems with few spacer sequences or within organisms that have few sequenced relatives and mobile genetic elements. To test these predictions, we used our recently developed *C. autoethanogenum* cell-free system to experimentally screen PAM sequences by depletion of variants from a 5-nucleotide randomized PAM library. The depletion assay and subsequent validation indicate a conserved CCW PAM 5' to the protospacer, different from what previous sequence alignment-based methods have predicted. We anticipate that this new pipeline and experimental strategy will aid in standardizing PAM prediction and in characterizing PAM preferences in new CRISPR-Cas systems.

We acknowledge the Department of Energy grant DE-SC0018249 for funding of this project.

Establishing an Automated High-throughput Screening Platform

Rasmus O. Jensen (rasmus.jensen@lanzatech.com),[1]* Kathrine Barkus,[1] Asela Dassanayake,[1] Sean D. Simpson,[1] Michael Köpke[1] and **Michael C. Jewett**[2]

[1]LanzaTech Inc, Skokie, IL; [2]Northwestern University, Evanston, IL

http://www.lanzatech.com; http://jewettlab.northwestern.edu

Project Goals:

Automation of molecular biology workflows is accelerating strain development cycles in model organisms; however, such processes have yet to be fully integrated in non-model organisms such as gas-fermenting clostridia. Implementation of automation technology to these organisms has been impeded by the biological limitations in transforming, modifying and screening the organism in presence of anaerobic conditions and toxic flammable gases. As a part of establishing a new interdisciplinary venture, the clostridia Foundry for Biosystems Design (cBioFAB) that combines advancements in cell-free and *Clostridium* engineering metabolic engineering we are developing a fully automated system for strain generation and screening of *Clostridium autoethanogenum*.

Rising levels of greenhouse gases in atmosphere, and resulting instability in climate, pose significant economic and social challenges at a global scale. Technologies that enable capture and conversion of waste gases, such as carbon dioxide and carbon monoxide, into useful product streams can help mitigate negative effects of climate change while enabling a new carbon (neutral) economy. LanzaTech was founded with this mission in mind, and in its fifteen years of existence, it has demonstrated successful conversion of waste gaseous streams into fuels and chemicals (such as ethanol, acetone, isopropanol) at scale and commercialization of gas-to-ethanol production with over 11 million gallons of ethanol produced to date. LanzaTech has, over the past two years, developed a custom built, fully automated strain engineering and screening platform in context of anaerobic conditions and toxic and flammable gases.

Currently, the system is capable of performing fully integrated and automated workflows including transformation, picking colonies, liquid handling operations. This enables us to do automated strain engineering, knock-outs/ins, plasmid introduction and performing high-throughput growth experiments and make and review freezer stock of the resulting strains. The system has a current capacity of screening thousands of strains at a time.

We have validated the capabilities of the system through growth experiments in several different multiwall plate formats, and through testing thousands of strains for various combinatorial libraries for the production of target molecules such as acetone, butanol, 3-hydroxybutyrate (3-HB), 1,3-butanediol (1,3-BDO) or monoethylene glycol (MEG).

The large amount of generated data is automatically captured in a custom-built LIMS system and feeding into genome scale model and machine learning tools.

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.

Integrating Proteomic and Metabolomic Analyses to Optimize Cellular Extract Preparation for Enhanced Cell-Free Protein Synthesis

Payal Chirania,[1,2] Richard J. Giannone[1*] (giannonerj@ornl.gov), Nancy L. Engle,[1] Grant A. Rybnicky,[3] Blake Rasor,[4] 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]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL; [4]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL.

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. 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.

Although biocatalysts offer economic and efficient solutions for many biomanufacturing processes, there are limitations with using live cells for industrial production of potentially toxic compounds. Cell-free expression (CFE) systems provide an alternative solution for high-throughput enzyme prototyping and subsequent biomanufacturing of valuable chemicals that may or may not be compatible with living cells. A key component of CFE systems are cellular extracts, which provide the molecular machinery (i.e. transcription, translation) and metabolic intermediates required for biosynthesis. The overall activity, productivity and/or capability of these systems is thus intrinsically dependent on the extract preparation methodology itself. Routine procedures employed during extract preparation include sonication, ribosomal runoff, dialysis, and application of exogenous energetic molecules – slight modifications to any of these can affect CFE system output. Although the sonication energy applied to lyse cells has been optimized for generating highly productive extracts, the effect of runoff and dialysis has not been studied in detail. Proteins and metabolites form key components of any extract and improper processing can result in the loss of important factors required for optimal biosynthesis. To further investigate the impact of these processing steps on extract quality and to better inform future optimization efforts, an integrated omics analysis was performed. Cell-free extracts from *E. coli* BL21 utilizing different processing steps – no processing, after runoff, after dialysis, and after dual runoff and dialysis – for two commonly used salt conditions (acetate and glutamate) were prepared for proteomic and metabolomic analysis. Initial results indicate 87 proteins to be differentially abundant in the acetate extracts after both runoff and dialysis were applied; the condition in which the most dramatic decrease in activity was observed. Three of these proteins were related to transcription, a critical step for CFE, and were also altered after application of runoff alone. The metabolite profile was

also altered after dual processing, with the greatest effect driven by dialysis. The largest reduction was observed in nitrogen-containing metabolites, including amino acids, polyamines, purines, and pyrimidines, all of which are likely important for lysate productivity. Overall, this integrated analysis will help optimize cell-free extract preparation and provide information about the molecules required for cell-free activity in CFE systems.

This work is sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. Award number DE-SC0018249.

Kinetic Modeling Tools Using Cell-Free Experiments to Predict Metabolic Network Behavior in Non-Model Systems

Jacob Martin* (jacobmartin2022@u.northwestern.edu), Ashty Karim, Blake Rasor, 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 a predictive model of metabolism in bacterial cell-free systems for the purpose of rapidly prototyping heterologous metabolic pathways. This model may be used to both optimize the production of metabolites in cell-free systems, as well as to understand how results in these systems should inform design in living organisms.

Metabolic engineering efforts with non-model organisms present many opportunities in small molecule production due to the 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. Additionally, the same complexity that allows these microbes to produce so many interesting compounds makes the task of cloning and testing all possible factors within the cell experimentally infeasible. While the maturation of cell-free technologies has allowed the rapid testing of many experimental conditions for a heterologous pathway outside the cell, there remains a need for methods to analyze these data and understand the connection between these cell-free results and how these experimental conditions will perform once translated to *in vivo* production strains. To accomplish this goal, we are using kinetic models, based on the metabolic ensemble modeling (MEM) framework, to elucidate the underlying kinetic parameters of the pathways of interest in these cell-free systems. Currently, we are working towards a model of cell-free metabolism which will both predict conditions to optimize production in cell-free, as well as recommend future experiments which will most efficiently train our model. Ultimately, we aim to construct a model framework which will utilize these high-throughput cell-free experiments to predict, for a given heterologous pathway of interest, which enzyme homologs and expression levels are most likely to maximize target production rate when applied *in vivo.* We have developed several novel features to the MEM framework which will allow us to uniquely utilize this cell-free data to accomplish these goals.

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.

Predicting Novel Biosynthetic Pathways with Generalized Enzymatic Reaction Rules

Zhuofu Ni* (joseph.ni@u.northwestern.edu), Linda Broadbelt, **Keith Tyo**

Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL

tyolab.northwestern.edu; broadbelt.northwestern.edu

Project Goals: Curate a set of generalized enzymatic reaction rules that can predict the entire space of possible metabolic reactions, and create a cheminformatics framework to predict novel biosynthetic pathways leading to bioproduction of valuable molecules

Enzyme promiscuity, where enzymes may catalyze a range of side reactions in addition to its main reaction, is a widely recognized yet still largely unexplored phenomenon in biological systems. This could open up vast possibilities for bioproduction of valuable chemicals, where novel biosynthetic pathways leading to molecules of interest can be constructed based on enzymes with desired promiscuous reactions. In order to design these biosynthetic pathways, we curated a set of generalized enzymatic reaction rules capable of describing most enzymatic transformations, which can be applied in the Biochemical Network Integrated Computational Explorer (BNICE) platform to predict the entire space of possible metabolic reactions. We have created a minimum number of these rules verified to 1) comprehensively cover known reactions across metabolic databases, 2) describe reactions with the maximum level of promiscuity, and 3) represent unique enzymatic transformations. Based on these rules, a cheminformatics workflow is developed to efficiently predict and prune novel biosynthetic pathways towards molecules of interest, and systematically identify the most promising ones for experimental validation. By leveraging the entire knowledge of possible metabolic reactions through enzymatic reaction rules, we are able to accelerate novel pathway design and enable bioproduction towards a wide range of new molecules.

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.

Sequencing and Gene Mining the Largest Collection of Industrially used Acetone-Butanol-Ethanol (ABE) Fermentation Strains

Steven D. Brown[1]* (Steve.Brown@LanzaTech.com), Rasmus O. Jensen,[1] James Winkler,[1] Shilpa Nagaraju,[1] Wayne Mitchell,[1] Vinicio Reynoso,[1] Dawn M. Klingeman,[2] Zamin K. Yang[2], Nicole Shapiro,[3] Natalia Ivanova,[3] Frederik Schulz,[3] Tanja Woyke,[3] David Paez-Espino,[3] Nigel Mouncey,[3] Sean D. Simpson,[1] Michael Köpke,[1] and **Michael C. Jewett[4]**

[1]LanzaTech Inc, Skokie, IL; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]DOE Joint Genome Institute, Berkeley, CA; [4]Northwestern University, Evanston, IL

http://www.lanzatech.com; http://jewettlab.northwestern.edu

Project Goals: The Clostridium Foundry for Biosystems Design (*c*BioFAB) is developing an integrated framework that includes computational modeling, cell-free technologies, system-level omics data, high-throughput anaerobic strain construction and cultivation to rapidly model, design, and predictably engineer industrial clostridia strains to manufacture a variety of fuels (e.g., butanol) and building block chemicals. *Clostridium autoethanogenum* is a model acetogen used in commercial scale gas fermentation to produce bioethanol and it is being further developed by *c*BioFAB and related projects.

Clostridium autoethanogenum can use a wide range of CO, CO_2 and hydrogen containing gases for carbon and energy via the Wood-Ljungdahl pathway. To date, over 11 M gallons (>42 M liters) of fuel ethanol have been produced, avoiding nearly 60k metric tons of CO_2 emissions. A genetic toolbox exists that includes systems for heterologous expression via plasmid and chromosomal integration, gene deletion, CRISPR systems, validated genetic parts and codon adaptation algorithms.

The Acetone-Butanol-Ethanol (ABE) fermentation was one of the first large-scale industrial chemical production processes, widely deployed for the first part of the 20th Century and until 1983 in South Africa (1). The ABE process is a sugar-based fermentation using species in genus *Clostridium,* but not *C. autoethanogenum*. Acetone and butanol are non-native products for *C. autoethanogenum.* In collaboration with the US Department of Energy Joint Genome Institute (JGI), an historic collection of commercial solventogenic clostridia strains have had their genome sequences determined to survey ABE gene variants for heterologous expression, to potentially identify industrial robustness traits, regulatory networks and to investigate CRISPR systems and relationships between bacteriophage and prophage and immunity.

We have generated 273 new ABE genome sequences, with most being *Clostridium beijerenckii* (179) and *Clostridium saccharobutylicum* (75) strains and other species including *Clostridium acetobutylicum, Clostridium butyricum, Clostridium saccharoperbutylacetonicum* and *Clostridium tetanomorphum*. There are 208 strains where ten or fewer contigs represented the

genome sequences and majority of remaining genome assembled from Illumina only data. The 16S rRNA gene count ranged from 1 to 23 copies. Each genome was scanned for genes in the clostridial acetone fermentation pathway, including thiolases (*thlA*), acetoacetate:butyrate/acetate CoA transferases (*ctfAB*), and acetoacetate decarboxylases (*adc*), after which identical sequences were dereplicated with CD-HIT (100% clustering threshold) to identify the unique candidate genes. A total of 30 (from 508 gene calls) thiolases, 32 (506) CtfA, 30 (506) CtfB, and 13 (272) Adc-like unique proteins were identified, which are being screened via cell-free protein synthesis (CFPS) and through strain characterization. In a later study, the collection was mined for unique gene variants for the butanol pathway with 52 thiolases (*thlA*), 40 3-oxoacyl-CoA reductases (*hbd*), 49 3-hydroxyacyl-CoA dehydratases (*crt*), 43 trans-enoyl-CoA reductases (*ter*), 9 acyl-CoA/aldehyde reductases (*acr/ald*) and 236 phosphobutyrylases/butyrate kinases (*ptb-buk*) sequences identified. These gene variants (*thlA-hbd-crt-ter*) are being screened, along with termination enzymes (*acr/ald-ptb-buk*), via cell-free protein synthesis to identify the most promising candidates for *in vivo* butanol production.

About a third of the sequenced strains in collection have a complete CRISPR system with arrays and *cas* operons. All CRISPR systems across the different species are Type-IB, meaning they require multi-subunit effector complex or CASCADE proteins and a 5' protospacer acquisition motif (PAM) for interference. Except for five *C. beijerinckii* strains, all other CRISPR carrying strains have indications of phage and/or prophage infecttion. Of the putative infected, ten strains have spacers targeting the integrated phage and/or prophage. One *C. tetanomorphum* strain has 12 spacers targeting putative phage and/or prophage within its genome with a likely PAM of "CCN".

The new genome resources will facilitate biofuel and chemical production for a variety of biocatalysts and will enable broader biosystems design.

References

1. Jones DT and Woods DR (1986). Acetone-butanol fermentation revisited. *Microbiol Rev.* 50: 484–524.

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. The work conducted by the U.S. Department of Energy Joint Genome Institute (JGI) is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Transforming our understanding of chloroplast-associated genes through comprehensive characterization of protein localizations and protein-protein interactions

Lianyong Wang, Princeton University
Bibin Paulose, Michigan State University
Danny J. Schnell, Michigan State University
Martin C. Jonikas, Princeton University

Our project aims to generate a map of protein localizations and protein-protein interactions for 4,262 genes associated with the chloroplast, the energy-producing organelle which is a hallmark of plants. We consider these genes a high priority set 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.

This project consists of two synergistic elements, one in the unicellular model green alga *Chlamydomonas reinhardtii* and the other in 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 will 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.

The project is based on extensive preliminary data from the PI and Co-I demonstrating feasibility, quality and value of the work. Significant progress has already been made towards tagging and localizing many of the target proteins. The team has a strong track record of developing large-scale resources for the community and advancing our basic understanding of chloroplast biology.

This study will transform our understanding of photosynthetic eukaryotes and will open a broad range of engineering opportunities. The localization data has the potential to reveal novel classes of protein localization, providing insight into the sub-organellar structure of chloroplasts. 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 data will improve annotations for thousands of other genes. The project will also have a long-term impact as the scientific community utilizes the resource of strains, constructs and data.

Ecosystem responses in switchgrass monoculture stands across a latitudinal gradient

Roser Matamala[1], Julie D. Jastrow[1], Michael P. Ricketts[1*] (mricketts@anl.gov), Felix Fritschi[2], Philip A. Fay[3], Joel Reyes-Cabrera[4], Timothy W. Vugteveen[1], Scott M. Hofmann[1], Jason Bonnette[5], **Thomas E. Juenger[5]**

[1]Environmental Sciences Division, Argonne National Laboratory, Lemont, IL; [2]University of Missouri, Columbia, MO; [3]Agricultural Research Service, United States Department of Agriculture, Temple, TX; [4]Eastern New Mexico University, Portales, NM; [5]University of Texas, Austin, TX

https://sites.cns.utexas.edu/juenger_lab/switchgrass

Project Goals: As part of a collaborative effort, our project aims to evaluate the potential for sustainable switchgrass (*Panicum virgatum*) production as a biofuel product in North America and better understand the potential ecosystem services associated with large-scale production of perennial bioenergy crops. Specifically, by linking switchgrass growth responses and physiology, belowground plant and microbial activity, and key ecosystem carbon (C) pools and fluxes to variations in climate and edaphic properties, we seek to provide a conceptual mechanistic framework that explains the interactions between switchgrass productivity/morphology/phenology and ecosystem C dynamics.

We explored stand level ecosystem responses of multiple switchgrass cultivars planted along a continental gradient that varied in environmental and soil properties. Stands of switchgrass consisting of 3 lowland cultivars (Alamo, Kanlow, and Liberty) and 3 upland cultivars (Blackwell, Cave-in-Rock, and Carthage) were established during spring 2016 in 6m x 6m plot monocultures arranged in a randomized design with 5 replicates at TX, MO, and IL field sites. Baseline characterization of total soil organic C (SOC), permanganate-oxidizable C (POxC; representing the biologically active SOC pool), and total soil nitrogen (TN) contents to a depth of 2m were determined at the time of planting for each of the sites (Figure 1). Measurements of plant growth, biomass, phenology, and tissue chemistry were coupled with collection of associated soil samples starting at 100% green-up and continuing at 14-20 day intervals through all stages of plant development. To determine the effects of the different physiological and phenological characteristics of switchgrass cultivars on ecosystem dynamics, we measured CO_2 gas-fluxes including net ecosystem exchange (NEE) and soil respiration (R_s) at all sites for Cave-in-Rock and Alamo, representing upland and lowland cultivars respectively. In addition, R_s measurements were made in root exclusion zones to partition heterotrophic from autotrophic respiration (R_h and R_a, respectively). Associated soil samples were analyzed for POxC to investigate plant and seasonal effects on the biologically active SOC pool.

We found initial evidence to suggest that differences in soil properties could lead to variations in long term nutrient availability and sustainability among sites that have not yet been manifested in plant productivity. In MO, where SOC and TN concentrations were lowest, aboveground plant

tissues of all cultivars had far wider C:N ratios throughout most of the growing season. Phenological differences were most noticeable between northern sites (IL & MO) and the southern site (TX) where green-up for all cultivars occurred an average of ~50-days later at the northern sites. Flowering times also differed between upland and lowland cultivars, particularly in TX where upland cultivars flowered an average of 68-days earlier than lowland cultivars. In northern sites, lowland cultivars also flowered later (~24-days), and in 2018 Alamo failed to completely senesce before harvest. Overall, after 3-4 years in switchgrass, the active SOC pool was smaller relative to pre-planting levels in TX and IL, and slightly greater in MO where initial total SOC concentrations were already very low. Variations in active SOC among cultivars are evident, however observed patterns are so far not attributable to any definitive environmental factors. Soil respiration differences between the lowland cultivar Alamo and the upland cultivar Cave-in Rock were most noticeable between the IL and MO sites where overall R_s was greater in IL for Alamo, but greater in MO for Cave-in Rock. However, partitioning out R_h from these measurements shows similar R_a for both cultivars regardless of site and suggests greater soil microbial activity in IL relative to MO. In 2018, a noticeable drop in the amount of active SOC at the TX site was linked to a drought that reduced both biomass production and ecosystem respiration, which likely also inhibited root exudation and soil microbial activity compared to the other sites.

These findings suggest that switchgrass ecosystem dynamics are affected by ecotype responses to latitudinal variations in climatic conditions that are also intimately linked with belowground soil C and N processes. Ultimately, understanding the connections between the environmental factors underlying plant and soil linkages may have implications for determining optimal management strategies for the implementation and sustainability of large-scale switchgrass production in North America.

This research is funded by the Office of Biological and Environmental Research within the Department of Energy Office of Science.

Genetics of Climate Adaptation Using Genome-Wide Association in Switchgrass

Alice MacQueen[1]* (alice.macqueen@utexas.edu), Jason Bonette[1], John Lovell[2], Sujan Mamidi[2], Jeremy Schmutz[2], **Thomas Juenger**[1]

[1]The University of Texas at Austin; [2]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama

Project Goals:

- **Establish common gardens of clonally replicated switchgrass genotypes to study ecotype divergence, local adaptation, and the spatial scale of genome-by-environment interaction 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. In this poster, I present genetic analyses of climate adaptation in switchgrass using climate variables from each genotype's location of origin, and genetic analyses of phenology data from the 2019 growing season. In particular, I present preliminary results from genome-wide association studies aimed at detecting genes associated with specific climate adaptations of switchgrass across the species' latitudinal range.

References

1. https://github.com/Alice-MacQueen/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 to TEJ.

Host Genetics Control the Composition of Root-associated Microbiota in Switchgrass (*Panicum virgatum*)

Joseph Edwards*[1](j_edwards@utexas.edu), Usha Saran[1], Jason Bonnette[1], Jane Grimwood[2], Felix Fritschi[3], **Thomas Juenger**[1]

[1]University of Texas, Austin; [2]Hudson Alpha Institute for Biotechnology, Huntsville, AL; [3]University of Missouri, Columbia

• Establish common gardens of clonally replicated switchgrass genotypes to study ecotype divergence, local adaptation, environmental gradients (e.g., precipitation, temperature, and soils).

• Identify genomic regions underlying adaptation and sustainability in switchgrass using genome-wide associations.

• Characterize the relative roles of switchgrass genotype and local environments in the assembly of the switchgrass bacterial and fungal microbiome communities in natural habitats.

Plant root-associated microbiota mediate important below-ground processes such as nutrient availability, pathogen inhibition, and general growth promotion. Assembly of root microbiota is significantly impacted by the genotype of the host plant and the host's environment; however, specific loci in the host genome influencing the composition and relative abundance of root microbiota have not been identified. In this study, we use experimental populations of the bioenergy crop switchgrass (*Panicum virgatum*) planted across a latitudinal gradient to identify genomic loci affecting the composition of root-associated microbiota. We find many bacterial strains display heritable variation, and by using a QTL mapping approach, we identified genomic regions in the host plant associated with the abundance of specific bacterial taxa. A majority of the identified QTL showed consistent effects across locations, and only a few QTL displayed environmental deviations. Interestingly, many of the identified QTL coregulate multiple microbes, suggesting that either the same host signaling and response mechanisms govern the acquisition and maintenance of multiple bacterial strains, or that QTL may act on microbial keystone species, which in turn affect the abundance of other microbial taxa. Together, these results indicate that root-associated bacteria are under genetic control of the host and that altering the microbiome for greater yields or ecosystem sustainability may be possible through manipulation of host plant genetics.

This research was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research award number DE-SC0014156 to TEJ.

Spatiotemporal dynamics of a microbiome on *Panicum hallii* under drought stress

Esther Singer[1,2,*] (esinger@lbl.gov), Joseph Edwards[3], Joel Reyes-Cabrera[4] , Tanja Woyke[1], **Thomas E. Juenger**[3]

[1]Joint Genome Institute, Berkeley, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]University of Texas, Austin, TX; and [4]Eastern New Mexico University, Portales, NM

https://sites.cns.utexas.edu/juenger_lab/switchgrass

Project Goals: Short statement of goals.

Our project aimed at establishing a high-throughput pipeline for characterizing diversity and community composition of the switchgrass and *Panicum hallii* microbiome. The pipeline we aimed to establish includes the optimization of 1) sampling techniques of various plant tissue types, 2) sample processing protocols as well as 3) the streamlined analysis of large 16S rRNA and ITS amplicon sequence datasets. The results of this project contribute essential information for the exploration of plant-microbe-soil interactions across continental scale environmental gradients.

Abstract text. Please limit to 2 pages.

Part of the DOE's strategy to ensure American energy independence is to produce biofuels from dedicated biomass crops. Achieving DOE's ambitious goal of displacing 30% of 2004 gasoline demand with biofuels by 2030 will require major increases in plant productivity. Switchgrass (*Panicum virgatum*) has been championed as a promising bioenergy species, but widespread commercial use has partly been challenged by its genetic complexity. *Panicum hallii* is a close relative of agronomic switchgrass with a diploid genome and seed-to-seed time of 8 weeks, offering researchers a model system for exploring *Panicum* genetics, genomics, and adaptation for agronomic improvement. Furthermore, plant microbiomes are known to influence many aspects of plant health. We present biogeochemical dynamics including amplicon and shotgun sequencing data, soil chemistry and metabolite profiles as well as plant phenotypic characteristics. Interestingly, we observe significant impacts of soil depth on rhizosphere microbiome, but not on bulk soil communities. Drought treatment did not significantly change microbial community composition or soil chemistry. Plant development stages significantly affected individual bacterial species within depth horizons. This study is rare in its comprehensive biogeochemical characterization of environment and biological players and provides a framework for soil ecosystem science conducted in future EcoPod experiments. All data is integrated in an interactive we application that is publicly available.

This work was funded by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility supported under Contract No. DE-AC02-05CH11231 and by the Office of Science (BER), U.S. Department of Energy, Grant no DE-SC0014156.

The Genomic Basis of Ecotype Evolution in Switchgrass

Thomas E. Juenger[1], John T. Lovell[2,3]* (jlovell@hudsonalpha.org), Adam Healey[2,3], Shengqiang Shu[3], Jason Bonnette[1], Kerrie Barry[3], Dan Rohksar[3], Jane Grimwood[2] and Jeremy Schmutz[2,3]

[1]University of Texas, Austin, TX, USA; [2]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; [3]Department of Energy Joint Genome Institute, Berkeley, CA, USA;

Project Goals: Land use change, resource limitation and climate extremes threaten the sustainability of both agricultural and natural ecosystems. In plants, breeding can improve crop resilience to novel stresses, and successful ecological restoration can buffer the effects of ever-shrinking natural habitats; however, these efforts require sufficient knowledge of the traits, genes and environments that underlie productivity and adaptation. Here, we demonstrate how the development of genomic resources in the biofuel crop and widespread tallgrass prairie species, Switchgrass, permits inference of the processes of climate adaptation and definition of genetic loci that underlie climate-dependent growth.

Funding Statement: This research was supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research Award DESC0014156 to T.E.J. and DE-SC0017883 to D.B.L. Funding was provided by National Science Foundation Plant Genome Research Program Awards IOS0922457 and IOS-1444533 to T.E.J. and NSF/IOS-1402393 to J.T.L. This research was also based upon work supported in part by the Great Lakes Bioenergy Research Center, US Department of Energy, Office of Science, Office of Biological and Environmental Research under Awards DESC0018409 and DE-FC02-07ER64494. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract DE-AC02-05CH11231.

Using Machine Learning to Identify Cultivar x Site Interaction and Environmental Variable Affecting Aboveground Biomass

Li Zhang[1]*, Kathrine D. Behrman[1], Jason Bonnette[1], Thomas E Juenger[1]

[1]University of Texas at Austin, Austin TX

Project Goal:

Multi-Scale Modeling will define conditions of a sustainable biofuel system and identify key tradeoffs between genetic diversity, productivity, and ecosystem services. The aim is to use multiscale modeling to predict switchgrass performance and sustainability under future climate change.

Abstract:

Switchgrass, a perennial grass native to North American, is a promising second generation biofuel crop. Previous studies have showed that switchgrass yields are sensitive to climatic variability associated with temperature and precipitation in space and time. Switchgrass yields are often cultivar-dependent and exhibit cultivar x environment interactions as well. In this study, we grew six switchgrass cultivars (two lowlands: Alamo and Kanlow; two uplands: Blackwell and Cave-in-Rock; and two hybrids: Liberty and Carthage) at three locations across US (TMPL: Temple TX; CLMB: Colombia, MO; and FRMI: Fermin Lab, IL) for three consecutive years (2017-2019). Seasonal data on destructive leaf area index (DesLAI), tiller count (TC), plant height (HT) was collected and aboveground biomass (DesBIO) was measured each year. Our preliminary statistical analyses show that end of season aboveground biomass is strongly affected by site and genotype but not year. To further investigate these interactions, multivariate time series clustering was used to determine how two phenotypes, seasonal TC and DesBIO, are related for all cultivars for all locations. Not surprisingly, the two upland cultivars and the two lowland cultivas make two distinct clusters regardless of site. The hybrid Liberty always clutters with the lowland cultivar, indicating that phenotypically it is a lowland. On the other hand, the other hybrid, Carthage, phenotypically resembles an upland in the northern sites (CLMB and FRMI) and a lowland in the southern site (TMPL), thus indicating this cultivar is displaying strong cultivar x environment interactions. Next, machine learning (i.e., random forest algorithm) was used to identify the environmental variable(s) affecting these phenotypes. Average temperature between each sampling interval was identified as the major factor influencing aboveground biomass. There is a positive linear relationship between temperature and biomass when average temperature is between 15 to 25°C. At average temperature values between 25 to 30°C, biomass plateaus and is constant. This study is an example of how intra-annual time series data and machine learning methods (multivariate time-series clustering and random forest) can help identify cultivar by site interaction and the shape between environmental variables and phenotypes.

Funding Statement:

This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER), award number DE-SC0014156 to TEJ.

A Droplet Microfluidic Platform for Lab Automation

Kosuke Iwai[1,2]* (KosukeIwai@lbl.gov), Maren Wehrs,[1] Megan Garber,[1] Jess Sustarich,[1,2] William R. Gaillard,[1,2] Peter W. Kim,[1,2] Kai Deng,[1,2] Trent R. Northen,[1,4,6] 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 establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops 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 have attracted a significant attention for performing biochemical reactions and analysis as they provide improvement over their macroscale counterparts in cost, amounts of reagents required, speed, and integration.

We are developing microfluidic devices for many biofuel research applications including enzyme screening, enzyme evolution, and synthetic biology. 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 (references). 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. Electroporation is achieved by placing pairs of electrodes in each chamber to apply voltages to the arrayed droplets. This configuration allows us to customize the electroporation condition at each droplet and scale-up the numbers of

reactions by making high-density electrode arrays. We integrate optical fibers in the microchannels for fluorescence-based detection of encapsulated cells and enzymatic activities in the discrete droplets, and for triggering sorting of droplets. We used the microfluidic platform for automating CRISPR-based MAGE recombineering in *E. coli* to optimize the biosynthetic pathway of an example molecule, indigoidine.

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.

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.

Adaptive Laboratory Evolution as an Efficient Technology for Strain Construction

Adam M. Feist (afeist@ucsd.edu) [1,2,3]*, Hyun Gyu Lim[1,3], Thomas Eng[3], Aindrila Mukhopadhyay[3], Steven W. Singer[3,4], Bernhard O. Palsson[1,2,3], Jay D. Keasling (**Project PI**) [1,3]

[1]Department of Bioengineering, University of California at San Diego, San Diego, CA, USA [2]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark [3]Joint Bioenergy Institute, Lawrence Berkeley National Lab, Berkeley; [4]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA [5]University of California at Berkeley, Berkeley, CA, USA

http://jbei.org

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Harnessing the process of natural selection to obtain and understand new microbial phenotypes has become increasingly possible due to advances in culturing techniques, DNA sequencing, bioinformatics, and genetic engineering. Accordingly, Adaptive Laboratory Evolution (ALE) experiments represent a powerful approach to both investigate the evolutionary forces influencing strain phenotypes, performance, and stability, and to acquire production strains that contain beneficial mutations. This poster describes a full technology platform focused on ALE and successful use cases applying the platform to solve important issues in industrial biotechnology. Further, a list of current and future application areas is provided highlighting how ALE can be utilized as an efficient design and build tool for strain construction.

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

An Automated Sample Preparation Workflow For High-throughput, Quantitative Proteomic Studies of Microbes

Yan Chen,[1,2] Joel M. Guenther,[1,3] Jennifer W. Gin,[1,2] Leanne Jade G. Chan,[1,2] Zak Costello,[1,2] Tadeusz L. Ogorzalek,[1,2] Huu M. Tran,[1,3] Jacquelyn M. Blake-Hedges,[1,2,4] Jay D. Keasling,[1,2,5,6] Paul D. Adams,[1,2,5] Hector García Martín,[1,2] Nathan J. Hillson,[1,2] **Christopher J. Petzold** [1,2,*]

[1]Joint BioEnergy Institute, Emeryville, California
[2]Lawrence Berkeley National Laboratory, Berkeley, California
[3]Sandia National Laboratories (NTESS), Livermore, California
[4]Department of Chemistry, University of California Berkeley
[5]Department of Bioengineering, University of California Berkeley
[6]Department of Chemical and Biomolecular Engineering, University of California, Berkeley

www.jbei.org

Project goals: To develop an automated sample preparation workflow for quantitative proteomic analysis of engineered microbes.

Mass spectrometry-based quantitative proteomic analysis is key for biotechnology-related research and development. Driving this value have been improvements in sensitivity, resolution, and robustness of mass analyzers. However, manual sample preparation protocols are often a bottleneck for sample throughput and can lead to poor reproducibility. To alleviate these issues, we developed a "cells-to-peptides" automated workflow for Gram-negative bacteria and fungi that includes cell lysis, protein precipitation, resuspension, quantification, normalization, and tryptic digestion[1]. The workflow takes 2 hours to process 96 samples from cell pellets to the initiation of the tryptic digestion step and can process 384 samples in parallel. We measured the efficiency of protein extraction from various amounts of cell biomass and optimized the process for standard-flow LC-MS systems. The automated workflow was tested by preparing 96 Escherichia coli samples that resulted in a median coefficient of variation of 15.8% for over 600 peptides that were measured. Similar technical variance was observed for three other organisms as measured by highly multiplexed LC-MRM-MS acquisition methods. These results show that this automated sample preparation workflow provides robust, reproducible proteomic samples for high-throughput applications.

References

(1) Chen, Y.; Guenther, J. M.; Gin, J. W.; Chan, L. J. G.; Costello, Z.; Ogorzalek, T. L.; Tran, H. M.; Blake-Hedges, J. M.; Keasling, J. D.; Adams, P. D.; et al. Automated "Cells-To-Peptides" Sample Preparation Workflow for High-Throughput, Quantitative Proteomic Assays of Microbes. *J. Proteome Res.* **2019**, *18*, 3752–3761.

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.

ART: a machine learning Automated Recommendation Tool for synthetic biology

Hector Garcia Martin[1,2,3,12,*](hgmartin@lbl.gov), Tijana Radivojevic[1,2,3], Zak Costello[1,2,3], Kenneth Workman[1,4], Soren Petersen[5], Jie Zhang[5], Andres Ramirez[7], Andres Perez[7], Eduardo Abeliuk[6], Benjamin Sanchez[5], Yu Chen[10,11], Mike Fero[6], Jens Nielsen[5,11,13], Michael Krogh Jensen[5], Jay Keasling[1,4,5,8,9] (**Project PI**)

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]DOE Agile BioFoundry; [4]Department of Bioengineering, University of California, Berkeley,CA, USA; [5]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, Denmark; [6]TeselaGen Biotechnology, San Francisco, CA 94107, USA; [7]TeselaGen SpA, Santiago, Chile; [8]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA; [9]Center for Synthetic Biochemistry, Institute for Synthetic Biology, Shenzhen Institutes of Advanced Technologies, Shenzhen, China; [10]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden; [11]Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, Gothenburg, Sweden; [12]BCAM, Basque Center for Applied Mathematics, Bilbao, Spain; [13]BioInnovation Institute, Ole Maaløes Vej 3, DK-2200 Copenhagen, Denmark

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Traditional synthetic biology approaches involve ad-hoc non systematic engineering practices, which lead to long development times. Here, we present the Automated Recommendation Tool (ART), a tool that leverages machine learning and probabilistic modeling techniques to guide synthetic biology in a systematic fashion. Using sampling-based optimization, ART [1] provides a set of recommended strains to be built in the next engineering cycle, alongside probabilistic predictions of their production levels. We demonstrate the capabilities of ART on a tryptophan producing strain and are able to improve production by 17% compared to best designs used for algorithm training and 106% compared to the initial strain [2].

References

1. Radivojević T, Costello Z, Martin HG (2019) ART: A machine learning Automated Recommendation Tool for synthetic biology. arXiv.

2. Zhang J, Petersen SD, Radivojevic T, Ramirez A, Pérez A, et al. (2019) Predictive engineering and optimization of tryptophan metabolism in yeast through a combination of mechanistic and machine learning models. BioRxiv. doi:10.1101/858464.

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.

Carbon Footprint and Economics of Integrating Biogas Upgrading Process and Carbon Capture Technologies in Cellulosic Biorefineries

Minliang Yang (minliangyang@lbl.gov)[1,2*], Aikaterini Anastasopoulou,[3] Nawa Raj Baral,[1,2] Hanna M. Breunig,[3] and **Corinne D. Scown**[1,2,3]

[1]Life-cycle, Economics, and Agronomy Division, Joint BioEnergy Institute, Emeryville, CA; [2]Biological Systems and Engineering Division, Lawrence Berkeley national Laboratory, Berkeley, CA; [3]Energy Analysis and Environmental Impacts Division, Lawrence Berkeley national Laboratory, Berkeley, CA

http://www.jbei.org

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Gaseous streams in biorefineries have been historically undervalued and underutilized. Biogas produced in existing biorefineries is assumed to be combusted directly on-site to generate process heat and electricity along with lignin. However, biogas can alternatively be upgraded to biomethane, which can be used as a transportation fuel. Biogenic CO_2 generated in biorefineries can also play a critical role in climate change mitigation due to the large amount (45 Mt annually from fermentation). With economic incentives, cellulosic biorefineries could be redesigned to make better use of biogas and concentrated CO_2 streams. To date, the trade-offs associated with biogas upgrading, bioenergy with carbon capture technologies and biogas on-site combustion in cellulosic biorefineries have not been thoroughly investigated. Here, we explore the economic and environmental impacts of upgrading biogas and capturing carbon at cellulosic biorefineries and identify opportunities to maximize value and environmental benefits. The results indicate that biorefineries using biogas upgrading technologies resulted in a similar minimum ethanol selling price as the base case. The amount of carbon captured in biorefineries could be ~90 Mt per year in the U.S. with a maximum 10% conversion of pastureland and cropland to biomass sorghum in future scenarios.

References

1. Sanchez, D. L.; Johnson, N.; McCoy, S. T.; Turner, P. A.; Mach, K. J. Near-term deployment of carbon capture and sequestration from biorefineries in the United States. *Proc. Natl. Acad. Sci. USA* 2018, 115(19), 4875–4880. DOI: 10.1073/pnas.1719695115

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.

Catabolism of Lignin Oligomers by Soil-Derived Microbiomes

Steven W. Singer[1,2]*(SWSinger@lbl.gov), Mee-Rye Park[1,2], Ji Peng[2,3], Jeffrey G. Pelton[4], Ning Sun[2,3], Blake A. Simmons[1,2]

[1]Joint BioEnergy Institute, Emeryville, CA, 94608;[2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720; [3]Advanced Biofuels Process Demonstration Unit, Emeryville, CA 94608; [4]QB3 Institute, University of California-Berkeley, 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 are developing strategies to convert aromatic-rich fractions obtained from lignin to biofuels and bioproducts.

Abstract

Overcoming the recalcitrance of lignin and developing conversion strategies for aromatics are key goals to maximize conversion of carbon in plant biomass. Biomass pretreatment solubilizes and depolymerizes lignin, generating mixtures of oligomeric and monoaromatic molecules. While monoaromatics are metabolized by a variety of bacteria, the metabolism of aromatic oligomers remains largely undescribed. Identifying bacteria that metabolize these oligomers will improve biomass conversions that include lignin as a substrate. Ionic liquid pretreatment with cholinium-ionic liquids generated aromatic-rich fractions that contained size-defined mixtures of monoaromatics and oligomers. *Pseudomonas putida* rapidly consumed the monoaromatics in these mixtures but did not metabolize the oligomers. Enrichment of microbiomes derived from soil generated microbial consortia that consumed the oligomers, as demonstrated by the disappearance of aromatic molecules in the molecular weight range of 300-2000 Da in gel permeation chromatograms and the reduction of characteristic aromatic peaks in the 2D-HSQC NMR spectra. Individual microbiomes consumed the oligomers at different rates and targeted variable bond linkages, indicating that a diverse range of bacteria were involved in the deconstruction of these oligomers. Current efforts are focused on identifying the microbes and enzymes involved in the metabolism of aromatic oligomers and developing methods to elucidate the molecular structures of these oligomers.

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.

Collaboration with the Experiment Data Depot

William Morrell[2,3,4], Mark Forrer[2,3,4], Garrett Birkel[1,3,5], Traci Lopez[2,3,4], Hector Garcia Martin[1,3,5]*, and **Nathan J. Hillson**[2,3,5]* (njhillson@lbl.gov)

[1]Biofuels and Bioproducts and [2]Technology Divisions, DOE Joint BioEnergy Institute; Emeryville, CA 94608, USA; [3]DOE Agile BioFoundry; Emeryville, CA 94608, USA; [4]Sandia National Laboratory; Livermore, CA 94550, USA; [5]Lawrence Berkeley National Laboratory; Berkeley, CA 94720, USA.

https://public-edd.jbei.org

Project Goals:

Although recent advances in synthetic biology allow us to produce biological designs more efficiently than ever, our ability to predict the end result of these designs is still nascent. Predictive models require a large corpus of high-quality data to be usefully parametrized and tested. Suitable datasets for these models are often not generally available. Here, we present the Experiment Data Depot (EDD), an online platform designed to act as a repository of experimental data and metadata. EDD provides a convenient way to upload a variety of data types, visualize these data, and export them in a standardized fashion for use with predictive algorithms. In this poster, we describe EDD and showcase its utility for three different use cases: the characterization of promoters for synthetic biology parts, leveraging proteomics data to improve biofuel yield, and the use of extracellular metabolite concentrations to predict intracellular metabolic fluxes.

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.

References

1. Morrell W, Birkel G, Forrer M, Lopez T, Backman T, Dussault M, Petzold C, Baidoo E, Costello Z, Ando D, Alonso Gutierrez J, George K, Mukhopadhyay A, Vaino I, Keasling J, Adams P, Hillson NJ*, Garcia Martin H*. (2017) "The Experiment Data Depot: a web-based software tool for biological experimental data storage, sharing, and visualization" *ACS Synthetic Biology* DOI: 10.1021/acssynbio.7b00204

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.

Conversion of Ionic Liquid Pretreated Poplar into Jet Fuel

Gina Geiselman[1,3], James Kirby[1,3], Carolina Barcelos[2,4], Gabriella Papa[2,4], Eric Sundstrom[2,4], Taek Soon Lee[1,4] Blake Simmons[1,4] and **John Gladden**[1,3]* (jmgladden@lbl.gov)

[1]Joint BioEnergy Institute, Emeryville, California; [2]Advanced Biofuels and Bioproducts Process Demonstration Unit, Emeryville, California; [3]Biomass Science and Conversion Technology, Sandia National Laboratories, Livermore, California; and [4]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California

https://www.jbei.org

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Microbial production of energy-dense terpene biofuels from ionic liquid (IL) pretreated lignocellulosic biomass is a promising route to sustainable fuels. One-pot IL pretreatment and saccharification effectively releases biomass sugars while streamlining the deconstruction process and conserving water. This work describes a one-pot process with sugar-to-fuel conversion using the versatile basidiomycete, *Rhodosporidium toruloides*. This organism is an attractive lignocellulosic conversion host due to its ability to utilize a wide range of carbon sources, tolerate various ILs, and accumulate acetyl-CoA-derived bioproducts such as fatty acids and terpenes. Researchers at JBEI have previously demonstrated the potential of *R. toruloides* for conversion of IL-treated biomass into terpene bioproducts by engineering a strain to produce bisabolene, a potential alternative biodiesel. Here, we have selected two tricyclic sesquiterpenes that may have applications as jet fuels, epi-isozizaene and prespatane, and demonstrated their production from IL-pretreated poplar using *R. toruloides*.

This work was carried out 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

Distillable Ionic Liquids/Deep Eutectic Solvents for an Effective Recycling and Recovery Approach

Ezinne C. Achinivu[1,2]* (eachinivu@lbl.gov), Mood Mohan[1,2], Anthe George[1,2], Blake Simmons[1,3] and **John Gladden**[1,2]

[1]Deconstruction Division, Joint Bio Energy Institute, Emeryville, California; [2]Biomass Science and Conversion Technology, Sandia National Laboratories, Livermore, California; and [3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California

https://www.jbei.org/research/deconstruction/biomass-pretreatment-process-development/

Project Goals:

Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts. Our current research focuses on developing recyclable IL pretreatment technologies while simultaneously facilitating the efficient depolymerization of both polysaccharides and lignin.

Developing a low-cost and high efficiency lignocellulosic biomass deconstruction process is a critical step towards the widespread adoption of lignocellulosic biofuels. Ionic liquids (ILs) and deep eutectic solvents (DESs) are novel alternative solvents for biomass pretreatment and conversion, and they are most notably one of the most effective methods for producing lignin and high yields of fermentable sugars for bioenergy production. Despite their commercial potential, the cost of IL/DES utilization (typically associated with their synthesis, purification and reuse/recycling) is a significant problem that must be addressed before an affordable IL/DES-based process is commercially viable. Therefore, this study features the use of ***distillable*** solvents for the development of an integrated biomass pretreatment approach that combines effective pretreatment with a simplistic and energy efficient recovery/recycling method.

Protic ionic liquids (PILs) that are formed with the combination of ***organic ammonium-based cations*** and ***organic carboxylic acid-based anions*** are an attractive group of solvents worth considering for this process. PILs are acid-base conjugate ILs that can be synthesized via the direct addition of their acid and base precursors. Additionally, when sufficient energy is employed, they can dissociate back into their neutral acid and base precursors, while the PILs are re-formed upon cooling. This presents a suitable way to recover and recycle the ILs after their application.

The PIL - hydroxyethylammonium acetate - [Eth][OAc] - has already been demonstrated as an effective solvent for biomass pretreatment and is also relatively cheap due to its ease of synthesis.[1] Therefore, this PIL and chemically analogous PILs were studied for their distillability, as well as, their effect on biomass deconstruction in a one-pot/consolidated process. Preliminary results indicate a PIL recovery of 98% for "neat" IL distillation of [Eth][OAc]), followed by PIL recovery [~80-85%] after biomass pretreatment with 15% biomass loading. Following the PIL removal, the

residual biomass was saccharified to generate ~74% total sugars (compared to ~78% sugars for -One pot and ~91% sugars- Early separation).

This is a promising proof of concept that supports our approach for distilling ionic liquids as a recovery method. Once optimized, this will launch our research into economic regimes, making an IL-based biorefinery a realizable goal.

References

1. Sun, J.; Konda, N. V. S. N. M.; Parthasarathi, R.; Dutta, T.; Valiev, M.; Xu, F.; Simmons, B. A.; Singh, S. One-Pot Integrated Biofuel Production Using Low-Cost Biocompatible Protic Ionic Liquids. *Green Chem.* **2017**, *19* (13), 3152–3163. https://doi.org/10.1039/C7GC01179B.

This work was carried out 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

Engineered Polyketide Synthases as Platform for Synthetic Chemistry

Jay Keasling (Project PI), Robert Haushalter, Zilong Wang, Tyler Backman, Alberto Nava, Jacquelyn Blake-Hedges, Amin Zargar

Departments of Chemical & Biomolecular Engineering and of Bioengineering, University of California, Berkeley, CA 94720

Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Polyketides are one of the largest classes of natural products, possessing immense structural diversity and complex chemical architectures. Many polyketides (PKs) are among the most important secondary metabolites for their applications in medicine, agriculture, and industry. Examples include anticancer drugs (epothilone), antibiotics (erythromycin), insecticides (spinosyn A) and antifungals (amphotericin B). These particular examples of polyketides are biosynthesized by multimodular enzyme complexes known as type I modular polyketide synthases (PKSs). Working in an assembly-line fashion, multimodular PKSs assemble and tailor readily available acyl-CoAs within the host cell into large, complex, chiral molecules. Each of these PKSs comprises a series of modules that can be further dissected into a series of domains responsible for the extension of the polyketide back- bone through condensation and selective reductive processing of an acyl-CoA building block. The collinear architecture of these modules, apparent by inspection of the domains present and the predictive selectivity motifs harbored within, provide insights into the chemical connectivity and stereochemical configuration of the polyketide metabolite from analysis of its coding sequence.

While PKSs have been traditionally studied for the production of pharmaceuticals, engineered modular PKSs have the potential to be an extraordinarily effective retrosynthesis platform for the bio-production of products from biofuels and commodity chemicals to both pharmaceutical and nonpharmaceutical fine and specialty chemicals. By rearranging existing polyketide modules and domains, one can exquisitely control chemical structure from DNA sequence alone. However, this potential has only just begun to be realized as the compounds that have been made using engineered PKSs represent a small fraction of the potentially accessible chemical space. We will highlight work from our laboratory where we have engineered PKSs to produce a variety of commodity and specialty chemicals and developed software and high throughput robotic platforms to enable design and construction of PKSs in high throughput.

The material presented in this poster is part of JBEI – Joint BioEnergy 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

Environmental Impacts of Biomass Sorghum Production in the Continental United States

Sagar Gautam[1,2], Umakant Mishra[1,2*] (umishra@anl.gov), Corinne D Scown[2,3,4], Yao Zhang[5]

[1]Environmental Science Division, Argonne National Laboratory, Lemont, IL 60439, United States
[2]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA 94608, United States
[3]Energy Analysis & Environmental Impact Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States
[4] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States
[5]Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523, United States

http://www.jbei.org/

Project goal: Assess biofuel/bioproduct pathway economic and environmental performance at U.S. national scale

Reliable estimates of bioenergy crop yields and their environmental impacts are essential to develop sustainable bioenergy-based land use strategies. In this study, we used the process-based Daily Century (DAYCENT) model with site specific environmental data to simulate Sorghum (*Sorghum bicolor* L. Moench) biomass yield, soil organic carbon (SOC) change, and nitrous oxide emissions across cultivated lands in the continental USA. The simulated rainfed biomass productivity for continental US ranged from 1.4 to 20 Mg ha^{-1} yr^{-1}, with a spatiotemporal average biomass yield of $10^{+0.9}_{-0.9}$ Mg ha^{-1} yr^{-1}, and a coefficient of variation 35%. The spatiotemporal average SOC sequestration and direct nitrous oxide emission rates were simulated as $0.46^{+0.33}_{-0.42}$ Mg CO_2e ha^{-1} yr^{-1} and $0.45^{+0.05}_{-0.03}$ Mg CO_2e ha^{-1} yr^{-1}, respectively. Model predictions were validated using multiyear observed biomass yield data at multiple locations. Compared with field observed data, model predictions of biomass productivity showed a root mean square error of 5.8 Mg ha^{-1} yr^{-1}. Our results suggest 17 million ha cultivated lands in the Southern United States will produce economic Sorghum biomass yield (>10 Mg ha^{-1} yr^{-1}) with net carbon sequestration under rainfed conditions. Cultivated lands of Upper Midwestern states including Iowa, Minnesota, Wisconsin and Michigan showed lower sorghum biomass productivity and net carbon emissions. Our national scale spatially-explicit results are critical for robust lifecycle and technoeconomic analysis of future bioenergy scenarios. Future studies should focus on representing genotypic variations of bioenergy crops and quantifying their total environmental impacts.

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 US Department of Energy under Contract No. DE-AC02-05CH11231.

Field Testing of Engineered Switchgrass with Improved Biomass Yield and Sustainability Traits

Mi Yeon Lee[1], Christopher De Ben[2], Jasmine Ortega[1], Chien-Yuan Lin[1], Americk Eudes[1], Pamela C. Ronald[1,3], Jenny Mortimer[1], Corinne Scown[1], Daniel Putnam[1,2] Jeffery Dahlberg[4] and **Henrik V. Scheller**[1*](hscheller@lbl.gov)

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [2]Department of Plant Science, University of California, Davis, CA 95616; [3]Department of Plant Pathology and the Genome Center, University of California, Davis, CA 95616; [4]Division of Agriculture and Natural Resources, Kearney Agricultural Research and Extension Center, University of California, Parlier, CA 93648

http://jbei.org

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Switchgrass (*Panicum virgatum* L.) is a promising perennial dedicated bioenergy feedstock. It can be grown on marginal lands and produce abundant biomass. Modification of cell wall composition for improved deconstruction is an important strategy for biomass improvement. The JBEI Feedstocks team previously developed engineered switchgrass lines overexpressing rice AT10, an acyltransferase affecting ferulic acid esterification of xylan in biomass, and lines expressing QsuB, a dehydroshikimate dehydratase from *Corynebacterium glutamicum*, which results in low lignin content. These engineering strategies result in increased saccharification efficiency according to our previous studies in switchgrass and other plant species. We now conducted field tests of the engineered plants in two locations. The plants grown in 2018 and 2019 showed changes in biomass composition in general agreement with previous results, although the effects were smaller. In both years we observed that plants expressing QsuB had higher biomass yield than control plants when grown at the Davis, CA, field site.

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.

Genome-scale metabolic rewiring to achieve predictable titers, rates and yields of non-native products at scale

Aindrila Mukhopadhyay*, Thomas Eng[1], Deepanwita Banerjee[1], Andrew Lau[1], Brenda Wang[1], Yan Chen[1], Yuzhong Liu[1], Deepti Tanjore[2], Christopher J Petzold[1], Jay D. Keasling (**Project PI)** [1,3]

[1]Joint Bioenergy Institute, Lawrence Berkeley National Lab, Berkeley; [2]Advanced Biofuel and Bioproducts Demonstration Unit, Lawrence Berkeley National Lab, Berkeley; [3]University of California at Berkeley, Berkeley, CA, USA

http://jbei.org

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

The grand challenge facing synthetic biologists today is understanding how any microorganism can be engineered to produce any desired final product. To meet this challenge, we have developed a new paradigm for host engineering, termed PrOSE (Product Obligatory Strain Engineering). Using genome scale metabolic models, we select a host which has the highest theoretical maximum yield with the added biochemical reactions from a given heterologous multi-gene pathway. Computational models predict gene targets for repression, which are realized using multiplex CRISPR interference (CRISPRi). We demonstrate that PrOSE successfully optimized production of the renewable dye, indigoidine, when produced using the emerging industrial host, *Pseudomonas putida* KT2440. Using PrOSE, production of the desired final product was shifted from stationary phase to exponential phase under optimized conditions, and close to 50% maximum theoretical yield indogidine was realized. In the absence of genome scale models, other systems biology methods can be used to query the solution space. Our results indicate that the careful selection of host/product pair along with computationally guided methods for rational strain engineering is possible. With the advent of facile tools for genetic engineering in nearly any organism, these methods may be generally applicable for any favorable host/product pair.

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 Screening of Lignocellulosic Biomass Degrading Enzymes Utilizing Mass Spectrometry

Nicole Ing[1,2], Noel Ha[1,3 *] (noelha@lbl.gov), Kai Deng[1,2], , Markus de Raad[1,3], Fangchao Song[3], Jose Henrique Pereira[1,3], Martina Aulitto[1,3], Yan Chen[1,3], Jennifer Gin[1,3], Carolina, Barcelos[1,2], Kenneth Sale[1,2], Blake Simmons[1,3] Chris Petzold[1,3] Steve Singer[1,3], Paul Adams[1,3], Anup Singh[1,2], and **Trent Northen**[1,3]

[1]Joint BioEnergy Institute, Technology Division, Emeryville, CA[2] Sandia National Laboratories, Livermore, CA [3] Lawrence Berkeley National Laboratory, Berkeley, CA

http://www.jbei.org/research/technology/high-throughput-biochemistry

Project Goal: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts. This work aims to develop an analytical platform for high-throughput analysis of lignocellulosic biomass deconstruction reactions. This platform combines surface-based mass spectrometry with bioconjugation chemistry to screen enzymatic activity toward lignin, cellulose, and hemicellulose degradation.

Lignocellulosic biomass is the most abundant raw material on the planet and is composed primarily of the cellulose, hemicellulose and lignin. Understanding how to deconstruct this complex material is critical to achieving a cost-competitive replacement for petrochemicals. Enzymes and microbes are capable of degrading this biomass in nature; however, the species, mechanisms, and pathways involved in deconstructing this material are vast and complex. To this end, we are developing highly sensitive nanostructure-initiator mass spectrometry (NIMS)-based assays to screen for novel enzyme activity and gain mechanistic insight into biomass degradation to support our efforts to develop optimal enzyme cocktails at the Joint BioEnergy Institute (JBEI).

Here we describe the development of a library of NIMS substrates representing major lignin linkages, the development of a multiplexed assay to screen for both ligninase and cellulase activity, and the development of a high throughput microfluidic device integrating droplet mixing with NIMS analysis. We have now extended our perfluorinated tagging technique targets for analysis of hemicellulose and cellulases to characterize lignin-modifying enzymes and synthesized a variety of lignin model probes representing the most common lignin linkages. Importantly, unlike existing colorimetric assays, this NIMS-based lignin assay provides information about specific bond-cleavage reactions.

We have combined our lignin and glycan probes to create a multiplexed assay to screen for enzyme activities on mixed. The vast combinations of lignin modifying enzymes and glycoside hydrolases necessitates increasing the throughput and reducing the volumes of our current assays. Hence, we have developed a droplet-based microfluidic device, which combines droplet sample loading and processing onto a NIMS surface. We have now demonstrated array generation, droplet merging, and NIMS analysis and will next extend this to screen the activities of single and mixed enzymes.

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.

Lowering Lignin Recalcitrance and Producing Value Bioproducts in Poplar

Yunjun Zhao[1], Shuncang Zhang[1], Kevin Lin[2], Henrik Scheller[2], Aymerick Eudes[2] and **Chang-Jun Liu**[1]* (cliu@bnl.gov)

[1]Biology Department, Brookhaven National Laboratory, Upton, NY 11973
[2]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Sustaining an economically viable cellulosic biofuel solution requires finely optimized lignocellulosic feedstock to overcome its biomass recalcitrance and to reduce the processing costs. Poplar as a dedicated bioenergy crop offers ample cellulosic resource convertible for biofuels and bio-chemicals. In order to achieve the cost-effective cellulosic biofuel production, as part of JBEI feedstock development program, we employ different synthetic biology strategies to modify the woody biomass of poplar with purpose to lower lignin recalcitrance and to produce value-added specialty chemicals: 1) Stacking monolignol modifying genes to potentially alter lignin structure. The genes encoding monolignol 4-*O*-methyltransferase (MOMT) and monolignol feruloyltransferase (FMT) were co-expressed in poplar. MOMT converts monolignols to the 4-*O*-methylated monomers and FMT transfers ferulate to the monolignols to form ester conjugates. We are investigating whether the combination of two enzymes would lead to the formation and incorporation of 4-*O*-methylated monolignol-ferulate conjugate into lignin polymer, thereby generating the *para*-methylated "dead end" of the polymer, with which to alter lignin structure. 2) Introducing a novel flavonoid tricin biosynthetic pathway in poplar. With recognition of a wide range of monocot grasses to utilize flavone tricin as a natural comonomer with monolignols for cell wall lignification, we are exploring the potential effect of introducing tricin biosynthesis in the dicot poplar on its cell wall lignification. A collection of tricin pathway genes including *CHS, CHI, FNSII (CYP93G1), F3'5'H (CYP75B4)* and *COMT*, driven by the individual xylem specific or preferential promoters, were stacked into an expression vector and transformed into hybrid aspen. 3) Producing industrial platform chemical muconic acid (MA). MA is a dicarboxylic acid used for the production of industrially relevant chemicals such as adipic acid, terephthalic acid, and caprolactam that are widely used in the nylon and thermoplastic polymer industries. The dual expression of plastid-targeted bacterial salicylate hydroxylase (NahG) and catechol 1,2-dioxygenase (CatA) together with plastid-targeted versions of bacterial salicylate synthase (Irp9) and the feedback-resistant 3-deoxy-*D*-arabino-heptulosonate synthase (AroG) were introduced into the hybrid aspen. Biochemical and genetic characterization of the resulted transgenic trees are ongoing.

*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.*

MOMT was previously created with the support of DOE office of Basic Energy Sciences through the Physical Biosciences program of the Chemical Sciences, Geosciences and Biosciences Division (DE-SC0012704 to C.-J.L.)

Machine Learning to Predict Biomass Sorghum Yields under Future Climate Scenarios

Tyler Huntington[1,2], Xinguang Cui[1,2,3], Umakant Mishra[1,2,4]*, **Corinne D. Scown**[1,2,5,6]*

[1]Biosciences Area, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, 94720, USA

[2]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA, 94608, USA

[3]Department of Aerospace Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, Hubei, 430074, China

[4]Environmental Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL, 60439, USA

[5]Energy Technologies Area, 1 Cyclotron Road, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

[6]Energy & Biosciences Institute, University of California, 2151 Berkeley Way, Berkeley, CA, 94704, USA

Project Goals: Predict future bioenergy sorghum yields across the continental United States, accounting for market and farmer behavior shifts, using a machine learning model trained on historical data to identify most promising regions for cultivation.

Crop yield modeling is critical in the design of national strategies for agricultural production, particularly in the context of a changing climate. Forecasting yields of bioenergy crops at fine spatial resolutions can help to evaluate near- and long-term pathways to scaling up bio-based fuel and chemical production, and to understand the impacts of abiotic stressors such as severe droughts and temperature extremes on potential biomass supply. We used a large dataset of 28,364 *Sorghum bicolor* yield samples (uniquely identified by county and year of observation), environmental variables, and multiple approaches to analyze historical trends in sorghum productivity across the U.S. We selected the most accurate machine learning approach (a variation of Random Forest) to predict future trends in sorghum yields under four greenhouse gas emission scenarios and two irrigation regimes. We identified irrigation practices, vapor pressure deficit, and time (a proxy for technological improvement) as the most important predictors of sorghum productivity. Our results showed a decreasing trend of sorghum yields over future years (on average 2.7% from 2018 to 2099), with greater decline under a high greenhouse gas emissions scenario (3.8%) and in the absence of irrigation (4.6%). Geographically, we observed the steepest predicted declines in the Great Lakes (8.2%), Upper Midwest (7.5%), and Heartland (6.7%) regions. Our study demonstrates the use of machine learning to identify environmental controllers of sorghum biomass yield and predict yields with reasonable accuracy. These results can inform the development of more realistic biomass supply projections for bioenergy if sorghum production is scaled up.

Production of Platform Chemicals in Bioenergy Crops: Stacking Low-Recalcitrance Traits with Co-Products

Chien-Yuan Lin,[1] Aymerick Eudes,[1,*] (ageudes@lbl.gov), Khanh Vuu,[1] Edward Baidoo, [1] Bashar Amer,[1] Patrick Shih[1], Chang-Jun Liu,[2] Henrik V. Scheller[1] and **Jay D. Keasling**[1]

[1]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Brookhaven National Laboratory, Upton, NY.

http://www.jbei.org

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Muconic acid (MA) is used for the production of important chemicals such as adipic acid, terephthalic acid, and caprolactam. 2-Pyrone-4,6-dicarboxylic acid (PDC) is a promising building block chemical used to make diverse biodegradable polyesters with novel functionalities. There is no chemical synthesis method currently available for manufacturing PDC, whereas synthesis of MA utilizes petroleum-derived chemicals. Therefore, the development of alternative strategies for bio-based production of MA and PDC has garnered significant interest. Plants represent advantageous hosts for engineered metabolic pathways towards the production of chemicals. We demonstrate that plants can be used for the bio-manufacturing of MA and PDC by re-routing intermediates of the shikimate pathway within chloroplasts. In particular, expression of bacterial 3-dehydroshikimate dehydratase (QsuB) in plastids results in concomitant reductions of lignin and accumulation of protocatechuate (PCA) in biomass. Additional engineering strategies are currently designed to enhance PCA titers and enable its conversion into MA and PDC in-planta. Specifically, bacterial feedback-insensitive 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase was overexpressed to increase carbon flux through the shikimate pathway, co-expression of PCA decarboxylase with catechol 1,2-dioxygenase allowed MA production, and co-expression of PCA 4,5-dioxygenase with 4-carboxy-2-hydroxymuconate-6-semialdehyde dehydrogenase enabled PDC synthesis. The implementation in bioenergy crops (switchgrass, poplar, and sorghum) of MA and PDC biosynthetic routes that divert phenylpropanoid pathway intermediates away from lignin biosynthesis will be presented. These engineering approaches combine in plant biomass the production of value-added chemicals with low-recalcitrance traits towards sustainable development of biorefineries.

This work is part of the DOE Joint BioEnergy Institute 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.

Redirecting metabolic flux via combinatorial multiplex CRISPRi-mediated repression for isopentenol production in *E. coli*

Taek Soon Lee[1,2,*] **(tslee@lbl.gov)**, Tian Tian[1,2], Jinho Kim[1,2], Jing Wei Kang [1,2,3], Aram Kang[1,2], Jay D. Keasling **(Project PI)**

[1]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA 94608, USA.
[2]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
[3]Department of Chemical Engineering, University of California, Berkeley, California, USA.

http://www.jbei.org/research/divisions/biofuels-and-bioproducts/pathway-metabolic-engineering/

Project Goals: We aim to establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts at JBEI. Re-directing carbon flux to the targeted pathway is an important approach to improve production titer, rate, and yield. CRISPR interference (CRISPRi) is a novel approach that can be used to knock down endogenous genes in competing pathways. In this study, we constructed a CRISPRi-mediated multiplex repression system to silence transcription of several endogenous genes in order to increase precursor availability in a heterologous isopentenol biosynthesis pathway. The result shows that multiplex combinatorial knockdown of competing genes using CRISPRi can increase production of target metabolite, while the repression level needs to be adjusted to balance the metabolic network and to achieve the maximum titer improvement.

CRISPR interference (CRISPRi) via target guide RNA (gRNA) arrays and a deactivated Cas9 (dCas9) protein has been shown to simultaneously repress expression of multiple genomic DNA loci. By knocking down endogenous genes in competing pathways, CRISPRi technology can be utilized to re-direct metabolic flux toward target metabolite [1]. In this study, we constructed a CRISPRi-mediated multiplex repression system to silence transcription of several endogenous genes in order to increase precursor availability in a heterologous biosynthesis pathway for isopentenol which is a drop-in biofuel and a precursor for commodity chemicals [2]. To identify genomic knockdown targets in competing pathways, we first designed a single-gRNA library with 15 individual targets, where 3 gRNA cassettes targeting gene *asnA*, *prpE*, *gldA* increased isopentenol titer by 18-24%. We then combined the 3 single-gRNA cassettes into two- or three-gRNA array and observed up to 98% enhancement in production by fine-tuning the repression level through titrating dCas9 expression [1]. Our strategy shows that multiplex combinatorial knockdown of competing genes using CRISPRi can increase production of target metabolite. In this approach, we also showed that the repression level needs to be adjusted to balance the metabolic network and achieve the maximum titer improvement.

References

1. Tian, T., Kang, J.W., Kang, A., **Lee, T.S.**, *ACS Synth Biol.* 2019, (doi: 10.1021/acssynbio.8b00429)

2. George, K.W., Chen, A., Jain, A., Batth, T.S., Baidoo, E., Wang, G., Adams, P.D., Petzold, C.J., Keasling, J.D., **Lee, T.S.**, *Biotech. Bioeng.* 2014, 111, 1648-1658 (doi:10.1002/bit.25226)

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.

Sorghum Secondary Cell Wall Nanoarchitecture Can Be Revealed By Solid State NMR

[1]Yu Gao, [2]Andrew Lipton, [3]Yuuki Witmer, [2]Nancy Washton, [1,4]Henrik Scheller, [3]Dylan Murray, [1,4]Jay Keasling, **[1]Jenny Mortimer*** (jcmortimer@lbl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]EMSL, Pacific Northwest National Lab, Richland; [3]University of California, Davis; [4]University of California, Berkeley;

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Lignocellulosic biomass (i.e. the plant secondary cell wall) is a promising sustainable feedstock for the production of biofuels and bioproducts. The cell wall is a complex material composed of lignin, diverse polysaccharides, and proteins that assembles outside of the cell membrane. Our goal is to develop a physical model of bioenergy crop cell walls which will guide engineering and breeding strategies.

Understanding the biosynthesis and nanoarchitecture of plant cell walls is key for the predictable engineering of bioenergy feedstocks. Multi-dimensional solid-state NMR (ssNMR) facilitates the close investigation of the native nanoarchitecture of plant cell walls[(1–4)]. By successfully generating mature plants with high ^{13}C incorporation (~90 %) in a customized growth chamber, the effect of modifying the cell wall on cell wall polysaccharide arrangement was investigated. Here, we present data on the nanoarchitecture of sorghum stem, root and leaf cell walls. We explore the effects of sample preparation (dried material vs. fresh), and propose a model for xylan-cellulose interactions. Future work will include testing the model using bioenergy crops with engineered cell walls.

References

1. *Wang T, Phyo P, Hong M. Multidimensional solid-state NMR spectroscopy of plant cell walls. Solid State Nucl Magn Reson. 2016 Aug 13;78:56–63.*

2. *Simmons TJ*, Mortimer JC*, Bernardinelli OD, Pöppler A-C, Brown SP, deAzevedo ER, et al. Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. Nat Commun. 2016 Dec 21;7:13902.*

3. *Dupree R, Simmons TJ, Mortimer JC, Patel D, Iuga D, Brown SP, et al. Probing the molecular architecture of Arabidopsis thaliana secondary cell walls using two- and three-dimensional (13)C solid state nuclear magnetic resonance spectroscopy. Biochemistry. 2015 Apr 14;54(14):2335–2345.*

4. *Kang X, Kirui A, Dickwella Widanage MC, Mentink-Vigier F, Cosgrove DJ, Wang T. Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR. Nat Commun. 2019 Jan 21;10(1):347.*

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. This work was also supported by the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at PNNL.

Towards Whole Biomass Utilization: Development of Ionic Liquid Technologies for Lignin

Seema Singh[1,2*] (ssingh@lbl.gov), Mood Mohan[1,2], Lalitendu Das[1,2], Hemant Choudhary[1,2], Alexander Landera[2], Anthe George[1,2], Blake A. Simmons[1,3], John Gladden[1,2], **Jay Keasling[1,3,4]**

[1]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, California 94608, United States

[2]Biological and Engineering Science Center, Sandia National Laboratories, 7011 East Avenue, Livermore, California 94551, United States

[3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, United States

[4]Department of Chemical & Biomolecular Engineering, Department of Bioengineering, University of California, Berkeley, Berkeley, California 94720

Project Goals: Establish the scientific knowledge and new technologies to transform the maximum amount of carbon available in bioenergy crops into biofuels and bioproducts.

Abstract: Lignin is a multifunctional polymer and integral part of the plant cell wall. Lignin's full potential as a renewable source of aromatic compounds can be, in part, unlocked only if an efficient and economic method for lignin depolymerization and valorization is developed. Alternative solvents such as ionic liquids (ILs) and deep eutectic solvents (DESs) have received increasing interest because of their high efficacy in fractionating and pretreating lignocellulosic biomass. However, the lignin-carbohydrate complex degradation mechanism in DES, especially metal containing DESs (mDESs), computational predictability of eutectic points, and mDESs' characteristics are not well understood. This study aims to i) develop and understand the atomistic behavior of known/newly designed ILs on lignin via multiscale simulation approaches, ii) predict the eutectic points for the mDES and their impact on lignin depolymerization, iii) experimental validation of predicted results for biomass fractionation and process parameter optimization, and iv) characterize and upgrade the lignin streams from ILs/mDESs treated biomass to value-added compounds. In this study, a quantum chemical-based molecular simulation namely COSMO-RS (*CO*nductor like *S*creening *MO*del for *R*eal *S*olvents) model was used to screen rational combinations of ILs (60 anions and 90 cations) for the solubility of lignin. The activity coefficient and excess enthalpy of IL and lignin mixtures were evaluated as reference property to describe the affinity of lignin for different ILs. Furthermore, based on the COSMO-RS results, the selected ILs were visualized by observing their structural properties and dynamics with lignin by performing the molecular dynamics (MD) simulations. Subsequently, we demonstrate *simultaneous fractionation of biomass and lignin depolymerization* using mDES. We studied the product profile of depolymerized biopolymers (qualitatively and quantitatively) along with the molecular weight distribution profile as a function of reaction coordinates. Notably, the employed DES system not only achieved ~95% glucose and ~25% lignin monomer yields, but also had limited enzyme inhibition and microbial toxicity (>5wt%), opening the possibility for a subsequent downstream

biological conversion of the depolymerized stream(s). This study provides a mechanistic understanding of biomass fractionation and lignin depolymerization in mDES and explores the potential of catalytic upgrading of lignin to value-added products.

Funding Acknowledgement:

"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. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes."

Deciphering N-fixing symbiosis signaling in *Medicago* with dynamic regulatory module networks (DRMNs)

Daniel Conde,[2] Sara Knaack[1,]* (saknaack@wisc.edu), 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: Our research goal is to identify genomic elements required for the symbiotic relationship between nodulating plants and nitrogen (N)-fixing bacteria. In order to identify these elements, we measured and analyzed novel RNAseq and ATACseq time course data obtained from Medicago plants subjected to treatment with lipo-chitooligosaccharide (LCO) Nod factors. The gene regulatory network involved in the response to LCOs were discovered by applying a novel computational method that defines dynamically transitioning genes and predicts key regulators of these genes. Prioritized regulators and their target genes will be validated experimentally.

Nitrogen fixation occurs naturally in a limited lineage of plants (N-fixing clade) through a symbiotic relationship where plant root nodules are colonized by N-fixing bacteria. The vast majority of plants do not have this ability, requiring the application of fertilizers for agricultural production. Some molecular components of the symbiosis pathways are known, yet the gene regulatory network involved in this process is not well understood. A detailed characterization of the regulatory network controlling symbiosis can provide important insights into the key regulators that are missing from non N-fixing species. More importantly, this can prioritize the elements and regulators that could be engineered into plants lacking the ability to form symbiotic relationships.

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 part of the early signaling processes needed to trigger the establishment of symbiosis in *Medicago*.

We analyzed the expression data using a time-series clustering algorithm (Escarole[1]) and identified groups of genes exhibiting coherent dynamic change over time, several of which are associated with coordinated changes in accessibility. In parallel, we examined the signal profile of peak calls from the ATAC-seq data collected over the same time points, and identified 81,114 putative regulatory regions, including clusters of those regions that exhibit coherent dynamic change over time. These dynamically changing peaks are proximal to 39,755 genes (of 51,316 in the v1.6 annotation of the v5 *Medicago* genome). Furthermore, the accessibility profile around the gene promoter is correlated (Pearson correlation of 0.50 or greater) to transcriptomic changes detected over the duration of the experiment in 28.5% of genes.

We next applied a novel approach, Dynamic Regulatory Module Networks (DRMNs), for integrating RNA-seq and ATAC-seq time courses along with known sequence-specific motif instances. With this approach we identified modules of co-expressed genes at each time point, in addition to per-module regulatory networks predictive of gene expression within each module. The network edges are based on predicting gene expression using gene promoter and motif site accessibility values. We examined known regulators of the nodulation pathway and found them to be enriched within specific groups of genes exhibiting transitions in their expression state. Several of our modules are enriched for root hair elongation, defense response to bacterium, chromatin organization and MAPK cascade processes. We identified transitioning genes using DRMN results and predicted regulators for these genes using a network inference algorithm[2]. The regulators include IBM1, EDN3, MTF1, EIN3, SHY2, BHLH, AB15, RRB15, MTF1, NSP1, RRB9, AB14, and SMB. Several of these regulators (EIN3, NSP1) are known to be involved in nitrogen fixation providing early promising support to our predictions. We are currently identifying targets for these regulators and experimentally testing them.

Taken together our dataset and associated analyses provide a valuable resource to the plant community to understand the gene regulatory programs controlling Nitrogen fixation.

References

1. Inferring Regulatory Programs Governing Region Specificity of Neuroepithelial Stem Cells during Early Hindbrain and Spinal Cord Development. Deborah Chasman, Nisha Iyer, Alireza Fotuhi Siahpirani, Maria Estevez Silva, Ethan Lippmann, Brian McIntosh, Mitchell D. Probasco, Peng Jiang, Ron Stewart, James A.Thomson, Randolph S. Ashton, Sushmita Roy. Cell Systems, 2019.

2. Integrated Module and Gene-Specific Regulatory Inference Implicates Upstream Signaling Networks. Sushmita Roy, Stephen Lagree, Zhonggang Hou, James A. Thomson, Ron Stewart, Audrey P. Gasch. PLoS Comput Biol, 2013.

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).

Evolution of root nodule symbiosis & engineering of symbiotic nitrogen fixation in *Populus* sp.

Thomas B. Irving[1*] (tbirving@wisc.edu), Lucas Gontijo Silva Maia[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. Genetic studies in model legumes such as *Medicago truncatula* and *Lotus japonicus* identified that (1) rhizobia colonize legume roots intracellularly through the recruitment of the arbuscular mycorrhizal (AM) signaling pathway and (2) root nodules evolved from the recruitment of the lateral root developmental pathway. Essential mechanisms connecting these two processes are cytokinin signaling and the NIN (Nodule INception) protein.

Recent comparative phylogenomic studies suggest that these root nodule symbioses appeared once in the last common ancestor of the Fabales, Fagales, Cucurbitales, and Rosales, and has been lost multiple times within this monophyletic group often called the "nitrogen-fixing clade"[1]. *Populus* sp. are bioenergy crops and close relatives to the "nitrogen-fixing clade". In particular, they seem to possess all the 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 identified rhizobia and their diffusible signals that can activate the AM signaling pathway in *Populus* sp., as they do in legumes. 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 the *NIN* genes in *Populus* sp. We overexpressed and knocked-down the *NIN* genes and cytokinin receptors in *Populus* sp. Overexpression of some *NIN* genes and cytokinin receptors led to an increase in lateral root development and, in some cases, the development of nodule-like structures on *Populus* sp. roots. Interestingly, the development of these root lateral structures is activated by cytokinins, which is a critical feature distinguishing root nodules from lateral roots. We are also characterizing the *cis*-elements targets of NIN and other symbiotic transcription factors using ATAC-seq and DAP-seq. We are actively working on trying to get these nodule-like structures colonized by rhizobia.

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.

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

Funding statement

This project is funded by the U.S. Department of Energy grant #DE-SC0018247

Global-Scale Phylogenomics of the Nitrogen-Fixing Clade

Heather Rose Kates*[1] (hkates@ufl.edu), Jean-Michel Ane[2], Kelly Balmant[3], Daniel Conde[3], Matthew Crook[4], Christopher Dervinis[3], Robert P. Guralnick[1], Thomas Irving[2], **Matias Kirst**[3], Sara Knaack[2], Lucas Maia[2], Sushmita Roy[2], Ryan Folk[1], Douglas E. Soltis[1], Pamela S. Soltis[1]

[1]University of Florida, Florida Museum of Natural History, Gainesville, FL; [2]University of Wisconsin, Madison, WI; [3]University of Florida, Gainesville, FL; [4]Weber State University, Ogden, UT

www.nitfix.org

Project Goals: The collaborative project "Phylogenomic discovery and engineering of nitrogen fixation into the bioenergy woody crop poplar" is focused on identifying the genomic novelties that enable the symbiotic relationship between nodulating plants and N-fixing bacteria to support genetically engineering this capability into bioenergy crops. The first aim of this project is a comparative phylogenomic study of the nodulating clade to uncover the genomic novelties that were required for the evolution of the root nodule symbioses. A massively improved phylogenetic framework for the N-fixing clade that includes genetic data for nodulation genes for all species in the phylogeny will provide a robust, revised understanding of the exact ancestral origin of nodulation and the evolution of the predisposition to nodulate. This framework will also inform a experiments comparing close relatives that nodulate or do not nodulate to identify the genes underlying this trait. To achieve these aims we developed a cutting-edge phylogenetic approach with an unprecedented sampling effort that is described below.

Robust phylogenetic inferences on the origin of the predisposition to nodulation, and events of nodulation gain and loss, are key to understanding the evolutionary lability and, consequently, the likelihood of successful transferability of N-fixing symbioses among lineages of angiosperms. A well-resolved and well-sampled N-fixing clade phylogeny is therefore a prerequisite to the discovery of genes that determine nodule development. Multiple phylogenetic analyses have been conducted on the N-fixing clade with the aim of elucidating the origins of N-fixing symbioses; however, these analyses have relied on trees estimated using a few genes and in which species sampling in the N-fixing clade was limited. We present our first steps toward a revised phylogeny based on deliberate and extensive sampling, phylogenomic data, and rigorous statistical analysis. These will allow more accurate inference of precursors of N-fixing symbioses, gain and/or loss events, and potential transferability of the capability to crop plants not in the N-fixing clade. The final phylogeny will include 15,000 of 30,000 species of the N-fixing clade, making it the best sampled and phylogenetically studied major clade of plants.

This ambitious phylogenetic study requires novel strategies for rapid specimen sampling and DNA data generation. Elements of our strategy that enable geographically and taxonomically comprehensive sampling include a protocol for rapid tissue sampling of 15,000 historical specimens, a high-throughput, high-yield DNA extraction protocol specifically suited to degraded DNA, a targeted-enrichment DNA sequencing kit that works across phylogenetic

scales and includes key functional genes, and scalable information management and processing using a project database and custom bioinformatics tools. The protocols and tools developed for this project can be leveraged as a toolkit for phylogenetic researchers to generate very large phylogenetic datasets and will facilitate greater utilization of historical specimens in phylogenetic research.

The results so far include our completion of rapid sampling, data generation, and data assembly and highlight preliminary phylogenetic results and their usefulness for meeting project goals. Our preliminary phylogeny is based on 100 nuclear loci sequenced for 9,876 species from the N-fixing clade (Fig. 1) and informs comparisons between nodulating and non-nodulating species based on novel phylogenetic relationships. We also highlight a set of functional genes that may be correlated with nodulation based on our preliminary results.

This result represents an initial analysis comprising two thirds of our genomic and taxonomic sampling effort. We will leverage the total phylogenetic and comparative genomic results to discover gene candidates that potentially underlie nodule development so that these genes can be tested for function in nodulating and non-nodulating model systems.

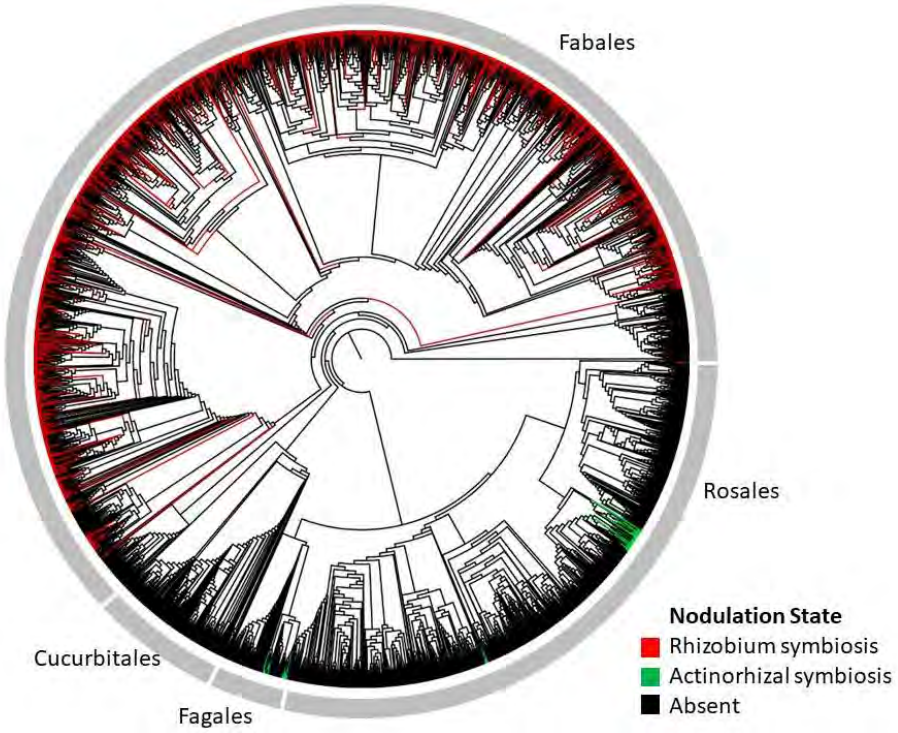


Fig. 1. Preliminary phylogeny of a subset of the N-fixing clade based on our gene capture data available at the time of the abstract submission. This tree represents two thirds of the complete phylogeny that will be available for meeting project goals. Terminal branches are colored by presence or absence of nodulation based on scorings from Werner et al. (2014) and Afkhami et al. (2018).

This work is supported by the Biosystems Design Program from the Biological and Environmental Research (BER) Office of Science at the U.S. Department of Energy (grant #DE-SC0018247).

Title: Lessons from the Field: How Sorghum and Its Microbiome Respond to Drought

Peggy G. Lemaux[1*] **(lemauxpg@berkeley.edu)**, Jeffery Dahlberg,[2] Devin Coleman-Derr,[1,3] Robert Hutmacher,[4] Christer Jansson,[5] Ronan O'Malley,[6] Elizabeth Purdom,[1] John Taylor,[1] Axel Visel[6] and John Vogel[1,6]

[1] University of California, Berkeley CA; [2]University of California, Kearney Ag Research & Extension Center, Parlier CA; [3]USDA-ARS Plant Gene Expression Center, Albany CA; [4]University of California, West Side Research & Extension Center, Five Points CA; [5]EMSL-Pacific Northwest National Laboratory, Richland WA; and [6]DOE Joint Genome Institute, Berkeley CA

https://njp-spin.jgi.doe.gov/epicon/

Project Goals:

Analysis of transcriptomic and epigenetic control mechanisms during spatiotemporal responses to water-limiting conditions is being performed on leaves and roots of field-grown *Sorghum bicolor* (L.) Moench, under control and pre- and post-flowering drought. Also investigated are changes in associated bacterial and fungal communities in bulk soil, rhizosphere, leaves and roots of drought-stressed sorghum. Goals of these efforts are to understand in sorghum mechanisms functioning in acclimation to and recovery from pre- and post-flowering drought, using RNA-Seq, BS-Seq, proteomics, metabolomics, and histone profiling. Additional insights into sorghum's responses are gleaned from studying impacts on microbial populations, using metagenomics, metatranscriptomics, and metabolomics. Cumulative data are being used to devise models to better predict and control roles and interactions of transcriptional regulation and of the microbiome in sorghum's response to drought. We hope to identify genes, molecular markers and microbes to develop strategies for improving drought tolerance in sorghum and other crops.

EPICON researchers aim to better understand impacts of drought on crops, due to increased frequency and severity of that stress with climate change. Both transcriptomic and epigenetic changes play major roles in regulating reactions to drought. To better understand spatiotemporal responses, samples of leaves and roots of field-grown *Sorghum bicolor* (L.) Moench were collected weekly over the growing season. RNA-Seq, BS-Seq, metabolomics, proteomics, and histone profiling were used to gain a mechanistic understanding of plant acclimation to and recovery from drought. Shotgun metagenomics, metatranscriptomics, and metabolomics were used to monitor sorghum's rhizosphere bacterial and fungal microbiome communities.

As a widely cultivated, drought- and flood-tolerant cereal, sorghum is a flexible bioenergy feedstock with a relatively reduced environmental footprint. EPICON studies were done in fields in California's Central Valley, where rare summer rainfall permitted controlled drought conditions during the four years of field trials. One pre-flowering drought-tolerant and one stay-green, post-flowering drought-tolerant variety were planted in a replicated, split plot design, with normal watering and pre- and post-flowering drought. Phenotypic observations were taken during the growing season, grain and biomass yields at harvest. Most impacted phenotypes during pre-flowering drought were later flowering, shorter stature and lower forage/grain yields; for post-flowering drought premature plant death, lodging and reduction in seed size were seen.

Year one transcriptional profiling of triplicate, weekly leaf and root samples revealed widespread adaptations at all developmental stages, as well as after watering pre-flowering droughted plants and after imposing post-flowering drought – with 44% of expressed genes significantly affected (Varoquaux et al., 2019). Based on 350 transcriptomes, fast, temporal transcriptomic responses were seen in leaves and roots, including changes in well-known drought pathways. Roots had greater transcriptional disruptions than leaves. Pre-flowering drought had more complex temporal changes than post-flowering drought; large differences were found between genotypes. Efforts now focus on detailed analyses of Year 2 and 3 data.

In-depth studies were done on two drought-related transcriptional responses. (1) Qualitative differences in transcriptomic and proteomic data were seen in responses of photosynthesis-related genes to pre- and post-flowering drought, suggesting mechanisms of (*i*) enhanced photoprotection in pre-flowering drought, (*ii*) changes to photosynthate partitioning, specific to pre- and post-flowering drought, and (*iii*) altered progression of leaf senescence in post-flowering drought. Quantification of photosynthetic traits in the field confirmed genotypic and drought-induced differences in photosynthetic performance, predicted based on transcriptomic and proteomic data. (2) Large-scale reduction in expression of sorghum genes critical to arbuscular mycorrhizal (AM) fungi symbiosis occurred in both pre- and post-flowering drought and co-occurred with a drop in AM fungal mass, suggesting drought leads to loss of AM fungi and their vital symbiotic interactions. Upon re-watering, plant gene expression and AM fungal mass increased to control levels. Gene expression differences were the largest genotype-specific drought responses for a single functional gene category, indicating AM symbiosis may explain some genotype differences during pre-flowering drought recovery (Varoquaux et al 2019).

Preliminary analysis of BS-Seq data, designed to explore DNA methylation patterns, revealed many regions in leaves where changes correlated with plant development. Also, strong expression from transposable elements occurred under drought, often continuing after water resumption. Further study awaits high accuracy genome sequencing and annotation of the two genotypes. Preliminary data from year 1 leaf samples suggest terminal clipping of histones H4 and H3 may regulate plant growth and drought tolerance differently in the two genotypes (Zhou

et al. 2019). LC-MS analysis of year 2 intact histones from leaves will enable discovery of novel drought- or development-related posttranslational modifications.

Distinct, differential protein profile changes between genotypes suggest protein level changes may relate to differential drought tolerances, especially in pre-flowering drought. Comparisons of protein ratios revealed functional categories distinct to each variety. During pre-flowering drought, identity of leaf proteins, significantly increased in the pre-flowering, drought-tolerant variety versus the post-flowering, drought-tolerant variety, suggest chaperones may play key roles in stabilizing key protein or enzyme functions involved in drought tolerance in the pre-flowering, drought-tolerant variety. Conversely, the post-flowering, drought-tolerant variety showed significantly higher enrichment in beta-glucoside metabolism and photosynthesis-related proteins. Regarding metabolites, increases in proline, glyceraldehyde, shikimic acid, sugar pentoses and hexoses, were seen during pre-flowering drought. During post-flowering drought, increases in 3-hydroxycinnamic acid, mevalonic acid, quinic acid and sugar monosaccharides were seen in leaves of both varieties. Significant increases in proline, glycine, betaine, arginine, lysine and serine were seen in roots of the post-flowering, drought-tolerant variety.

Using soil, root, leaf and rhizosphere samples, collected weekly from the same plants, dramatic shifts in bacteria and fungi followed drought and re-watering (Xu et al. 2019). Genome-resolved metagenomics within rhizosphere and soil samples allowed binning of shotgun metagenomic and metatranscriptomics reads into the most abundant organisms. Results show that rapid changes in bacterial community composition, occurring after pre-flowering drought, correlate with shifts in transcriptional activity of specific biological functions, i.e., genes related to carbohydrate and amino acid transport and metabolism. Also, organisms, enriched under drought stress, contain a larger repertoire of genes in the same categories. These data, coupled with root metabolomics, suggest interplay between plant metabolism and bacterial community activity that is in part sensitive to shifts, following drought stress, in nutrient profiles within the root and rhizosphere.

Total fungal diversity and community composition are significantly affected by both pre- and post-flowering drought (Gao et al 2020). Importantly, abundance of the fungal plant pathogens, *i.e., Fusarium*, *Gibberella* and *Sarocladium*, decreases in pre-flowering drought but increases in post-flowering drought, more strongly in rhizosphere than root (Gao et al 2020). AM fungi behavior was discussed above (Gao et al 2018, Varoquaux et al 2019). Expression of sorghum genes involved in communication with AM fungi (both strigolactones and Myc-LCOs) were also found to correlate with abundance of early or late AM fungal species. Lastly, the cross-kingdom network of fungal-bacterial co-occurrence is disrupted by drought, recovering after re-watering.

Scale and scope of EPICON data provide unprecedented platforms for in-depth exploration of molecular mechanisms of drought tolerance and its interplay with the plants' biotic environment. Data generated provide many avenues for future research on sorghum and drought – likely relevant to other crops. Ultimately, genes, molecular markers and microbes causally associated with drought tolerance will be identified that improve yield and fitness under drought.

References:

Xu et al., 2018. *Proc Natl Acad Sci USA.* www.pnas.org/cgi/doi/10.1073/pnas.1717308115

Gao et al., 2018. *Int Soc Micro Ecol J* https://doi.org/10.1038/s41396-018-0264-0

Varoquaux et al., 2019. *Proc Natl Acad Sci USA* https://doi.org/10.1073/pnas.1907500116

Gao et al., 2020. *Nat Commun.* 11:34 https://www.nature.com/articles/s41467-019-13913-9

Zhou et al., 2019. *Methods* 75_R1 https://data.mendeley.com/datasets/9j232f653t/1

Funding: DOE Office of Biological and Environmental Research (DE-SC0014081).

Title: Arctic Microbial Permafrost Degradation

Karen Lloyd (klloyd@utk.edu)[1], Katie Sipes[1], Tatiana Vishnivetskaya[1], Andrew Steen[1], Tullis C. Onstott[2], Renxing Liang[2], Zachary Garvin[2], Robert Hettich[3], Richard Giannone[3], John Cliff[4], James Bradley[5], Julia Boike[6], Sebastian Wetterich[6], Andrey Abramov[7], Elizaveta Rivkina[7]

[1]University of Tennessee, Knoxville, TN
[2]Princeton University, Princeton, NJ
[3]Oak Ridge National Laboratory, Oak Ridge, TN
[4]Environmental Molecular Sciences Laboratory, Richland, WA
[5]Queen Mary University of London, London, UK
[6]Alfred Wegener Institute, Potsdam, Germany
[7]Institute of Physicochemical and Biological Problems in Soil Science, Pushchino, Russia

Project Goals:
Permafrost is one of Earth's largest reservoirs of soil organic carbon (SOC). As permafrost thaws, heterotrophic microbial communities degrade the newly-available SOC, often resulting in large and variable fluxes of the greenhouse gases CO_2, CH_4, and possibly N_2O to the atmosphere. Our project will address some questions fundamental to understanding how permafrost thaw affects greenhouse gas emissions. What are the relative contributions of CO_2-consuming vs. net-CO_2-producing processes in thawing permafrost? As SOC degrades, what are the relative production or consumption rates of CO_2, CH_4, and N_2O, which metabolic pathways and microbes drive SOC degradation in thawing permafrost, how are the metabolic pathways affected by SOC composition, and how much of the SOC is degraded? We will address these questions with intact core incubations from Bayelva, Svalbard. This site of continuous permafrost has been warming and thaw depth has increased from ~1 m to ~2 m over the last 20 years.

To determine how permafrost thaw impacts SOC-degrading microbial communities in intact core incubations, we will collect two long cores that go through the current permafrost layer, subsample one of them for DNA and geochemical analyses while in Ny Ålesund, and send the other in coolers to Potsdam, Germany. We will collect nine 2 m cores for transport to Princeton, NJ, where we will monitor pore gas composition and surface fluxes, pore water chemistry and isotopic compositions of dissolved inorganic carbon, dissolved organic carbon, CO_2, CH_4, and N_2O. We will take subsamples for meta-omic analyses at times zero, 1 week after permafrost thaw, after core slumping, and after 12 and 18 months, from four depths in the cores in order to obtain high-resolution molecular information about the microbial functional activities and temporal changes.

Funding from US Department of Energy, Office of Biological and Environmental Research (DE-SC0020369).

Increasing photosynthetic efficiency of energycane under fluctuating lights

Moonsub Lee[1] (mlee128@illinois.edu), Fredy Altpeter[2], and **Donald Ort[1]**
[1]University of Illinois, Urbana-Champaign and [2] University of Florida, Gainesville
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 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

Engineering bioenergy crops to produce natural oils provides an alternative energy source that can help ensure energy security while mitigating environmental problems associated with traditional fossil fuels. The C_4 photosynthetic pathway of bioenergy grasses such as sugarcane, while the most efficient pathway, needs to be improved in order to accommodate the high energetic costs of oil biosynthesis without decreasing plant growth. Crop canopy modeling demonstrates that the rapidly changing light environment experienced by bioenergy crops in dense plantings limits photosynthetic efficiency. However, there has been little empirical research investigating the impact of rapidly changing light environments on C_4 bioenergy crops biomass production or on developing engineering strategies to improve the efficiency of bioenergy grasses in dynamic light environments. Previous modeling of C_4 photosynthesis in fluctuating light environments suggested that photosynthetic efficiency is reliant on the coordination of the C_4 and C_3 metabolic cycles. We hypothesize that large metabolite pools act as buffers to minimize changes to metabolite fluxes in rapid changes in light environments, and that this buffering helps maintain coordination of the C_4

and C_3 metabolic cycles. Increased chloroplast volume would increase this metabolite buffering capacity by increasing metabolite pool sizes, and thereby enhance the C4 photosynthetic efficiency in fluctuating light conditions. We are engineering increased sugarcane chloroplast volume by genetically manipulating various components of the chloroplast division machinery. We have already developed a methodology for estimating chloroplast volume of sugarcane leaves using confocal microscopy and a 3-D image program. Leaf microscopy, leaf photosynthetic gas exchange, and above ground biomass production in greenhouse trials will be used to select the best chloroplast modifications for future field trails.

References

1. 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.
2. 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.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Award Number DE-SC-0018254).

Stem Parenchyma Cell-specific Gene Characterization in Energycane

Jiang Wang[1] (jwang846@illinois.edu), Ya Chi Yu[1], and **Li-Qing Chen[1]**

[1]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. This work investigates mature stem parenchyma cell-specific genes with the aim to increase oil production in the stem.

Abstract

Triacylglycerols (TAG) are major components of plant oil. Engineering energycane to produce an abundance of TAG for further conversion into biodiesel, biojet fuel, and bioproducts, is an essential goal of ROGUE. Constitutively engineering TAG biosynthesis throughout the plant may produce pleiotropic effects, therefore manipulations should be restricted in a specific tissue or cell type, such as mature stem parenchyma cells, where tremendously stored photoassimilates are available for efficient TAG conversion. However, the information of mature stem parenchyma cell-specific or preferential genes still lacks up to now. Several stem highly expressed candidates were selected from the published sugarcane RNA-seq dataset. The real-time PCR analysis was conducted to confirm the candidate gene expression levels across various tissues at two developmental stages (immature and mature) in energycane. Spatial expressions of candidate genes in the stem were examined using optimized RNA *in situ* hybridization. Two candidate genes, namely SPCG1 (Stem Parenchyma Cell-specific Gene 1), and SPCG2, were strongly expressed in the mature stem of energycane. Additionally, abundant SPCG2 RNA transcripts were found in the pith parenchyma cells of the mature stem using two different gene-specific RNA probes. Further validating the promoter activities of SPCG1 and SPCG2 in energycane is needed for potential applications of engineering energycane.

References

1. Zhang, J., Zhang, X., Tang, H. et al. Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum L.. Nat Genet 50, 1565–1573 (2018). https://doi.org/10.1038/s41588-018-0237-2

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Award Number DE-SC-0018254).

Toward transgenic sustainable productivity increases in *Miscanthus giganteus*

Karolina Sobańska[1, 2] *, (sobanska@illinois.edu), Kher Xing Chan[1]* (cindyckx@illinois.edu), Steven J. Burgess[1] , Benjamin Haas[1], Joanna Cerazy-Waliszewska[2], Cameron Kern [1], Piotr Jedryszek[3], Tomasz Pniewski[2], Fredy Altpeter[4] and **Stephen P. Long[1]**

[1]University of Illinois at Urbana-Champaign, Urbana, Illinois, USA; [2]Polish Academy of Sciences, Poznań, Poland; [3]University of Oxford, Oxford, UK; [4]University of Florida, Gainesville, Florida, USA

https://rogue.illinois.edu/

Project Goals:

The Renewable Oil Generated with Ultra-productive Energycane (ROGUE) project aims to engineer the two most productive American biofuel 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

Miscanthus × *giganteus* is a more chilling-tolerant C_4 biomass feedstock in comparison to other phylogenetically-related C_4 crops such as maize, sorghum or sugarcane (1, 2). Photosynthetic activity of C_4 crops is limited by the amount of pyruvate orthophosphate dikinase (PPDK) and rubisco which restrain regenerate of phospho*enol*pyruvate (PEP) (1, 3, 4). At the same time, photosynthetic efficiency is shown to improve under fluctuating light when photoprotection response time is accelerated by overexpression of zeaxanthin epoxidase (ZEP), violaxanthin de-epoxidase (VDE) and Photosystem II subunit S (PsbS) (5). We hypothesize that alleviating rate limitation in C_4 photosynthesis by PPDK and accelerating relaxation of photoprotection will significantly raise photosynthetic efficiency in *Miscanthus*. Although *M.* × *giganteus* fits the characteristics of an ideal bioenergy crop with the added advantage of minimal invasive potential, the propagation of this highly productive feedstock is limited by its triploid genome and the sterility of the plant (6). Traditionally, biolistic transformation of *Miscanthus* uses embryogenic calli induced from immature influorescences as the main transformation material (7) which can only be collected once a year. In this study, we established a *M.* × *giganteus* transformation system at the University of Illinois at Urbana-Champaign using microparticle bombardment (8) and demonstrated this transformation method using vectors encoding genes related photoprotection in plants and PPDK, respectively. In order to increase the availability of material for transformation, we are also developing a system to induce embryogenic callus from shoot apices of *M.* × *giganteus*. With these in place, we hope to obtain a more robust system to study the effect of photosynthetic genes in transgenic *M.* × *giganteus*.

References

1. Long SP, Spence AK (2013) Toward Cool C4 Crops. *Annu Rev Plant Biol* 64(1):701–722.
2. Fonteyne S, et al. (2018) Physiological basis of chilling tolerance and early-season growth in Miscanthus. *Ann Bot* 121(2):281–295.
3. Naidu SL, Moose SP, Al-Shoaibi AK, Raines CA, Long SP (2003) Cold Tolerance of C4 photosynthesis in Miscanthus x giganteus: Adaptation in Amounts and Sequence of C4 Photosynthetic Enzymes. *Plant Physiol* 132(July):1688–1697.
4. Wang D, Naidu SL, Portis AR, Moose SP, Long SP (2008) Can the cold tolerance of C4photosynthesis in Miscanthus x giganteus relative to Zea mays be explained by differences in activities and thermal properties of Rubisco? *J Exp Bot* 59(7):1779–1787.
5. Kromdijk J, et al. (2016) Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science (80-)* 354(6314):857–862.
6. Boersma NN, Heaton EA (2014) Propagation method affects Miscanthus×giganteus developmental morphology. *Ind Crops Prod* 57(M):59–68.
7. Ślusarkiewicz-Jarzina A, et al. (2017) Effective and simple in vitro regeneration system of Miscanthus sinensis, M. × giganteus and M. sacchariflorus for planting and biotechnology purposes. *Biomass and Bioenergy* 107:219–226.
8. Sobańska K, et al. (2019) Optimised expression cassettes of hpt marker gene for biolistic transformation of Miscanthus sacchariflorus. *Biomass and Bioenergy* 127:105255.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Award Number DE-SC-0018254).

Towards Oil Cane: Engineering Energycane for Hyperaccumulation of Lipids and Improved Agronomic Performance

Guangbin Luo[1] (g.luo@ufl.edu), Baskaran Kannan[1], Viet Dang Cao[1], Hui Liu[2], Thaibinhduong Nguyen[1], Matthew Schneider[1], Paul South[3], Donald Ort[3], John Shanklin[2], Stephen 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) Overcoming recalcitrance in tissue culture and genetic transformation of energycane.
2) Engineer energycane to produce an abundance of lipids in the form of triacylglycerol which can be converted into biodiesel, biojet fuel, and bioproducts.
3) 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 towards hyper-accumulation of lipids [e.g. triacylglycerol (TAG)] in the vegetative tissues of high biomass yielding crops is a new strategy to improve lipid yields for biofuel production. Energycane is an ideal feedstock for this approach due to its superior biomass production and persistence. However, energycane is among the most recalcitrant crops in tissue culture, impeding its genetic transformation.

Visual browning of the newly excised energycane explants is a major hurdle that needs to be overcome to establish an efficient genetic transformation protocol for this target species. We investigated effects of several culture media supplements (e.g. anti-oxidants/anti-browning agents) in the tissue culture medium on visual tissue browning in energycane. The combination of 2 to 3 anti-browning agents significantly reduced visual tissue browning while increasing the number of regenerating plantlets from energycane callus.

Plant lipid is one of the most energy-rich and abundant forms of reduced carbon available from nature. 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.

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.

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.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (Award Number DE-SC-0018254).

Identification of Adaptive Fungal Pathogen Resistance Loci in Switchgrass

David B. Lowry[1,*] (dlowry@msu.edu), Acer VanWallendael[1], Shawn C. Kenaley[2], Chathurika G. Wijewardana[2], Gary C. Bergstrom[2]

[1]Michigan State University, East Lansing, MI; [2]Cornell University, Ithaca, NY

Project Goals:

Switchgrass (*Panicum virgatum*) is an important target species for domestic production of cellulosic biofuels, but it is susceptible to multiple fungal pathogens. The principle aim of our research is to identify the loci responsible for disease resistance in switchgrass and determine how effective those resistance loci are across geographic space. To accomplish this overarching goal, we are: 1) Characterizing the pathogens associated with disease in switchgrass and quantify their geographic distributions. 2) Discovering genetic loci for effective switchgrass disease resistance across different geographic locations. 3) Validating QTLs for pathogen resistance through controlled experiments. Overall, the discovery of loci and genes involved in resistance to specific pathogens will make crucial improvements of switchgrass cultivars possible through future breeding and gene editing efforts.

Abstract:

Switchgrass is an important target species for domestic production of cellulosic biofuels. The principal aim of most switchgrass breeding programs is to develop high-yielding cultivars. However, as feedstock plantings expand, so will pathogen pressure. Unless controlled, fungal pathogens with explosive disease potential will likely drive yield declines and economic losses. Pathogen resistance can be developed through breeding programs that exploit natural genetic variation in disease resistance. Much of the functional genetic variation in switchgrass, including pathogen resistance, is distributed clinally with latitude as well as between ecotypes. In general, southern lowland cultivars are more resistant to fungal pathogens than northern upland cultivars.

To identify quantitative trait loci (QTL) responsible for pathogen resistance, we have conducted QTL mapping of rust infection on a mapping population planted across the central United States. The mapping population was created by crossing two northern upland accessions with two southern lowland accessions to create a four-way outbred population. This outbred mapping population was planted in 2015 at an unprecedented geographical scale, spanning ten common garden field sites distributed over 17 degrees of latitude in the central United States. Over the following three years (2016-2018), we collected ~149,000 rust infection scores on the mapping population across eight field sites. This data allowed us to identify two major QTLs and many minor QTLs contributing to rust resistance across field sites. The two major QTLs had large effects at northern field sites, but had little or no detectable effect at the southern field sites. These two QTLs co-localize with biomass QTLs, which suggests that gains could be made in both important traits by selecting upon these loci. Further, there was a strong epistatic interaction between the two major effect QTLs. Thus, it will be very important to

consider both genotype x environment and genotype x genotype interactions when breeding disease resistance in switchgrass. The finding that major resistance loci generally only had effects in the north suggests that there are either different rust types in the north or that resistance is modified by local environmental conditions. To examine whether the abundance of different rust species was diverged between the north and the south, we used a combination of microscopy and molecular methods on field collected samples. This analysis found that *Puccinia panici* was by far the dominant rust species present at all of the sites we surveyed, which suggests that species-level differences in resistance are not responsible for the QTL differences between the north and the south. We are currently conducting population genetic aanalyses to determine whether population structure within *P. panici* could explain the latitudinal QTL patterns. In the summer of 2019, we began phenotyping a new genome-wide association study (GWAS) mapping population that was planted at 10 sites in the United States and three field sites in central Mexico. Our initial GWAS analysis has discovered additional loci involved in rust resistance. Overall, our efforts are providing foundational research that will facilitate the future development of more rust resistant cultivars of switchgrass.

Funding Statement:

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0017883.

Simultaneous 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 natural carbon sources composed of multiple substrates, there are two utilization strategies: single 'superbugs' (SS) capable of simultaneously utilizing all of these substrates and the division of labor (DOL) among strains with each specializing in one substrate. Our working hypotheses are that heavy metabolic cost, including burdens and toxicity, of substrate utilization favors DOL over SS and that DOL is more adaptive than SS to changing substrate compositions. To test the hypothesis, we will engineer a set of *Saccharomyces cerevisiae* strains to implement the both strategies and also build a multiscale mathematical model to quantitatively elucidate the general rules for designing substrate co-utilization.

Abstract

Abundant and inexpensive agricultural residues contain mixed sugars, such as glucose, xylose, and arabinose, that can be utilized by wild-type and engineered microbes for chemical production. Studies have shown that, during consumption, microbes often follow a sequential fashion for mixed sugars—preferential consumption of glucose over other sugars—which results in impaired yield and productivity of target molecules. Here, we dissect glucose and xylose consumption by constructing a yeast consortium of a glucose utilizing strain and a xylose utilizing strain. The former (Y_{G1}) was constructed by deleting endogenous hexose transporters (*HXT1-7* and *GAL2*) and introducing a heterologous glucose-specific transporter in a wild-type *Saccharomyces cerevisiae*. The latter (Y_{X1}) was constructed by deleting endogenous hexose transporters (*HXT1-7* and *GAL2*) and introducing a heterologous xylose-specific transporter in an engineered *S. cerevisiae* expressing xylose reductase (*XYL1*), xylitol dehydrogenase (*XYL2*), and xylulokinase

(*XYL2*).

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 major hexose transporters (*HXT1-7* and *GAL2*) were deleted, Y_{X1} could consume glucose after lengthy cultivations on glucose due to activation of minor hexose transporters (*HXT8-17*). In order to prevent glucose consumption of Y_{X1}, we deleted hexokinases (*HXK1-2*) and glucokinase (*GLK1*) which are responsible for glucose phosphorylation so that glucose consumption can be completely eliminated even after the activation of minor hexose transporters. Indeed, Y_{X2} (*HXK1-2* and *GLK1* deleted Y_{X1}) 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 sugars were 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}. As a result, the glucose and xylose uptake rates of Y_{G2} and Y_{X3} were comparable. Lastly, monitoring the populations of Y_{G2} and Y_{X3} during cultivation was difficult. As such, fluorescence proteins (GFP and RFP) were expressed in Y_{G2} and Y_{X3} to monitor each population using a real-time fluorescence image of the consortium. The resulting Y_{G2f} and Y_{X3f} were employed for studying the division of labor during fermentation of glucose and xylose via DOL.

In summary, we confirmed that the inoculation ratio and controlled sugar consumption rate of Y_{G2} and Y_{X3} can be used as tuning parameters to optimize simultaneous consumption of glucose and xylose at various concentrations. Our results demonstrate that the division of labor in ecosystems could be applied to consume mixed sugars and develop biochemical processes for efficient conversion of mixed sugars into value-added products.

Dissecting the Social Interactions of Yeast-Lactic Acid Bacteria Consortia

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:

Microbial social interactions are a major driving force that regulates the organization and functioning of microbial communities. Here we experimentally investigate cellular interactions in various yeast—lactic acid bacteria (LAB) consortia and determine how these interactions contribute to corresponding ecosystem dynamics.

Abstract

Engineered microbial consortia must be able to generate desired population behaviors for reliable and efficient industrial bioprocessing. One promising way to develop such ecosystems is through the design and construction of specific cellular social interactions. Here, we experimentally uncover microbial interactions in various yeast—lactic acid bacteria (LAB) consortia and determine how these interactions contribute to corresponding ecosystem dynamics. In a glucose-supplemented, chemically defined medium which supports only the growth of yeast, we found that co-culturing *S. cerevisiae* S90 with LAB, including *Lactococcus lactis*, *Lactobacillus brevis*, *Lactobacillus plantarum* and *Lactobacillus acidophilus*, promotes their growth while being minorly inhibited. In a lactose-supplemented, defined medium that supports LAB only, LAB are able to grow normally but *S. cerevisiae* S90 fail when they are co-cultured. One exception is the *Lb. brevis—S. cerevisiae* S90 co-culture where the both grow better than their monocultures, suggesting that *S. cerevisiae* S90 and *Lb. brevis* can form a mutualistic interaction. To create the same symbiotic interaction between yeast and Lactococcus species, we engineered a lactose-positive and galactose-negative *L. lactis* strain, MG2, using direct evolution and genetic engineering. Our subsequent co-culture experiment confirmed that *S. cerevisiae* S90 and MG2 indeed form a mutualism consortia. Together, our experiments show different modes of social

interactions in yeast-LAB consortia and their corresponding community dynamics, which provides insights into the organization of yeast-LAB consortia and future applications of these ecosystems for metabolic engineering purposes.

Funding

This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0019185

Constructing the Nitrogen Flux Maps (NFMs) of Plants

Yasuo Yoshikuni[1*](yyoshikuni@lbl.gov), Trent Northen[1], Taichi Takasuka[2], Zoran Nikoloski[3]
Hiroshi Maeda[4]

[1]Lawrence Berkeley National Laboratory (LBNL), Oakland, CA; [2]Hokkaido University, Sapporo, Japan; [3] University of Potsdam & Max Planck Institute of Molecular Plant Physiology, Germany; and [4]University of Wisconsin-Madison, WI

URL: https://nfluxmap.github.io/

Project Goals: To construct plant N flux maps (NFMs) from plant genomes and to determine both biochemical and system-level functionality of plant N metabolic network.

Abstract: Nitrogen (N) is an essential element of organic molecules, such as amino acids and proteins, but is often 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)**". The lack of understanding of NFM represents a fundamental and critical knowledge gap in plant biology and agriculture. **Our main objectives are to construct plant NFMs from plant genomes and to determine system-level functionality of 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.

While the early steps of N metabolism, i.e. inorganic N uptake and assimilation, have been extensively studied, much less is known how assimilated N is distributed throughout the plant metabolic network. The core of NFM is composed of different branches of amino acid metabolism interconnected by **aminotransferases (ATs)**. AT enzymes play pivotal roles in distributing reduced N for synthesis of various organonitrogen compounds, such as amino acids and their downstream products including proteins, nucleic acids, and alkaloids. Genes and enzymes responsible for specific AT reactions have been identified. However, an AT enzyme can potentially catalyze 380 different transamination reactions simply considering 20 proteogenic amino acids and corresponding keto acids as potential substrates. Yet, the multi-substrate specificities of ATs remain largely uncharacterized due to their poor sequence-function relationships and tedious aminotransferase activity assays that allow testing of only "two substrates at a time". As a result, we do not know the true functionality of AT enzymes and how ATs transfer N across the plant N metabolic network, i.e. NFM.

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. **Aim 1** will construct

a biochemically feasible NFM from the plant genomes using computational modeling. **Aim 2** will experimentally determine biochemical functions of AT enzymes through high-throughput gene/protein synthesis and enzyme assay platforms, followed by computational protein modeling to predict AT functions from other species. **Aim 3** will validate the system-level functionality of ATs and will determine N flux distribution under different N availability through kinetic stable isotope precursor feeding. 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 (Relevant prior publications by PIs)

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. Dornfeld, C., Weisberg, A. J., K C, R., Dudareva, N., Jelesko, J. G. & Maeda, H. A. Phylobiochemical characterization of class-Ib aspartate/prephenate aminotransferases reveals evolution of the plant arogenate phenylalanine pathway. *Plant Cell* **26,** 3101–3114 (2014).
4. Wang, M. & Maeda, H. A. Aromatic amino acid aminotransferases in plants. *Phytochem Rev* 1–29 (2017).
5. 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).
6. 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).
7. 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).
8. 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).
9. Yoshikuni Y., Ferrin T.E. & Keasling J.D. Designed divergent evolution of enzyme function. *Nature* **440,** 1078–1082 (2006).
10. Yoshikuni, Y., Martin, V. J. J., Ferrin, T. E. & Keasling, J. D. Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chem. Biol.* **13,** 91–98 (2006).
11. 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).
12. Arnold, A., Sajitz-Hermstein, M. & Nikoloski, Z. Effects of varying nitrogen sources on amino acid synthesis costs in Arabidopsis thaliana under different light and carbon-source conditions. *PLoS ONE* **10,** e0116536 (2015).
13. Szecowka, M., Heise, R., Tohge, T., Nunes-Nesi, A., Vosloh, D., Huege, J., Feil, R., Lunn, J., Nikoloski, Z., Stitt, M., Fernie, A. R. & Arrivault, S. Metabolic fluxes in an illuminated Arabidopsis rosette. *Plant Cell* **25,** 694–714 (2013).
14. Eloundou-Mbebi, J. M. O., Küken, A., Omranian, N., Kleessen, S., Neigenfind, J., Basler, G. & Nikoloski, Z. A network property necessary for concentration robustness. *Nat Commun* **7,** (2016).
15. Küken, A. & Nikoloski, Z. Computational Approaches to Design and Test Plant Synthetic Metabolic Pathways. *Plant Physiol.* **179,** 894–906 (2019)

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 a substantial portion of accumulated carbon from starch to oil metabolism in *Lemnaceae*, using resting fronds as the storage tissue.

Recent efforts in the Martienssen and Lam labs have produced three new reference quality *Lemnaceae* genome assemblies complete with chromosome structures, methylomes, small RNA transcriptomes, and structural variant analysis for these novel genomes. We performed single molecule long-read genome sequencing of diploid *L. gibba*, diploid *W. australiana*, and allotetraploid *L. minor* clones using Oxford Nanopore technology followed by Hi-C. Comparisons of the resulting chromosome-scale assemblies reveal that the 21 chromosomes of diploid *L. gibba* are highly colinear with each of the subgenomes of allotetraploid *L. minor*,

while the *W. australiana* has 20 chromosomes with significant architectural differences with the 20 chromosomes of the giant duckweed *S. polyrhiza*. *S. polyrhiza*, *L. gibba*, and the two subgenomes of *L. minor* all encode around 18,000 genes – significantly fewer than terrestrial monocots such as rice and Brachypodium, and comparable to the unicellular alga *Chlamydomonas reinhardtii*. *W. australiana* has undergone an even more striking reduction to only 14,000 genes. Whole methylome sequencing has shown that *Spirodela polyrhiza* has among the lowest cytosine methylation levels in plants and lacks CpG gene body methylation. Coincident with reduced methylation, *S. polyrhiza* has very little retrotransposon sequence, while *L. gibba* and *L. minor* retain a similar retrotransposon content to other monocot genomes, and *W. australiana* is intermediate. Small RNA sequencing has revealed dramatic differences between the three genera consistent with known pathways of RNA directed DNA methylation. We have analyzed orthologous gene content across the *Lemnaceae*, 11 other monocots and 9 non-monocots, revealing variations that likely account for some of these differences, as well as for reduced morphology, clonal reproduction, and aquatic growth habit.

The Birchler Lab has completed the design and construction of a transgene stacking system with alternating transformation vectors that enable iterative recombination into a specific genomic site specified by a previously integrated landing pad. This novel design is compatible with consecutive transformation of *Lemnaceae* undergoing strictly clonal propagation. In addition, antibodies against Centromeric histone H3 have been raised for the four sequenced species mentioned above, and efforts are underway to visualize centromere organization and validate predicted centromeric repeat sequences.

Key experiments in the Lam Lab have confirmed that natural genetic variation in *S. polyrhiza* leads to more or less turion production under the same induction by phosphate limitation. Further, methods for RNA extraction from high starch (>75%) turions have been optimized and applied to high throughput RNA-sequencing experiments of two genotypes at the extremes of turion yield along with their corresponding vegetative fronds. These experiments have identified turion specific genes associated with dormancy and starch biosynthesis along with others encoding putative transcription factors that may be involved in the developmental transition.

Critically, we have already developed engineered *L. minor* exhibiting a significant increase in oil content, building on the successful engineering of sugarcane to achieve 2-5% leaf TAG in the Shanklin Lab under ARPA-E support. Engineered lines include stable overexpressers of WRINKLED, DGAT and PDAT1, all exhibiting marked increases in TAG content. As expected, efforts to eliminate futile cycling and FA cytotoxicity in our first generation transgenics have resulted in dramatic increases in oil content. In our most recent multigene overexpression lines including OLE1, we now observed TAG accumulation up to 6% of dry weight in *L. minor* with no apparent growth rate defect.

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

Andreas E. Vasdekis,[1] * (andreasv@uidaho.edu), Tomislav Ticak,[1] Rabindra Khanal,[1] Andrew T. Johnson,[1] Cole S. Garrett,[1] Monica J. Pedroni,[1] Triana N. Dalia[2], Sergey Stolyar,[1] Ankur B. Dalia,[2] and **Christopher J. Marx**[1,PI]

[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 made significant progress on the development of the genetic tools for introduction of genome-edited gene clusters from *Vibrio* into *Methylobacterium*. Second, we have developed the capacity to grow *Methylobacterium* in patterned microfluidic devices to permit growth dynamics and PHB accumulation to be ascertained. Third, we have examined PHB accumulation across a variety of substrates, growth phases, and nitrogen levels. This allows comparison of image cytometry, HPLC analysis of total levels, and flow cytometry. Fourth, these PHB accumulation studies have been conducted for a variety of natural strains, revealing interesting diversity in both the mean and variation of PHB produced. Fifth, we have nearly completed genetic and physiological analysis of the novel pathway for aromatic utilization that we have found in *Methylobacterium*. Finally, we have evolved strains to effectively utilize novel methoxylated aromatic compounds. All of these steps forward move us toward our ultimate goal to develop *M. extorquens* for conversion of methoxylated aromatics to butanol, while simultaneously demonstrating a novel approach that combines the advantages of gene editing and

deep-sequencing, an internally-accumulated product as a proxy, and an analysis of phenotypic heterogeneity for both growth and production.

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

Gerald N. Presley[1], Leah H. Burdick[1], Jacob H. Cecil[1], and **Joshua K. Michener***[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; * michenerjk@ornl.gov

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.

In a typical biorefinery, biomass-derived sugars are fermented to fuels by microorganisms, while residual lignin is burned for process heat. Converting waste lignin into value-added bioproducts offers a potential source of additional revenue to improve the economics of biofuel production. While bacteria have been isolated and engineered to catabolize lignin-derived compounds, economically viable lignin valorization will require further work to discover and optimize these pathways in non-model bacteria. Extending beyond degradation pathway intermediates, new bioproducts must also be deployed that have sufficiently large markets and a competitive advantage for biosynthesis from lignin. Finally, engineered assimilatory and conversion pathways must be integrated with the host metabolic and regulatory networks to ensure efficient operation.

This project will optimize biochemical pathways for assimilation of lignin-derived compounds in *Novosphingobium aromaticivorans*. This strain is capable of efficiently catabolizing more lignin-derived compounds than were previously recognized, and we continue to identify and characterize catabolic pathways for relevant compounds including syringate, guaiacol, and β-1 aromatic dimers.

New pathways will then be engineered into *N. aromaticivorans* for conversion of lignin-derived aromatic compounds into valuable bioproducts. Most industrial biosynthetic pathways have been optimized for production from glucose or glycerol, and the alternate stoichiometry from aromatic catabolism offers potential advantages for certain classes of bioproducts.

To function efficiently, the assimilatory and conversion pathways described above must be carefully integrated into the host metabolism. To this end, we are applying novel techniques for bacterial quantitative trait-locus mapping to perform genome-wide screens for loci that affect these functions. This information will be used to build a predictive genetic model and identify additional targets for further metabolic optimization.

In combination, these efforts will result in new methods to predictively model and engineer a promising microbe for lignin valorization. The same methods can ultimately be applied to a wide range of emerging microorganisms relevant for BER's mission in sustainable bioenergy.

References

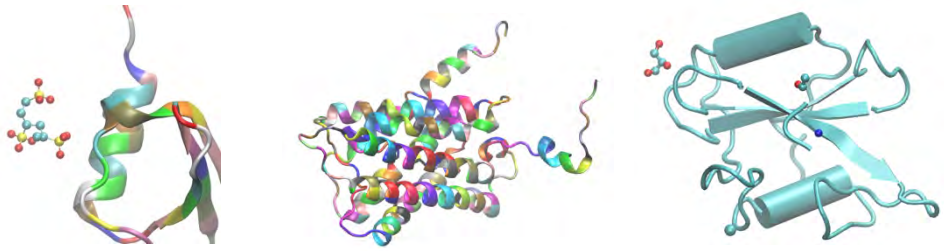
1. Cecil JH, Garcia DC, Giannone RJ, Michener JK. Rapid, parallel identification of catabolism pathways of lignin-derived aromatic compounds in *Novosphingobium aromaticivorans*. *Appl Environ Microbiol* 2018 84:e01185-18.

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.

Structure to Function: Bringing Protein Structure and Ligand Screening to KBase

Ada Sedova[1]* (sedovaaa@ornl.gov), Tianhao Gu[2], Loukas Petridis[1], Brian H Davison,[1]
Christopher S. Henry[2], and **Julie Mitchell**[1]
[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee
[2]Argonne National Laboratory, Argonne, Illinois
http://cmb.ornl.gov/index.php/research/bioenergy/dynamic-visualization-of-lignocellulose

Project Goals: The annotation of gene function benefits greatly from information on protein structure and binding. We have been developing new functionality for structural analysis and virtual ligand screening in KBase, in support of the Dynamic Visualization of Biological Structures SFA at Oak Ridge National Laboratory. The new structural branch of KBase includes tools with which to upload and view protein three-dimensional structure, and to create co-evolutionary analysis-based contacts maps for use as restraints in protein folding algorithms. An open-source protein folding tool has been developed, and a protein-ligand docking workflow allows for the use of the AutodockVINA[1] software to screen databases of metabolites, connecting metabolic-level function via physical molecular interactions to genomic information.



The annotation of gene function benefits greatly from information on protein structure and binding. We are developing a new structural branch for KBase, in support of the Biofuels SFA at Oak Ridge National Laboratory. We have created the software infrastructure in KBase that uploads a model protein structure or downloads an experimental structure from the RCSB Protein Data Bank, and a web-based three-dimensional viewer to display the protein. This infrastructure paves the way for a number of structure-based analysis tools to be used in KBase that can add important information to computational biology research. The ligand screening application we have developed uses the AutodockVINA[1] protein-ligand docking program to allow the protein structure to be screened against a set of small molecule metabolites. A batch downloader to interface with the ZINC[2] database allows for the import of three-dimensional mol2 ligand files, providing robust starting geometries for the ligands. This application can be used as a part of a functional annotation of uncharacterized proteins, as ligand screening will help researchers anticipate likely functions. In addition to ligand docking, we have developed applications for protein structure prediction using co-evolutionary analysis. This method infers protein residue contacts from a multiple sequence alignment. The CCMPredPy[3] method has

been chosen for this application. The contacts can then be used as restraints in protein folding algorithms to generate a three-dimensional structural model for the protein from sequence. Information from the contact map an also provide value independent of any structure prediction, as the results indicate amino-acid residues that drive folding stability which is useful for experimental strategies in protein redesign and can also be used to refine the MSA and to create a phylogenetic tree. Finally, we have been developing the only open-source protein folding tool to take a set of distance restraints and arrive at a final three-dimensional structure.

Lignocellulosic biomass is a complex substrate that requires the synergistic action of a variety of enzymes for its efficient deconstruction. Biomass pretreatment generates byproducts, including solubilized lignin-derived aromatics, that inhibit enzymatic hydrolysis of cellulose[4]. Which bioproducts are formed depends on the biomass feedstock as well as the details of the pretreatment process. Applying the Ligand Screening App to predict which specific byproducts affect which particular enzymes can lead to an optimal selection of cellulolytic enzyme cocktails that minimize inhibition. A further barrier to biofuels and bioproduct production is that fermentation products, pretreatment solvents and byproducts can be toxic to microorganisms. Determining which proteins, and which protein residues, the small molecules bind to may lead to rational genetic engineering of those proteins and to microbes exhibiting improved tolerance to toxic pretreatment byproducts and solvents.

1) Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry. 2010 Jan 30;31(2):455-61.

2) Sterling T, Irwin JJ. ZINC 15–ligand discovery for everyone. Journal of chemical information and modeling. 2015 Nov 23;55(11):2324-37.

3)Vorberg S, Seemayer S, Söding J. Synthetic protein alignments by CCMgen quantify noise in residue-residue contact prediction. PLoS computational biology. 2018 Nov 5;14(11):e1006526.

4) Ximenes E, Kim Y, Mosier N, Dien B, Ladisch M. Deactivation of cellulases by phenols. Enzyme and microbial technology. 2011 Jan 5;48(1):54-60.

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.

Evaluating Biogeochemical Processes Facilitated by Plant and Microbial Interactions within the Rhizosphere

James Moran[1]* (james.moran@pnnl.gov), Vivian Lin[1], Kumar Swatantar[1], Monee McGrady[1], Ying Zhu[1], Jamie Nuñez[1], Elias Zegeye[1], Sarah Fansler[1], Yuliya Farris[1], Ryan Renslow[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.

We hypothesize that localized regions within the rhizosphere act as foci for exchanging root-derived organic carbon with soil-derived nutrients made available by a combination of microbial activity and inherent soil resource availability. Further, we hypothesize that the locations of these resulting nutrient exchange hotspots are not stochastically distributed throughout the rhizosphere but are controlled by microenvironmental conditions resulting from a combination of plant-derived carbon, microbiological activity, and soil geochemistry. To test these hypotheses, we are applying a suite of tools to evaluate the rhizosphere within a series of microcosms constructed with natural soil (Kellogg Biological Station, Hickory Corners, Michigan, USA) and *Panicum virgatum*, (switchgrass, variety Cave-in-Rock) seedlings. We are using these investigations to evaluate drivers of spatial heterogeneity in root exudation rates, rhizosphere microbial membership and activity, and localized shifts in soil geochemistry and nutrient availability.

In the first investigation, we are imposing spatially resolved variation in phosphorus availability to track the degree of plasticity in root physiology and how this can be leveraged to improve plant fitness. We applied spatially constrained phosphorus resources [either inorganic (calcium triphosphate) or organic (phytic acid) sources] with and without the addition of a known plant growth promoting organism implicated in phosphorus mobilization (*Flavobacterium johnsoniae*) to assess localized plant response. Monitoring levels of soluble phosphorus in pore water extracted both proximal and distal to roots allowed us to demonstrate an increase in orthophosphorus availability at the interface of plant roots and the applied resource, suggesting plant linkage to enhancement of solubilization. We further observed the greatest levels of phosphorus solubilization under experimental conditions supplemented with the *Flavobacterium*, highlighting the potential bacterial role in phosphorus cycling within the system which is consistent with previous studies. We are currently pushing beyond this association by assessing the ability of plant roots to preferentially direct organic carbon exudation into areas of higher phosphorus ability to help stimulate further solubilization of this resource. We are also assessing the variability in the composition of soluble organic compounds extracted from pore waters at various locations in the

rhizobox microcosm to provide insight to the mechanism of interaction between plant host and phosphorus solubilizing bacteria which is facilitated by plant root exudation.

Secondly, we developed a suite of spatially resolved techniques designed to elucidate microbial activity and distribution. We are spatially mapping phosphatase activity within our rhizobox microcosms using fluorescent, substrate-specific activity probes. This approach enabled us to identify increased phosphatase activity in the spatial vicinity of a root growing proximal to an applied organic phosphorus sources in comparison to a control root (Figure 1). A key feature of the approach is its non-destructive nature which allows for tracking shifts in phosphatase activity during plant growth and at different phases of root development. We are currently working to link this phosphatase approach with a similar probe targeting chitinase activity to help spatially map intersections of phosphorus, carbon, and nitrogen cycling within rhizosphere systems. These hotspots of nutrient cycling are hypothesized to be key locations of nutrient transfer to plant roots.

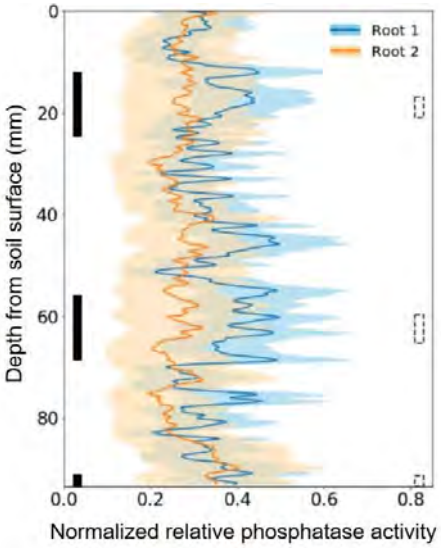


Figure 1: Vertical variation in phosphatase activity surrounding two roots was tracked as the roots grew through soil sections supplemented with an organic phosphorus source (Root 1, approximate areas of enrichment depicted by the shaded box) and a control (Root 2, approximate area depicted by dashed box).

In recognition of the dominant role of microbial processes in nutrient cycling, we are developing protein extraction approaches coupled with proteomic analysis to permit spatially resolved assessment of taxonomic and, in some cases, functional distribution within this complex system. This approach is non-destructive when applied in our rhizobox systems and enables us to track variation in the observed metaproteomes associated with root development as well as spatially focused variation arising from microenvironments within the rhizobox. We are currently employing the approach to differentiate proteomic expression in the rhizospheres associated with roots grown under two different nutrient conditions – phosphorus replete and deplete. We are using a split root experimental design for these experiments where we allow roots from a single plant to grow into these controlled nutrient conditions and are just beginning to assess resulting adaptation of the plant and microbial proteome in response to shifted nutrient conditions.

Our developments are enabling us to interrogate the high spatial heterogeneity of the rhizosphere system to begin elucidating key interactions being leveraged to facilitate nutrient transfer to a host plant. This knowledge will provide insights on how to strengthen nutrient transfer processes to increase plant performance under varied and challenging growth conditions.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER). This contribution originates from an Early Career Research Award granted at the Pacific Northwest National Laboratory (PNNL).

Metabolic and membrane adaptations of the hydraulically fractured shale isolate Halanaerobium in response to temperature and growth rate fluctuations under continuous culture

Fabrizio Colosimo, [1] Samuel O. Purvine,[2] Heather M. Brewer, [2] Allison R. Wong, [2] Elizabeth K. Eder, [2] David W. Hoyt, [2] Stephen J. Callister,[2] and **Paula J. Mouser**[1]* (Paula.Mouser@unh.edu)

[1]University of New Hampshire, Durham, [2] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington

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: Here we characterized the physiology of *Halanaerobium congolense* WG10, a dominant taxon isolated from a 2.5-km deep hydraulically fractured natural gas well in Ohio, 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 ^{1}H-NMR and GC-FID) under steady state growth rates of 0.021 to 0.052 1/hr. Our experiments highlighted 1.5 fold increased biomass at higher dilution rates (HRT 19.2 hrs) and warmer temperatures (40ºC) as compared with slower dilution rates (HRT 48 hr) and cooler temperatures (25ºC).

Proteomics analysis showed a total of 2,227 out of 2,800 predicted protein-coding genes (79.5%) were identified in our data set. Among those proteins identified, 356 were found to be significantly higher in abundance in one or more treatments (Student's t test, $p < 0.05$). Of these, 91 proteins were identified during cooler temperature growth (25ºC) while 109 were identified in growth at warmer temperature (40ºC). An additional 71 were in greater abundance when *H. congolense* WG10 was grown at 40°C in the highest dilution rate (HRT 19.2 hrs). When grown at 25ºC, cells exhibited cold shock proteins (CspA family, WG10-13112 -12818), a typA, bipA GTP binding protein involved in stress response (WG10-10337), and a nucleotide-binding universal stress protein (UspA family, WG10-10469). 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 stress. The propanediol dehydratase is a cobamide-dependent enzyme that has been shown to also dehydrate ethylene glycol to acetaldehyde. The formation of both propionaldehyde and acetaldehyde was confirmed with both ^{1}H-NMR and 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 ^{1}H-NMR and GC-FID. Lipidomics analysis and fatty acid methyl ethyl analysis is currently underway to characterize key membrane lipids for these treatments. In parallel with experimental efforts, produced fluid samples were collected from hydraulically fractured shale wells from MSEEL II, a DOE NET-funded field research site in West Virginia for analysis of intact polar lipids and phospholipids. Our continuous culture MPLEx approach sheds new light on the metabolism and membrane features of *Halanaerobium* under biogeochemical drivers relevant to engineered shale.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER) and the Established Program to Stimulate Competitive Research (EPSCoR) program under Award Number DESC0019444.

Genome sequencing reveal structural and nucleotide-level divergence among immunosuppressing G-type Lectin Receptor kinases across multiple *Salix* species.

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 host immunosuppression during endophyte recruitment in the genus *Salix* and access speciation-driven divergence of these molecular process at the genome level.

In revealing a novel role for self-immunosuppression in plants during host-cell invasion by microbes, Plasminogen-Apple-Nematode (PAN) domain proteins, D-mannose lectin receptor kinases (G-LecRKs), were shown to function as negative regulators of defense signaling during pathogenesis by 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]. Moreover, PAN domain carrying S-locus kinases, reported to mediate self-incompatibility during pollination, fall under the same class of G-LecRKs[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 cysteine residues, is a unifying feature that is found in association with proteins involved in immunosuppression across highly divergent organisms. Further, we reveal that PAN domain proteins are used or targeted by pathogens, parasites and symbionts during host-cell invasion. As such, the PAN domain is a reliable biomarker for the host-cell invasion machinery. Species divergence in the PAN domain-containing G-LecRKs was evaluated using Nanopore assemblies of 11 genomes sampled to represent diverse speciation events in the genus *Salix* and implications on species-level differences in endosphere microbial community diversity will be illustrated.

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).

Elucidating Nutrient-Dependent Effects on Regulation of Photosynthesis and Metabolism

Tim L. Jeffers[1]* (tim_jeffers@berkeley.edu), Ryan McCombs,[1] Sam Purvine,[2] Daniel Westcott,[1] Mary Lipton,[2] **Krishna K. Niyogi**,[1,3,4] and Melissa S. Roth[1]

[1]Department of Plant and Microbial Biology, University of California, Berkeley; [2]Pacific Northwest National Laboratory, Richland; [3]Howard Hughes Medical Institute, University of California, Berkeley; [4]Lawrence Berkeley National Laboratory, 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 the 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.

Both photosynthesis and the generation of neutral lipids that can be used as biofuels are heavily regulated by nutrient availability in algae. The oleaginous green alga *Chromochloris zofingiensis* displays the ability to reversibly shut off photosynthesis entirely ($F_v/F_m = 0$, no oxygen evolution), independent of light, in the presence of glucose (+Glc) and insufficient iron (-Fe), leading to heterotrophic cultures that survive solely on glucose as a carbon source (1). This photosynthetic switch is also dependent on a functional hexokinase (HXK1) gene (2). Photosynthesis can be restored after glucose removal (1). Recently, we have found that replete iron supplementation (+Fe) can also rescue photosynthesis in the presence of glucose. With the goal of understanding the regulation of trophic states and the accumulation triacylglycerols (TAGs) in this system, we conducted a full combinatorial proteomic analysis of the wild type vs. two independent *hxk1* mutant strains grown with and without glucose and in iron-replete and iron-limiting conditions (n = 3-4, 47 total samples). Through mass spectrometry of isobaric labelled peptides, we quantified the relative abundance of 11,282 proteins across our samples, which is ~70% of the annotated proteome. In our preliminary analyses, we note large decreases, specific to the heterotrophic state, in almost all protein subunits of the photosynthetic electron transport chain, consistent with photosynthesis being turned off **(Figure 1)**. In addition, wild-type cells treated with glucose have

increased abundance of most of the lipid biosynthesis pathway. Our goal is to investigate the differentially abundant proteins, including several of unknown function, in heterotrophy and TAG-accumulating conditions to find gene targets for enhanced biofuel potential of this organism.

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. doi: 10.1105/tpc.18.00742.
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. doi: 10.1038/s42003-019-0577-1

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.

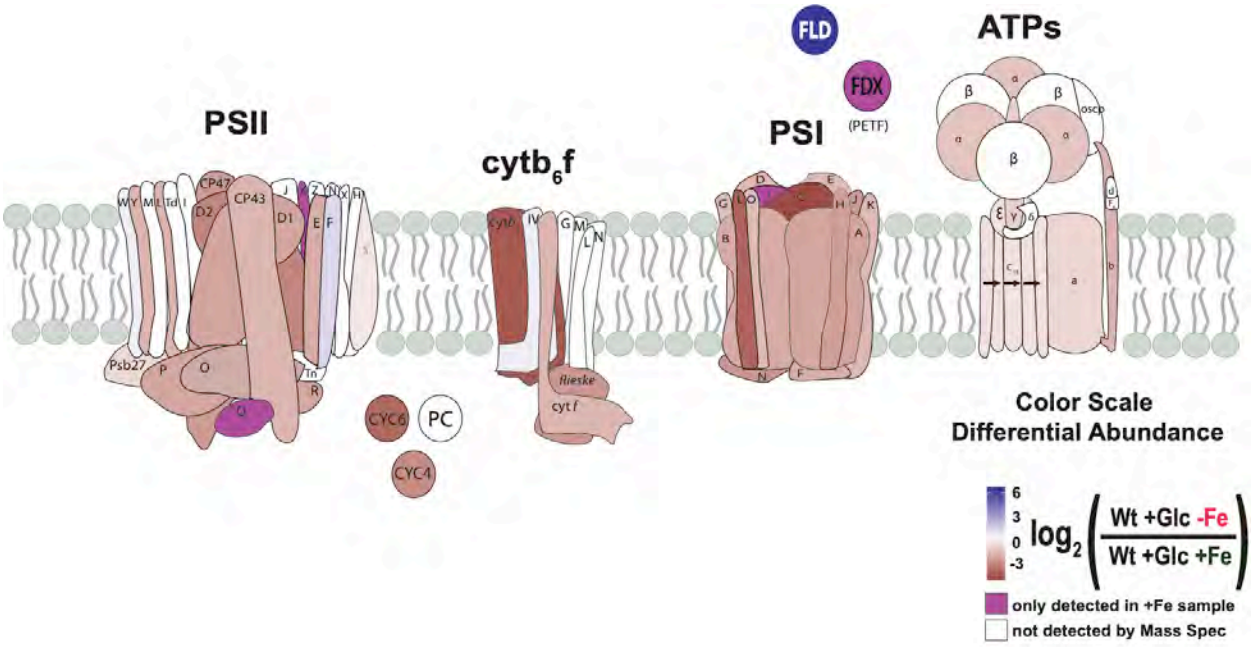


Figure 1. Universal downregulation of photosynthetic electron transport in heterotrophy (-Fe+Glc) vs. mixotrophy (+Fe+Glc). *Almost all photosynthetic complexes and electron carriers have lower protein abundance (increasing red color) in iron-deficiency-induced heterotrophy* (WT -Fe+Glc) *in comparison to the iron-replete, mixotrophic state* (WT+Fe+Glc) *that maintains photosynthesis. An exception is Flavodoxin (***FLD***) (blue circle, 90X more abundant in -Fe), an electron carrier that functions as Ferredoxin (***FDX***), but without an iron cofactor.* **PSI and PSII:** *Photosystems I and II,* **cytb$_6$f:** *cytochrome b$_6$f complex,* **ATPs:** *chloroplastic ATP synthase,* **CYC6 and CYC4:** *soluble cytochrome-c electron carriers,* **PC:** *plastocyanin.*

Genome-based Protein Function Discovery in the Eukaryotic Alga *Chromochloris zofingiensis*

Fatima Foflonker[1], Sean D. Gallaher[2], Sean McCorkle[1], Sabeeha Merchant[3] and **Crysten E. Blaby-Haas[1]*** (cblaby@bnl.gov)

[1]Biology Department, Brookhaven National Laboratory, Upton, NY; [2]Department of Chemistry and Biochemistry and Institute for Genomics and Proteomics, University of California, Los Angeles; [3]Quantitative Biosciences Institute, Department of Plant and Microbial Biology, University of California, Berkeley

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. 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.* Toward these objectives, we are implementing a phylogenomics-guided approach that leverages evolutionary relationships between genomes and between proteins encoded on those genomes for contextualized and evidence-based protein function discovery. For more information about the project and our team, please visit: https://sites.google.com/view/czofingiensis/home

The classic, but outdated, view of eukaryotic genomes is of gene islands randomly situated in a sea of non-coding DNA. This picture is derived from the observation that in contrast to prokaryotes, where functionally cooperating proteins are often encoded by operons, such structural organization does not appear to be necessary for co-regulating functional units in eukaryotes, in part because transcription and translation are uncoupled. 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, such as physical clustering of pathway members and co-regulated genes, has emerged as a characteristic of eukaryotic genomes. Current methods for identifying functionally cooperative gene neighborhoods in eukaryotes rely 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, while functionally annotated genes may be mis-annotated, inaccurate or vague, because of the evolutionary distance between algae and well-characterized model organisms, such as yeast and *E. coli*. 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. While *C. zofingiensis* has a high quality, chromosome-complete genome assembly, the currently available structural gene annotations for the species are replete with observable mis-annotations and fragmented genes. In an effort to correct this shortcoming, we collaborated with the Joint Genome Institute to generate whole-molecule, long-read sequencing of the transcriptome (IsoSeq) on the PacBio Sequel platform. With

a combination of computational analysis and manual curation, this IsoSeq data was used to inform the production of a more highly accurate set of gene annotations. Here, we present a greatly improved *C. zofingiensis* transcriptome and the identification of physical gene clusters in *C. zofingiensis*, using a method independent of functional annotation or coexpression data. We identified over 300 neighborhoods with potential functionally related neighbors including genes involved in carotenoid biosynthesis, photorespiration, nitrogen recycling, and oxidative stress responses. We also illustrate the value of conserved gene neighborhood identification for the discovery of gene function in algae with the discovery and testing of a novel arsenic detoxification pathway that evolved by co-opting glycolysis genes.

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.

Genome-scale Metabolic Model of *Chromochloris zofingiensis*, an Emerging Model Green Alga for Sustainable Fuel Production

Alexander Metcalf[1]* (metcalf@mines.edu), Yuntao Hu[2,3], Melissa S. Roth[3], Michelle Meagher[1], **Trent R. Northen**[2,4], Nanette 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 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. 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 laboratory is tasked with developing and utilizing a genome scale metabolic network reconstruction to predict intracellular carbon fluxes which will then be compared to fluxes measured experimentally using ^{13}C-MFA. The Northen group is focused on the exometabolomics and lipidomics analysis of the *C. zofingiensis* to help us understand the algal metabolite uptake preference and intracellular lipid synthesis 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) that has uses in the pharmaceutical, nutraceutical, and cosmetic industries[1-3]. It also demonstrates high levels of triacylglycerol accumulation and low photosynthetic productivity when additional organic carbon sources are provided[4], making it ideal for metabolic or genetic engineering focused on increasing algal lipid production.

In order to investigate the metabolic capacity of this organism for both biofuel and astaxanthin production, we generated a genome-scale metabolic network reconstruction. The current reconstruction includes 3500 metabolic reactions and 2832 metabolites. In order to formulate an accurate biomass formation equation, we measured both the macromolecule composition of *C. zofingiensis* (DNA, RNA, protein, lipid, carbohydrate) and the composition of each in photoautotrophic and photoheterotrophic growth modes. By combining these compositions with

growth curves and uptake measurements, we also predicted fermentation products and flux distributions. These predicted products are currently being experimentally validated.

To gain phenotypic data for model refinement, we conducted a *C. zofingiensis* time-course experiment, by measuring alterations in media composition resulting from algal growth and by measuring changes in algal lipid composition under different conditions. These experiments show a clear nutrient (carbon/nitrogen source) preference order during the growth of *C.zofingiensis.* Interestingly, we also observe utilization of organic compounds as energy and nitrogen sources under photoautotrophic and heterotrophic states. During these same transitions we find that the abundance of triacylglycerols increased up to 26-fold, whereas there are significant decreases in monogalactosyldiacylglycerols, digalactosyldiacylglycerols, and sulfoquinovosyl diacylglycerols, the major lipids that constitute thylakoid membranes. Additionally, we found an increase in phosphatidic acid but a decrease in phosphatidylinositol and phosphatidylserine. We are now incorporating these exometabolomic and lipidomic data to improve our understanding of the algal nutrient demand and lipid metabolisms, and the combination of the model and the data has great potential to elucidate dramatic metabolic shifts within the organism.

References

1. Hussein, G., Sankawa, U., Goto, H., Matsumoto, K. & Watanabe, H. Astaxanthin, a Carotenoid with Potential in Human Health and Nutrition. *Journal of Natural Products* **69**, 443-449 (2006).
2. Yuan, J. P., Peng, J., Yin, K. & Wang, J. H. Potential health-promoting effects of astaxanthin: A high-value carotenoid mostly from microalgae. *Molecular Nutrition & Food Research* **55**, 150-165 (2011).
3. Liu, J. *et al.* Chlorella zofingiensis as an Alternative Microalgal Producer of Astaxanthin: Biology and Industrial Potential. *Marine Drugs* **12**, 3487-3515 (2014).
4. Roth M.S. *et al*. Regulation of oxygenic photosynthesis during trophic transitions in the green alga *Chromochloris zofingiensis*. *Plant Cell*, tpc.00742.02018 (2019).

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018301

Machine Learning Guided Design of Safeguards That Operate Under Various Bacterial Physiologies

C.M. Mann[1], R. Weinberg[2], S. Forrester[2], G. Babnigg[2], PE. Larsen[2], MF. Gros[2], **A. Ramanathan**[1], and **P. Noirot**[2]* (pnoirot@anl.gov)

[1] Data Science and Learning Division, Argonne National Laboratory, Lemont, Il; [2] Biosciences Division, Argonne National Laboratory, Lemont, Il.

Project Goals: Our overarching goal is to build and train machine learning models that can predict the activity of the complexes between the CRISPR Cas9 nuclease and guide RNAs (Cas9-gRNA) in various genomes. To reach this goal, we are developing a machine learning approach that predicts the activity of Cas9-gRNA complexes based on features from the targeted DNA sequences and from the context of these targets, such as genome organization, annotation, and gene expression. The rules learned by the model will be transferred to another genomic context and validated experimentally. Finally, gRNAs that are highly efficient at killing bacteria will be used to develop safeguard systems.

The growing deployment of engineered organisms for environmental, bioenergy, and industrial applications represents an ever-increasing risk of releasing these organisms in the environment. To limit this risk, efficient biocontainment systems must be developed to prevent the survival of even a small number of released genetically engineered organisms. A safeguard system based on the controlled activation of a CRISPR nuclease, which breaks the chromosome at multiple targeted sites and kills the cell, is a design that is portable between organisms, including in non-model bacterial species of relevance in many environmental and biotechnological processes of interest to DOE. However, Cas9-gRNA complexes exhibit cleavage efficiencies that vary considerably along the genome [1, 2], limiting the use of CRISPR/Cas9 in safeguard systems.

We hypothesize that physiological conditions drive key factors that influence Cas9-gRNA activity in bacteria. To test this hypothesis, we are developing a machine learning approach, called CRISPRAct, to predict the activity of Cas9-gRNA complexes based on features from the targeted DNA sequences and from the context of these targets, such as genome organization, annotation, and gene expression. We are generating genome-wide Cas9-gRNA activity profiles in *E. coli* by screening a library of ~ 200,000 gRNAs under different physiological conditions. The datasets will be used to train and validate our CRISPRAct model to predict condition-specific Cas9-gRNA activity along the *E. coli* chromosome. The gRNAs predicted to be highly efficient at killing *E. coli* in different physiological conditions will be used to build safeguards.

The prediction of Cas9-gRNA activity not only as a function of DNA sequence but also including specific features from genome context and environmental conditions goes beyond how current models predict optimal gRNA design. These additional features will be key to apply a transfer learning approach to predict gRNA-Cas9 activity in other bacterial genomes. In this

project, we will apply transfer learning to predict Cas9-gRNA activity in *Pseudomonas fluorescens*. Our models will be made publicly available through an integrated Python software package. Our results and tools will open the way to the portable design of CRISPR-based safeguards in multiple non-model bacterial species.

References

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

This work is funded by the Department of Energy, office of Biological and Environmental Research.

Designing Synthetic communities for dissecting plant-microbe interactions in fabricated ecosystems (EcoFABs)

Kateryna Zhalnina,[1,2]*(kzhalnina@lbl.gov), Lauren K. Jabusch,[1] Peter F. Andeer,[1,2] Dawn M. Chiniquy,[1] Spencer Diamond,[3] Benjamin P. Bowen,[1,2] **N. Louise Glass,**[1,3] **Adam M. Deutschbauer**[1,3] and **Trent R. Northen**[1,2]

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]Joint Genome Institute, Berkeley; [3]University of California, Berkeley

https://mcafes.lbl.gov

Project Goals: Microbial Community Analysis and Functional Evaluation in Soils (m-CAFEs) uses fabricated ecosystems (EcoFABs) in combination with CRISPR-Cas and phage-based approaches for interrogating gene and microbial functions *in situ* to gain critical new insights into the rhizosphere thus advancing a mechanistic understanding of microbial ecology. We use 'bottom-up' defined microbial assemblies that enable detailed characterization of both constituent isolates and synthetic communities. This complements 'top-down' investigations of native soil-derived enriched microbial communities enabling extension of our approaches to more diverse communities that include uncultivated microbes. Predictive models will be developed and iteratively refined through integrated simulations and experimentation.

Plants release a large fraction of photosynthetically-derived carbon into the rhizosphere. The soluble metabolites and root biopolymers released by plants serve as primary carbon sources for supporting microbial growth resulting in the well-known "rhizosphere effect" in the soil surrounding its root. Metabolite exchange is thought to be important in both the recruitment of microbial communities to plant roots as well as a driver in the formation and stability of microbial communities. Already, several plant exudates involved in the putative recruitment of rhizosphere colonizing bacteria have been identified in the grass *Avena barbata*. However, the specific molecular mechanisms of microbial community assembly in the rhizosphere remain elusive. **Here we hypothesize that specific root exudate components are selectively used by rhizosphere bacteria *in situ*, enabling plant modulation of community structure using exudate composition**. **In this work we assemble and perturb defined rhizosphere communities to investigate metabolic networks, interactions, localization, and activities.**

To dissect plant-microbe interactions, we use EcoFABs technologies that enable precise and reproducible control and characterization at a level that is not yet possible in complex soil systems (https://eco-fab.org). We design synthetic communities (SynComs) of rhizosphere microbes, analyze chemical signaling between these SynComs and model grasses, and connect dynamics and activities of the microbial taxa to the plant growth phenotypes observed in EcoFABs.

Design of synthetic communities for investigation in EcoFABs.

To identify rhizosphere colonization by soil bacteria and plant phenotypes in response to this inoculation, we inoculated seedlings with a defined 105-member synthetic community (SynCom) and analyzed dynamics of the different microbial taxa within this SynCom in response to plant growth. We found that after a week of incubation in EcoFABs with introduced seedlings the rhizosphere SynCom was significantly enriched with specific taxa (e.g. *Dyella*, *Leifsonia* and *Burkholderia*). Liquid chromatography mass spectrometry-based metabolomics was used to analyze the modulation of plant exudates by the added SynCom. We observed that inoculation of plants with the SynCom increased relative abundances of *p*-coumaric acid and dehydroshikimic acid, while some primary metabolites (e.g. arginine) decresed when compared to the non-inoculated plants. We used our 105-member SynCom to test the dynamics of SynCom members in response to the added mix of aromatic acids, as well as consumption of these acids by SynCom isolates as C source. Correlation analysis between aromatic acid consumption and rhizosphere community restructuring shows the enrichment of several isolates in the SynCom via 16S rDNA gene sequencing correlated with uptake of specific aromatic acids.

This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) a Project led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231

Fabricated Ecosystems (EcoFABs) design for controlled and reproducible habitats to investigate plant-microbe-soil interactions

Dawn Chiniquy*[1] (dmchiniquy@lbl.gov), Lauren Jabusch[1], Peter Kim[2], Peter F. Andeer[1], Trenton K. Owens[1], Kateryna Zhalnina[1], Anup K. Singh[2], N. Louise Glass[1,3], Jenny Mortimer[1], Adam M. Deutschbauer[1] and Trent R. Northen[1]* (TRNorthen@LBL.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]Sandia National Laboratory, Livermore; [3]University of California, Berkeley;

https://mcafes.lbl.gov/

Project Goals: Microbial Community Analysis and Functional Evaluation in Soils (mCAFEs) will use fabricated ecosystems (EcoFABs) in combination with CRISPR-Cas and phage-based approaches for interrogating gene and microbial functions *in situ,* gaining critical new insights into the rhizosphere and advancing a mechanistic understanding of microbial ecology. We will use 'bottom-up' defined microbial assemblies that enable detailed characterization of both constituent isolates and synthetic communities. We will complement this with 'top-down' investigation of native soil-derived microbial community enrichments, expanding our approaches to more diverse communities that include uncultivated microbes. Predictive models will be developed and iteratively refined through integrated simulations and experimentation.

Most of what we know about microorganisms, including their gene functions and growth dynamics have been derived either from single-organism studies in benchtop experiments or from surveys from complex natural ecosystems. To bridge the gap between fully-constrained single organism laboratory experiments and high-dimensional ecosystem analyses, we have developed EcoFABs (Ecosystem Fabrication; www.eco-fab.org; https://doi.org/10.3791/57170) that are small plant growth chambers constructed using common microfluidic procedures that can be used to study and engineer microbiomes living on and around the root surface. The devices are designed to be amenable to a number of analyses including microscopy of the root zones and soils, metabolomics analyses, and spatial sampling of the microbial communities. Over a dozen different plant varieties have been grown in EcoFABs, and a number of growth conditions are supported including hydroponics, synthetic soil and field soils. A multi-laboratory EcoFAB study growing the model plant *Brachypodium distachyon* demonstrated reproducible plant phenotypes and metabolomics including significant differences between treatments that were retained across laboratories (https://doi.org/10.1111/nph.15662). This highly controlled, reproducible plant growth platform will allow for more detailed studies on the complex interactions that take place in the plant root-microbe interface.

Over eight different iterations on the original EcoFAB design have allowed for specialized investigations, including mycorrhizal plant interactions (MycoFAB), chemical and oxygen gradients (μEcoFAB), and *in situ* imaging of fluorescently labeled bacteria on plant roots (imaging EcoFAB). We have demonstrated reproducible rhizosphere assembly and similar microbial community structure between *B. distachyon* and switchgrass in EcoFABs, and have shown that EcoFABs can support *B. distachyon* through flowering and senescence for longitudinal studies. Additionally, we have performed both chemiluminescent imaging and multi-color fluorescent rhizosphere imaging with single bacterial cell resolution within this highly reproducible plant growth system.

This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) a Project led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231

m-CAFEs Applications of Targeted Editing in Microbial Networks

Matthew A. Nethery[1]*(manether@ncsu.edu), Claudio Hidalgo[1], Benjamin M. Rubin[2], Spencer Diamond[2], Trenton K. Owens[3], Jill Banfield[2], Jennifer A. Doudna[2], N. Louise Glass[3], Adam M. Deutschbauer[3], Rodolphe Barrangou[1], and **Trent R. Northen[3]**

[1]North Carolina State University, Raleigh; [2]University of California, Berkeley; [3]Lawrence Berkeley National Laboratory, Berkeley.

https://mcafes.lbl.gov/

Project Goals: Microbial Community Analysis and Functional Evaluation in Soils (mCAFEs) will use fabricated ecosystems (EcoFABs) in combination with CRISPR-Cas and phage-based approaches for interrogating gene and microbial functions *in situ* to gain critical new insights into the rhizosphere thus advancing a mechanistic understanding of microbial ecology. We will use 'bottom-up' defined microbial assemblies that enable detailed characterization of both constituent isolates and synthetic communities. This will be complemented by 'top-down' investigation of native soil-derived enriched microbial communities enabling extension of our approaches to more diverse communities that include uncultivated microbes. Predictive models will be developed and iteratively refined through integrated simulations and experimentation.

Through targeted editing, the m-CAFEs program seeks to elucidate the complex biological roles and functional genomics of individual members of microbial networks within the rhizosphere. Current methods for the targeted editing of mixed communities are bound by low delivery efficiencies, as well as the breadth of target specificities. To overcome these limitations, the m-CAFEs program is leveraging the characteristically high delivery efficiencies and narrow host range of bacteriophages to deliver DNA-targeting CRISPR-Cas endonucleases to select microbes of interest. Through the characterization of ablation efficiency and host specificity of various phages, as well as the isolation of novel phages, we have assembled a library of candidate phages for the manipulation of our synthetic community. We demonstrate the efficiency of phage-based ablation both *in vitro* and within EcoFAB devices and show that these phages can successfully target a host of interest without perturbing other members of the community. Further, sequencing candidate phages within our library, we have identified target loci for the engineering of CRISPR-Cas endonucleases, a critical step toward the implementation of targeted editing. Continued development of this platform will extend our capabilities beyond targeted ablation to enable the addition of desirable traits for functional enhancement of rhizosphere communities, which could be of particular significance to the evaluation of plant fitness in future models. Although these tools will have an immediate impact on the research of synthetic communities, the ultimate utility of these tools lies in its translational application to uncultivable microbial communities, facilitating the functional characterization of microbes that cannot be readily isolated.

This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) a Project led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231

Targeted DNA Editing Within Microbial Communities

Benjamin E. Rubin[1] (brubin@berkeley.edu), Spencer Diamond[1], Brady Cress[1], Claudio Hidalgo[2], Matthew Nethery[2], Trenton K. Owens[3], Christine He[1], Alex Crits-Christoph[1], Zeyi Zhou[1], Kimberly Tang[1], Dylan C. Smock[1], N. Louise Glass[3], Rodolphe Barrangou[2], Jill Banfield[1], Adam M. Deutschbauer[3], and Jennifer A. Doudna[1], **Trent R. Northen[3]** (trnorthen@lbl.gov).

[1]University of California, Berkeley; [2]North Carolina State University, Raleigh; [3]Lawrence Berkeley National Laboratory, Berkeley.

http://m-CAFEs.lbl.gov

Project Goals: Microbial Community Analysis and Functional Evaluation in Soils (mCAFEs) will use fabricated ecosystems (EcoFABs) in combination with CRISPR-Cas and phage-based approaches for interrogating gene and microbial functions *in situ* to gain critical new insights into the rhizosphere thus advancing a mechanistic understanding of microbial ecology. We will use 'bottom-up' defined microbial assemblies that enable detailed characterization of both constituent isolates and synthetic communities. This will be complemented by 'top-down' investigation of native soil-derived enriched microbial communities enabling extension of our approaches to more diverse communities that include uncultivated microbes. Predictive models will be developed and iteratively refined through integrated simulations and experimentation.

Our understanding of microbial life has largely been achieved by observing the effect of targeted genetic manipulations on isolated microbial species in laboratory settings. Unfortunately, this provides limited insights into the complex and societally relevant communities of microorganisms that exist in nature. We have created a generalizable toolset for targeted genetic manipulation within laboratory microbial communities that will allow microorganisms to be altered and studied in a community context. As a first step, we have developed environmental transformation sequencing (ET-Seq) to determine *in situ* what microbes can be edited within a community with what efficiency. The roadmap for tractability produced by ET-Seq is then used to guide CRISPR-Cas based targeted editing within the community. Editing events have currently been measured *in situ* in a 9 member synthetic soil community, and an ~60 member thiocyanate degrading community. The ability to make targeted DNA perturbations to microbial communities will open up the field of *in situ* microbial genetics which will greatly improve our basic science understanding of microbial genomics and microbial interactions.

This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) a Project led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231

Methane and nutrient cycling by sediment-hosted archaeal-bacterial syntrophic consortia and their viral predators

Grayson Chadwick,[1] Ranjani Murali,[1] Kyle Metcalfe,[1] Hang Yu[1], Rex Malmstrom,[2] Tanya Woyke,[2] Aditi Narayanan, [1] Alon Philosof, [1] Yongzhao Guo,[1] Fabai Wu,[1] Mark Ellisman,[3] and **Victoria J. Orphan**[1*] (vorphan@gps.caltech.edu)

[1]California Institute of Technology, Pasadena, California; [2]Joint Genome Institute, US Department of Energy, California; [3]University of California, San Diego, California

http://orphanlab.caltech.edu

Project Goals: Our overarching goal is to develop a systems-level understanding of the interactions and fundamental activities by syntrophic methane-oxidizing archaeal-bacterial consortia and viral assemblages involved in cycling of C and nutrients in anoxic sedimentary environments. As part of this work, we are developing single cell and virus-targeted analytical and metagenomics techniques in tandem with high resolution imaging methods, stable isotope analysis, and modeling for application both in the laboratory and field settings.

Microbially-mediated anaerobic oxidation of methane (AOM) represents a significant methane sink in anoxic sediments. This process is catalyzed primarily by consortia of methanotrophic 'ANME' archaea and syntrophic bacteria coupled with the reduction of sulfur, nitrogen, iron, and manganese. These methane-consuming consortia are also involved in the transformation of essential nutrients, through processes including nitrogen fixation and intracellular storage of iron, carbon, and phosphate. The nutritional requirements and interdependencies of different methanotrophic archaeal-bacterial partnerships and their collective impact on nutrient cycling within the sedimentary ecosystem is poorly understood and represents a major focus of this work. Viruses are also important mediators of nutrient cycling in diverse environments. As selective agents of microbial mortality, viruses have been shown to enhance the bioavailability of essential nutrients such as N and P and stimulate microbial growth. The degree to which viruses influence syntrophic AOM consortia and their potential role in transforming methane-derived carbon and nutrients within sediment communities has not yet been investigated. Here we are developing a multi-modal analytical imaging pipeline and comparative 'omics datasets for studying the ecophysiology of diverse uncultured ANME-bacterial consortia and associated viral predators in sediment ecosystems.

Comparative 'omics analysis of metagenome assembled genomes and BONCAT-based single consortia sequencing show notable differences in gene content and predicted physiologies of different ANME and associated syntrophic bacterial lineages, including differences in the predicted ability to fix nitrogen among the syntrophic bacterial partners of ANME. Application of mRNA-FISH and FISH-nanoSIMS analyses with $^{15}N_2$ support the involvement of syntrophic sulfate-reducing bacteria and ANME archaea in nitrogen fixation, but active diazotrophy appears to be unevenly distributed among archaeal and bacterial partners within individual consortia. Initial analyses of viral activity and morphological diversity within methane saturated sediments using viral-BONCAT and TEM surveys revealed active viral-like particle production and distinct morphotypes in sediments with active ANME-SRB consortia compared to sediments where only

the ANME were active, with an increase in ‘lemon-shaped’ morphologies reminiscent of viruses infecting archaea.

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]

Microbial Interactions at Micro-scale and Pore-scale Revealed by Process-based Reactive Transport Modeling

He, Xiaojia[1]* (xiaojia.he25@uga.edu), Jung, Heewon[1], Chadwick, Grayson[2], **Kempes, Christopher**[3], **Orphan, Victoria**[2], and **Meile, Christof**[1]

[1]University of Georgia, Athens, GA, USA; [2]California Institute of Technology, Pasadena, CA, USA; and [3]Santa Fe Institute, Santa Fe, NM, USA

https://faculty.franklin.uga.edu/cmeile/microbial-metabolism

Project Goals

The overarching scientific goal of this multidisciplinary research project 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.

The focus of this work is the metabolism of AOM consortia and microbial interactions with the environment. We present a three-dimensional model that simulates microbial activities in methane-oxidizing consortia and is validated using co-registered FISH-nanoSIMS observations [1]. Model results show that direct interspecies electron transfer (DIET) between archaeal and bacterial partners yields cell-specific activities that are consistent with observations, with little impact of the spatial distribution of bacterial and archaeal cells at commonly observed aggregate sizes (diameter 3 - 25 μm). Next, we explore the controls on interspecies electron transport of anaerobically methane-oxidizing consortia at a larger aggregate size [2]. Our simulations of metabolic interactions through DIET showed that ohmic resistance and activation loss are the two main factors causing the declining metabolic activity, where activation loss dominated at distance < ~6 μm. These simulations indicated that bacterial cells remain metabolically active at distance < ~30 μm from the archaea-bacteria interface, suggesting a maximum spatial distance between segregated syntrophic partners.

Moreover, we further expand our understanding of extracellular electron transfer in electroactive biofilms [3]. A one-dimensional reactive transport model representing cellular metabolism across a *Geobacter sulfurreducens* biofilm growing on an electrode is established to synthesize existing knowledge and provide a quantitative framework of the extracellular electron transfer that may guide further experimental studies. The model is able to reproduce high-resolution activity measurements under different anode potentials. At high anode potential (+0.24V), two metabolic activity peaks - one near the electrode and another one further away from the electrode – are simulated, consistent with observations using nanoSIMS. Our model attributes this to H^+ accumulation close to

the electrode, negatively impacting metabolic activity. The second peak approx. ~12 μm um away from the anode surface, is attributed to a shift between two redox-active systems allows *G. sulfurreducens* cells to respond to external electric potential.
Lastly, the role of heterogeneous distribution of microbial aggregates at the pore scale for upscaled microbial reaction rates is investigated. Our pore-scale reactive transport simulations reveal that the accuracy of macroscopic rate estimates depend strongly on the flow conditions and reaction kinetics, and to a lesser extent on the distribution of microbial aggregates [4]. Our modeling efforts further evaluate the physicochemical conditions suitable for microbial processes mediated by signaling molecules [5]. Modeling results show that advection dilutes signaling molecules so that faster flow conditions require higher microbial densities, faster signal production rates, or higher sensitivities for effective communication through signaling molecules. We present approximate analytical solutions of the advection-diffusion-reaction equation that allow to quantitatively estimate the effective communication distances amongst multiple microbial aggregates without further numerical simulations.

References

1. He, X., Chadwick, G., Shi, Y., Orphan, V., Kempes, C. and Meile, C. 2019 *Microbial interactions underlying anaerobic oxidation of methane: Constraints on mechanisms from process rates, activity patterns and modeling*. Environmental Microbiology, 21(2): 631-647.
2. He, X., Chadwick, G., Orphan, V., Kempes, C. and Meile, C. *Controls and size limitations of anaerobically methane oxidizing consortia metabolism*. In preparation.
3. He, X., Chadwick, G., Orphan, V. and Meile, C. *Spatially resolved electron transport through anode-respiring Geobacter biofilms: controls and constraints.* In preparation.
4. Jung, H. and Meile, C., 2019. *Upscaling of microbially driven first-order reactions in heterogeneous porous media*. Journal of Contaminant Hydrology, 224:103483.
5. Jung, H. and Meile, C., 2020. *Mathematical investigation of microbial quorum sensing under various flow conditions*. bioRxiv. doi:10.1101/2020.01.09.900027

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 Awards Number DE-SC0016469 and DE-SC0020373. The authors declare no conflict of interest.

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), Zi Ye,[1] John I. Hendry[2], Hoang V. Dinh[2], Debolina Sarkar[2], Lin Wang[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.

Abstract

Anabaena sp. ATCC 33047 is a fast-growing heterocystous cyanobacterium that utilizes high light and exhibits the highest recorded conversion rate of CO_2 into biomass in an oxygenic photosynthetic organism[1]. Unlike most diazotrophic strains, this cyanobacterium has the ability to maintain similar growth rates both in the presence and absence of combined N_2 sources and displays a doubling time as short as 3.8 h under photoautotrophic and N_2-fixing conditions. These characteristics render *Anabaena* 33047 an attractive platform for cost effective production of nitrogen-rich compounds. However, this strain was known to be genetically intractable and hence not widely studied. In the initial phase of our project we developed a genetic manipulation system that enabled us to make targeted modifications in its genome. We also focused on the development of a genome scale metabolic model for *Anabaena* 33047.

Nitrogenase activity in heterocystous cyanobacteria rely on ATP generated by PSI-mediated cyclic phosphorylation in the heterocysts and light is essential for this process. Under N_2 limited growth conditions, cyanobacteria exhibit an initial degradation of phycobilisomes (PBSs), the major light harvesting complex. This helps to maintain a more balanced C/N ratio under N_2 limitation by reducing photosynthetic activity of the cells and also ensures mobilization of cellular nitrogenous resources until the heterocysts engage in fixing nitrogen. PBS is resynthesized in the vegetative cells after fixed nitrogen becomes available but its fate in heterocysts remains poorly understood. The degradation of PBS is mediated by the NblA protein. In the *nblA* deficient strain of *Anabaena* sp. PCC 7120, nitrogen step down does not trigger PBS degradation and does not affect growth[2]. To assess the effect of this deletion on N_2 fixation in a high light tolerant strain, we engineered a Δ*nblA* mutant of *Anabaena* 33047. Intriguingly, under high light intensities, the Δ*nblA* mutant exhibited up to 2.5-fold higher rates of nitrogenase activity compared to the wild

type. Spectroscopic analysis shows higher PSI activity in the mutant possibly aided by higher levels of PBS in the mutant heterocysts. Thus, retaining higher levels of PBS appears to be an effective strategy to enhance rates of N_2 fixation in this high light-tolerant, fast growing diazotrophic strain.

We have also developed a comprehensive genome scale metabolic model for *Anabaena* 33047 by mining annotation from diverse data bases such as KEGG[3], MetaCyc[4] and ModelSEED[5] and a recently published model for the closely related *Anabaena* 7120[6]. A species-specific biomass equation was formulated based on biomass composition of *Anabaena* 7120[7]. To account for the heterocystous lifestyle of this species, the model accounts for two different cell types: vegetative cell (950 reactions) and heterocyst (942 reactions). The model accounts for 892 genes and 953 unique reactions. The vegetative cell performs photosynthesis but lacks the nitrogenase enzyme. The heterocyst performs N_2 fixation but lacks the capacity to perform oxygenic photosynthesis. The model was able to predict the growth rate under N_2 fixing and N_2 sufficient conditions. Under both the conditions, the PSI/PSII flux ratio predicted by the model was above 1.2 as observed for cyanobacteria in general. The model is being used to design metabolic engineering strategies for the overproduction of nitrogen containing compound valerolactam.

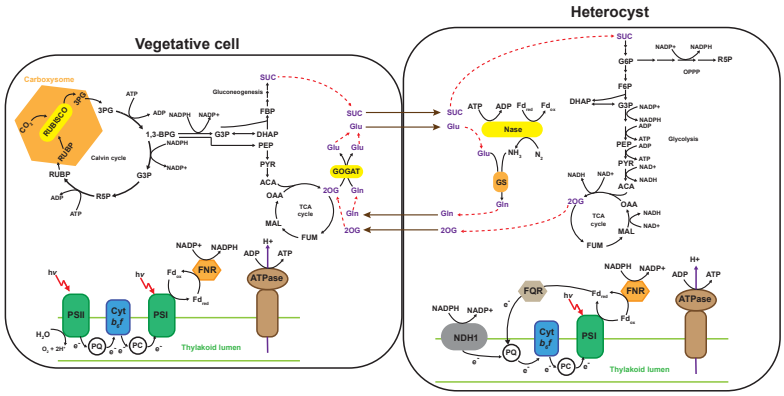


Figure 1: Two cell model of *Anabaena* 33047. The genome scale metabolic model, iAnC915, has two super-compartments, the vegetative cell and the heterocyst, in-order to capture the diazotrophic and heterocyst forming life style of this cyanobacterium.

References

1. Gonzalez Lopez et al., Utilization of the cyanobacteria *Anabaena* sp. ATCC 33047 in CO_2 removal processes. Bioresource technology. 2009;100(23):5904-10.
2. Baier et al., NblA is essential for phycobilisome degradation in *Anabaena* sp. strain PCC 7120 but not for development of functional heterocysts. Microbiology. 2004; 150:2739–49.
3. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32: D277-280
4. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, Paley S, Subhraveti P, Karp PD (2018) The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res 46: D633-D639
5. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28: 977-982
6. Malatinszky D, Steuer R, Jones PR (2017) A comprehensively curated genome-scale two-cell model for the heterocystous cyanobacterium *Anabaena* sp. PCC 7120. Plant Physiol 173: 509-523
7. Vargas M, Rodriguez H, Moreno J, Olivares H, Campo JD, Rivas J, Guerrero M (1998) Biochemical composition and fatty acid content of filamentous nitrogen-fixing cyanobacteria. Journal of Phycology 34: 812-817

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

Modeling growth kinetics and metabolism of *Clostridium acetobutylicum*/*Clostridium ljungdahlii* co-culture with cell fusion

Charles J. Foster[1]* (cjf33@psu.edu), Kamil Charubin,[2,3] Eleftherios Terry Papoutsakis,[2,3] and **Costas D. Maranas**[1]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA

[2]Department of Chemical and Biomolecular Engineering, University of Delaware, 15 Innovation Way, Newark, DE 19711, USA

[3]Delaware Biotechnology Institute, University of Delaware, 15 Innovation Way, Newark, DE 19711, USA

http://www.maranasgroup.com/

http://www.papoutsakis.org/

Project Goals: The goal of this project is to develop syntrophic *Clostridium* co-culture systems, involving three *Clostridium* organisms, for producing C4-C8-chain length metabolites that can be used as chemicals or serve as biofuels and biofuel precursors. Part of the effort includes modeling the behavior of these triple co-cultures. To build this capability, we first model the binary co-culture of *Clostridium acetobutylicum (C. ac)* with *C. ljungdahlii* (*C. lj*). This first modeling sub-goal then is to develop a method for modeling the co-culture growth kinetics that accounts for novel cell fusion events observed in this binary co-culture. Using the resulting kinetic model in conjunction with a community genome-scale metabolic model and the SteadyCom community modeling framework, we aim to explain the observed co-culture metabolism/redox state, and identify genetic intervention strategies maximizing production of isopropanol and 2,3-butanediol.

Clostridia organisms have been of interest for decades due to their ability to ferment a wide range of carbon sources to useful bioproducts. The metabolic repertoire of these anaerobes has been further expanded in co-cultures due to the diversity of substrates they can consume and unexpected syntrophic behaviors that are still being discovered. One such example lies in the syntrophic co-culture of *Clostridium acetobutyllicum* (*C. ac*) and *C. ljungdahlii* (*C. lj*). In addition to the discovery of an upregulation of *C. lj sadh* and *23bdh* gene expression in the presence of *C. ac* (allowing *C. lj* to convert acetone and acetoin produced by *C. ac* to isopropanol and 2,3-butanediol, respectively) [1], *C. ac* and *C. lj* cells were recently shown to fuse membranes and exchange proteomes. This work aims to characterize the resultant change in growth kinetics due to the observed fusion/protein exchange event using a kinetic model which characterizes both the pure and mixed-proteome *C. ac* and *C. lj* growth rates and the cell fusion/proteome exchange rate. The parameterized kinetic model is used to inform the construction of a community genome-scale metabolic model of pure and mixed-proteome *C. ac* and *C. lj* cells using the SteadyCom framework, and characterize the dynamic shift in co-culture metabolism and redox state related to the observed fusion event required to support the experimentally measured isopropanol and 2,3-butanediol production. Single organism strain design tools (i.e. optKnock, optForce) are being adapted to support the inclusion of multi-organism models in order to understand how the *C. ac* and *C. lj* genomes can be manipulated to maximize the production of fermentation products of interest (i.e isopropanol, 2,3-butanediol) under the newly discovered *C. ac*/*C.lj* co-culture conditions.

References

1. Charubin K, Papoutsakis ET. Direct cell-to-cell exchange of matter in a synthetic Clostridium syntrophy enables CO2 fixation, superior metabolite yields, and an expanded metabolic space. Metab Eng. 2019;52:9-19.

SUPPORTED by the U.S. Department of Energy (Award No. DE-SC0019155).

Syntrophic co-cultures of *Clostridium* organisms to produce higher alcohols and other C6-C8 metabolites

Kamil Charubin[1,2,*](kamilcha@udel.edu), Charles Foster[3], Jonathan Otten[1,2], Michael Dahle[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.

Clostridium organisms are of major importance for developing new technologies to produce biofuels and chemicals. Three major types of *Clostridium* organisms have been the focus of studies for the sustainable production of fuels 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.[1] Acetogenic clostridia can fix inorganic H_2, CO_2, and CO to generate C2 acids and alcohols.[1] 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.[1]

The majority of previous work with clostridia focused on optimizing single-organisms systems for production of biochemical. In nature, microorganisms live in complex communities where syntrophic interactions result in superior resource utilization. Here, we first examined a synthetic syntrophy consisting of the solventogen *Clostridium acetobutylicum*, which converts simple and complex carbohydrates into a variety of chemicals, and the acetogen *C. ljungdahlii*, which fixes CO_2.[2] 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 syntrophic co-culture produced robust metabolic outcomes over a broad range of starting population ratios of the two organisms. Finally, 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.[2] Furthermore, 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 *C. acetobutylicum* and *C. ljungdahlii* expressing fluorescent FAST[3] and HaloTag® proteins, respectively. When co-cultured together both fluorescent strains showed the whole-cell exchange of protein material, a unique and never-observed before phenomena.

To expand our synthetic co-culture system, we will form a triple co-culture including *C. kluyveri*, which can metabolize ethanol and acetate to produce C6 and C8 carboxylic acids. Both *C. acetobutylicum* and *C. ljungdahlii* produce ethanol and acetate, which makes *C. kluyveri* and ideal partner for a triple synthetic co-culture system capable to converting biomass-derived carbohydrates to C6 and C8 biochemicals. ^{13C}-based Metabolic Flux Analysis (MFA) will be used to gain insight into the regulation of cell growth and product formation pathways, and to identify metabolic bottlenecks. Currently, use of stable-isotope (e.g. ^{13}C) tracers combined with measurements of isotopic labeling by mass spectrometry represents the state-of-the-art in flux determination. Metabolic fluxes will be studied using ^{13}C MFA in *C. kluyveri*, *C. acetobutylicum*, and *C. ljungdahlii* under mono- and co-culture conditions to identify key changes in metabolism of each organism. The consortium model consisting of *C. kluyveri*, *C. acetobutylicum*, and *C. ljungdahlii* will be constructed using the SteadyCom framework. This will be done by standardizing the biomass equations and metabolite naming conventions for existing genome-scale models (GSMs), and updating each GSM using RNAseq and ^{13}C-fluxomics procured under varying experimental conditions. ^{13}C-fluxomics and RNAseq data will be used to infer regulatory events in each organism to simulate and compare transient monoculture and co-culture population dynamics.

SUPPORTED by the U.S. Department of Energy (Award No. DE-SC0019155).

Publications

1 Charubin, K., Bennett, R. K., Fast, A. G. & Papoutsakis, E. T. Engineering Clostridium organisms as microbial cell-factories: challenges & opportunities. *Metabolic engineering* **50**, 173-191, (2018).

2 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, (2019).

3 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, (2019).

Symbiotic niche mapping reveals nutrient specialization and functional complementarity among ectomycorrhizal fungi

Michael Van Nuland[1], and **Kabir G. Peay**[1]* (kpeay@stanford.edu)

[1]Department of Biology, Stanford University, Stanford CA 94305-5020

Project Goals: Despite strong geographic patterns in the dominant form of mycorrhizal symbiosis[1] 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 early career award (ECA) I am determining the mechanisms by which mycorrhizal symbiosis influences the distribution of tree species across North America.

Abstract text. Mutualisms are ubiquitous in natural systems, but research on ecological interactions has focused almost exclusively on antagonism and we know little about how these positive interactions influence species distributions. The niche concept is one useful approach for thinking about factors that control species distributions but has generally been considering without recognizing the growing importance of mutualisms. In particular, understanding how the nature of positive interactions between plants and mycorrhizal fungi change across large environmental gradients remains a significant research frontier. Here, we used a continuous niche mapping approach to examine how ectomycorrhizal fungi impact plant growth across a two-dimensional soil nitrogen (N) and phosphorus (P) gradient. We found that one ectomycorrhizal fungus, *Thelephora terrestris*, improved seedling growth most at high N:P ratios, suggesting that members of the Thelephoraceae may be P specialists. By contrast, the presence of a second ectomycorrhizal fungus (*Suillus pungens*) improved seedling growth most at the lowest nutrient levels and N:P ratios. Mycorrhizal colonization by *T. terrestris* increased plant niche volume (calculated as the volume of convex hulls comprising total plant biomass across N and P gradients) compared to non-mycorrhizal control plants and shows the positive effects of mutualisms on plant niche size. However, despite growth benefits at low nutrient conditions, the presence of both fungi decreased plant niche volume compared to *T. terrestris* alone, indicating the costs of maintaining multiple mycorrhizal symbioses exceed benefits in some environments. The niche mapping approach we present has the potential to answer fundamental questions about the dimensions of functional diversity in ectomycorrhizal fungi and the global distribution of mycorrhizal symbioses.

References

1. Steidinger, BS, Crowther, TW, Liang, J, Van Nuland, ME, Werner, GDA, Reich, PB, Nabuurs, G, de-Miguel, S, Zhou, M, Picard, N, Herault, B, Zhao, X, Zhang, C, Routh, D, GFBI Consortium, & Peay, KG (2019)

Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* **569**, 404-408

Funding statement. This research was funded by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research, Early Career Research Program, Award Number DESC0016097.

A Workflow for Generating and Polishing Nanopore Reads from Low Biomass Samples

Olivier Zablocki[1*] (Zablocki.4@osu.edu), Michelle Michelsen[2], Marie Burris[1], Natalie Solenenko[1], Romik Ghosh[1], **Jennifer Pett-Ridge**[3], Ben Temperton[2], and **Matthew Sullivan**[1]

[1]The Ohio State University, Columbus; [2]School of Biosciences, University of Exeter, UK; [3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA

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.*

Numerous technological and analytical advances have caused a revolution in the life sciences and revealed that microbes and their viruses represent hidden drivers of the nutrient and energy currencies that fuel our planet and our bodies. Arguably, the most impactful advance in the past decade has been the ability to view these micro- and nano-scale entities via sequencing rather than cultivation-based approaches—giving use a significantly broader window into the functional capabilities and interactions of microbe-based communities. Though powerful, current metagenomic sequencing approaches are limited in that short-read assemblies likely miss microdiverse populations and niche-defining hypervariable genomic islands, and routine long-read sequencing requires high-biomass inputs and has high error rates.

Here we optimized laboratory and informatics protocol of our VirION[1] approach to generate, analyze and polish Oxford Nanopore long reads (>10kb) from low (1ng) biomass samples. For method optimization, we used a three phage mock community of known genomes (size range: 38-130kb) and evaluated DNA extraction and sequencing library preparation options including: choice of DNA polymerase, DNA shearing size, number of PCR amplification cycles, input DNA amount and DNA cleanup strategies. This revealed that sheared DNA (15kb size) that was amplified by LA Takara (long-range) polymerase yielded the most consistent and significant increase in read lengths compared to other treatments. The optimized protocol achieved a median read length of 7095bp (up to 67kb), which represents a ~5,000 bp increase over that documented in the official Nanopore kit. The number of amplification cycles tested (15, 18, 20 and 22) did not significantly alter the number of chimeric reads produced (<1% in all treatments), which enables much lower input biomass than in the official Nanopore kit. Together these optimized protocols are now being applied to Hopland soils for this SFA project to better document short-read recalcitrant viral OTUs and their niche-defining hypervariable genomic islands.

References

1. Warwick-Dugdale, J. *et al.* Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* **7**, e6800 (2019).

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 Ohio State University. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

How Drought Modulates Formation and Persistence of Microbial-Derived Soil Carbon from Rhizosphere, Detritusphere, and Bulk Soil Microbial Communities

Noah Sokol*[1] (sokol1@llnl.gov), Megan Foley[2], Alex Greenlon[3], **Bruce Hungate**[2], **Jill Banfield**[3], **Steve Blazewicz**[1], **Mary Firestone**[3], **Jennifer Pett-Ridge**[1]

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory; [2]Center for Ecosystem Studies, Northern Arizona University; [3]Department of Environmental Science, Policy, and Management, University of California, Berkeley

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 dominant 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.*

Microbial residues are dominant ingredients of persistent soil organic matter (SOM). Via the 'microbial carbon pump,' plant carbon is processed by the microbial community *en route* to microbial-derived, mineral-associated pools of SOM—which are critically important to the Earth's 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 regions of the soil with distinct microbial communities (i.e. the rhizosphere, detritusphere, and bulk soil) and under different moisture conditions. Our SFA team is using stable-isotope labeling techniques to synthesize microbial ecology (via metagenome sequencing) with measurements of SOM formation and persistence under varying moisture regimes.

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 carbon pools under normal moisture vs drought conditions. Using a continuous $^{13}CO_2$-labeling system, we grew the Mediterranean annual grass *Avena barbata* in controlled growth chambers and measured the formation of organic matter 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 soil carbon pools and microbial community dynamics. In a second set of microcosms, we tracked the formation of soil carbon 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, seperately tracking the contributions from each root C source using a reciprocal ^{13}C-labeling design.

Here, we present initial data from the greenhouse experiment and outline our broad experimental goals. In all microcosms, our soil moisture manipulations generated significant differences in drought (8 ± 2%) and 'normal moisture' (15 ± 4.2 %) treatments. The magnitude of this difference increased through time, and manifested as differences in soil respiration, as well as in aboveground plant biomass, plant height, root architecture. We have also extracted DNA from rhizosphere, detritusphere, rhizosphere + detritusphere, and bulk soil communities, and are measuring a range of microbial traits on these same communities, including carbon use efficiency, growth rate, and the production of extrapolymeric substances. The overarching aims of this project will be to determine how microbial community composition and microbial community assembly through time are influenced by soil moisture status, and to connect these community-level differences with the dominant microbial traits that may affect the formation, chemical composition, and long-term persistence of microbial-derived, mineral-associated SOM.

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.

Trait-based Modeling of Mineral-associated Soil Organic Matter Formation in Distinct Soil Habitats

Gianna Marschmann[1*](glmarschmann@lbl.gov), Jinyun Tang[1], and **Eoin Brodie**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, California

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 will characterize 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. The ultimate goal of our SFA "Microbes Persist: Systems Biology of the Soil Microbiome" 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.

Trait-based soil C models are a useful tool for exploring the interactions between genomic traits and environmental conditions that drive variation in ecological strategies of microorganisms. Here, we use Equilibrium Chemistry Approximation kinetics and Dynamic Energy Budget Theory to model relationships between consumer identity, substrate preference, and bacterial growth efficiency of microbial guilds specialized for different soil habitats (rhizosphere, bulk soil). We synthesize genome-informed trait data of soil isolates, literature-based allometric scaling relations, and biophysical modeling approaches to constrain the model parameter space for plant-derived carbohydrate depolymerization, low molecular weight carbon uptake, and cell metabolism. Scenario simulations are used to investigate the interaction between plant substrate identity and soil microbial density on the formation of mineral-associated soil organic matter with focus on model sensitivities to microbial carbon use efficiency and biomass stoichiometry. Our allometric framework allows quantification of the influence of cell morphology on resource-based niche parameters and relevant microbial functional traits across different environmental conditions, thus providing a model-based link between microbial phylogeny, specific genes, phenotypic traits and environmental preferences.

References

1. Tang, Jinyun, and William J. Riley. "A theory of effective microbial substrate affinity parameters in variably saturated soils and an example application to aerobic soil heterotrophic respiration." *Journal of Geophysical Research: Biogeosciences* 124.4 (2019): 918-940.

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 for the "Microbes Persist: Systems Biology of the Soil Microbiome" SFA, and a subcontract to Lawrence Berkeley National Lab where work was performed under contract DE-AC02-05CH11231.

Unearthing the Active Microbes, Viruses and Metabolites in Dynamic-Redox Tropical Soils with Quantitative SIP and Metagenomics

Jennifer Pett-Ridge[*1] **(pettridge2@llnl.gov),** Ashley Campbell[1], Rachel Hestrin[1], Gareth Trubl[1], Amrita Bhattacharyya[2], Yang Lin[3], Ben Bowen[2], Trent Northen[2], Jeffrey A. Kimbrel[1]

[1]Lawrence Livermore National Laboratory, Livermore CA, [2]Lawrence Berkeley National Laboratory, Berkeley CA, [3]University of California Berkeley, Berkeley CA

https://pls.llnl.gov/people/staff-bios/nacs/pettridge-j

Project Goals: This Early Career research examines the genomic potential and activity of tropical soil microorganisms as they experience shifts in soil temperature, moisture, depth and oxygen availability. Associated fluctuations in redox potential are proximal controls of mineral-organic matter interactions in humid tropical soils. By tracking the degradation and fate of organic ^{13}C labeled compounds during shifts in soil redox status, this work will improve our understanding of microbial metabolic flexibility, and how microbial processes affect the fate of organic carbon in wet tropical systems. The mechanistic understanding produced by this research will also improve the predictive capacity of mathematical models that forecast future tropical soil carbon balance.

In soils, anaerobic processes are dynamic and widespread—even in bulk-oxygenated upland environments—and exist within soil aggregates, near perched water tables, and in zones of abundant labile C. However, the drivers and dynamics of these anoxic volumes remain poorly constrained, particularly in wet tropical soils where we know little about the metabolic capacities of soil microorganisms, and the mineral-organic matter relationships that regulate many aspects of soil C cycling. Using a 44 day redox manipulation experiment with soils from the Luquillo Experimental Forest, Puerto Rico, we examined patterns of tropical soil microorganisms and metabolites when soils were exposed to different redox regimes - static oxic, static anoxic, high frequency redox fluctuation (4 days oxic, 4 days anoxic), or low frequency redox fluctuation (8 days oxic, 4 days anoxic). Replicate microcosms were harvested throughout the incubation to measure the impact of redox condition on microbial community structure and activity, organic matter turnover, and soil metabolites. An addition of ^{13}C enriched plant biomass allowed us to distinguish decomposition of fresh plant litter vs native organic matter and conduct Stable Isotope Probing (SIP) with genome resolved metagenomics and viromics to identify active microorganisms and their viruses.

Our amplicon data show that bacterial and fungal community composition in the two fluctuating redox treatments was indistinguishable from the native soil community, while the static redox communities were distinct, suggesting the microbes in these soils are highly adapted to dynamic redox conditions. Using differential abundance analysis, we found that fluctuating redox enriched for relatively more bacterial and fungal taxa –compared to the static redox conditions. However, the anoxic treatment had a distinct iron-cycling microbial community relative to the other treatments. The majority of taxa adapted a facultative strategy in the first weeks of the incubation (when litter decomposition activity was highest) –suggesting they maintained mechanisms to tolerate inhospitable redox periods. However, by the end of the experiment, labile

C had become limiting, and obligate anaerobes had increased in their relative abundance. Using ^{13}C quantitative SIP of over 1100 16S rRNA libraries, coupled to CO_2 and DOC measurements, we measured reduced microbial carbon use efficiency (CUE) under static redox conditions compared to fluctuating redox conditions. SOM-C respiration was highest under static oxic conditions, litter-C respiration was highest under static anoxic conditions, and litter-C assimilation was highest under fluctuating redox conditions. Intriguingly, some taxa remained active (i.e., assimilating litter-derived C) under all redox conditions, and abundance was not always correlated with activity; many relatively 'rare' taxa assimilated a high ratio of litter C.

We also analyzed 95 metagenomes (85 SIP fractions and 10 bulk samples) from our 4 redox treatments, using sequences generated by the JGI (22 billion reads and over 3.3 trillion base pairs). Metagenome assemblies produced over 6,000 genome bins (MAGs), and co-assemblies produced 326 medium-to-high-quality MAGs from ^{13}C enriched DNA fractions. Over the 44 day incubation, we saw large differences in the active (^{13}C enriched MAGs) from the static anoxic vs static oxic soil, and a strong response in the Fe-reducer community. Overall, the fluctuating soil MAGs had generally more ^{13}C incorporation, particularly in Proteobacteria and Actinobacteria. In the static anoxic soils, we only observed ^{13}C incorporation by Bacteroidetes and Firmicutes.

Viruses were detected with VirSorter and VirFinder and clustered into viral populations (vOTUs). Active vOTUs were identified as those present in ^{13}C samples but not the paired ^{12}C sample. We recovered nearly 48,000 viruses clustering into 640 vOTUs >10 kb. SIP-fractionated samples recovered ~6% vOTUs that were not observed in the un-fractionated bulk samples. Viral diversity was highest in the oxic samples and decreased as follows: oxic>high frequency>low frequency>anoxic. Beta diversity suggests fluctuation between oxic and anoxic conditions had the largest impact on the active viral community structure. Only 27% of the vOTUs were active, with 33% active in all treatments, and 16% only active in the anoxic samples. Nearly 21,000 genes were predicted from the vOTUs, 15% of which had a known function and of the 85% of known genes, 65% were novel. From the 15% of known genes, 10% were structural, 86% were genes involved in replication, and 4% were putative auxiliary metabolic genes (AMGs). We categorized the AMGs into two distinct groups: host survival and C cycling. Host survival genes included oxidative stress, sporulation, heat shock, whereas C cycling genes included central C metabolism and CAZy enzymes. Almost 30% of the vOTUs were linked to the 326 MAGs across four phyla (Acidobacteria, Actinobacteria, Proteobacteria, and Bacteroidetes).

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 Early Career Research Program Award Number SCW1478 to J. Pett-Ridge at Lawrence Livermore National Laboratory. Work was performed at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231 and at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Using Quantitative Stable Isotope Probing to Link Precipitation Regimes of Mediterranean-Grassland Ecosystems to Soil Microbial Ecophysiology

Megan Foley*[1] (mmf289@nau.edu), Alex Greenlon[2], Dinesh Adhikari[3], Karis McFarlane[3], **Steve Blazewicz[3]**, **Mary Firestone[2]**, **Jill Banfield[2]**, **Bruce Hungate[1]**, **Jennifer Pett-Ridge[3]**

[1]Center for Ecosystem Science and Society, Northern Arizona University; [2]Department of Environmental Science, Policy, and Management, University of California, Berkeley; [3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA

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.*

Given the primacy of microorganisms in the formation, stabilization, and breakdown of soil organic matter, identifying how soil moisture regimes shape microbial ecophysiology may be insightful for understanding the link between microbial communities and soil C in a changing climate system. We characterized soils from three Mediterranean grasslands along a precipitation gradient and used quantitative stable isotope probing (qSIP) to assess the active microbial communities at each site. We used qSIP with iTag sequencing to compare patterns of bacterial growth across sites, and we have also developed an approach to combine qSIP with genome-resolved metagenomes to explore variation in the expression of microbial traits that may be shaped by soil moisture and relevant for understanding the fate of soil C.

We characterized soils from our three primary SFA sites during the 2018 wet season, when water is least limiting to growth. Triplicate soil cores were collected from Hopland Research and Extension Center, Angelo Coast Range Reserve, and Sedgwick Reserve. To establish soil chemistry and minerology, we used solid state ^{13}C NMR and quantitative XRD on bulk soils, and performed ^{14}C analyses on bulk soil and respired soil C. Based on solid-state ^{13}C NMR there, was slight variation in bulk soil C composition. Angelo, the site with the highest precipitation (2160 mm yr^{-1}) has the highest aromatic content, while Sedgwick (383 mm yr^{-1}) has the highest carbohydrate content; Hopland (956 mm yr^{-1}) has the highest lipid content. Lipid was the most well represented fraction of bulk soil C for all sites, followed by carbohydrate and protein. Based on quantitative XRD analysis, Angelo has the highest fraction of clay minerals while Hopland has the lowest. In addition to muscovite and chlorite that were present in all soils, Sedgwick soil contained about 15% halloysite and Hopland soil contained about 10% kaolinite. Samples were also analyzed for their ^{14}C content using LLNL's accelerator mass spectrometer. The average age of soil C increased with mean annual precipitation ($p<0.01$) and total C content,

suggesting that higher precipitation supports accumulation of SOM across the three sites. The ^{14}C values of CO_2 produced during six-day lab incubations were similar across sites and younger than bulk SOM—reflecting a microbial preference for recently fixed C.

To assess the active microbial communities at our sites, soils were incubated for 8 days with either natural abundance ^{16}O-H_2O or 98 atom % ^{18}O-H_2O. DNA was then extracted, separated by ultracentrifugation, and density-gradient fractionated. Total community DNA from 9 fractions, as well as unfractionated DNA from each incubation, was used for sequencing of the 16S rRNA marker gene, and for metagenomic library preparation and sequenced on the Illumina NovaSeq platform to an average depth of 8 Gbp per fraction for each sample and treatment (180 metagenomes).

Using ^{18}O enrichment of bacterial DNA as a proxy for growth, our amplicon-based qSIP analysis highlights distinct patterns in growth between the three sites. The intermediate site, Hopland, has the highest proportion of actively growing to total taxa, as well as the highest average enrichment of the significantly enriched taxa. Growing communities are not only distinct between the three sites, but also diverge from the total communities at a given site identified by sequencing alone. The growing communities at each site included many taxa within the phyla Acidobacteria, Actinobacteria, and Proteobacteria, and the drier two sites additionally had growing taxa within the phyla Armatimonadetes and Bacteriodetes. Relatively few bacterial taxa show growth at both Sedgwick and Angelo, the driest and the wettest sites. Only 8 taxa, all of which belong to the phylum Proteobacteria or Actinobacteria, showed growth at these two sites, compared to 48 and 20 shared growing taxa between Angelo and Hopland, and Sedgwick and Hopland, respectively.

Genome-resolved metagenomic analyses of the ^{18}O-H_2O SIP experiment enable quantitative assessment of metabolic pathways and microbial functional traits corresponding to growth and stasis in our study soils. Genomes with pathways for metabolizing 1-carbon molecules (in particular methanol and carbon monoxide) are overrepresented in ^{18}O-enriched genomes, corroborating previous observations that C1 metabolism plays an important role in soil microbial community functioning. We have not observed variation in isotopic enrichment of genomes with the capacity for compatible-solute synthesis by site, likely reflecting high moisture levels during the incubation. Isotopic enrichment values calculated on metagenome-assembled genome bins average 0.09, 0.15, and 0.06 atom fraction excess for Angelo, Sedgwick, and Hopland respectively, with upper ranges of 0.29, 0.36, and 0.21. The most commonly enriched taxa at each site include Acidobacteria, Actinobacteria, and Proteobacteria, as was observed in our amplicon sequences.

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 subcontracts to the University of California, Berkeley and Northern Arizona University.

Viral Diversity and Potential Carbon Cycling Impacts Across a Soil Climate Gradient

Christine L. Sun[1*] (sun.2508@osu.edu), Lindsey Solden[1], Alexa Nicolas[2], Ahmed Zayed[1], Alex Greenlon[2], Erin Nuccio[3], Olivier Zablocki[1], Jillian F. Banfield[2], Mary Firestone[2], Steven Blazewicz[3], **Jennifer Pett-Ridge[3]**, **Matthew B. Sullivan[1]**

[1]The Ohio State University, OH, USA; [2]University of California, Berkeley, CA, USA; [3]Lawrence Livermore National Laboratory, CA, USA

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.*

Microorganisms play many important roles in carbon (C) cycling in soils, and though viruses are known to modulate these roles in other systems (via lysis, gene flow and metabolic reprogramming) little is known about the impact of soil viruses due to technical challenges. Here, we examined viruses in well-characterized soil environments using data from purified viral particles, sequencing, and viral ecogenomic analyses. The soils were sampled from four sites, UC Hopland Research and Extension Center, Angelo Coast Range Reserve, Sedgwick Natural Reserve, and Luquillo Experimental Forest in Puerto Rico, which represent a gradient of soil moisture regimes.

This revealed 28,608 unique viral populations (≥ 10kb contigs dereplicated at 95% average nucleotide identity and 80% coverage) that sorted into nearly three thousand viral genera via gene-sharing networks. This dataset increases known soil virus populations by ~15-fold and soil viral genera 9-fold. Viral communities were strongly and statistically separated based on location, with few (0% - 8%) viral populations shared between sampling sites, suggesting that the heterogenous environmental conditions in soil give rise to very distinct communities. To assess potential viral roles in carbon cycling, we *in silico* predicted hosts, which, buoyed by metagenome-assembled-genomes (MAGs) from co-sampled bulk soil metagenomes, revealed putative hosts for 5% of the ~28K viruses, including 47 MAGs that are key C cyclers, including multiple *Actinobacteria* and *Gammaproteobacteria*. Further, we explored whether these soil viruses directly encoded key C cycling enzymes by identifying carbohydrate active enzymes (CAZymes), which catalyze the biosynthesis or breakdown of complex carbohydrates. This revealed 273 viral populations containing plant polysaccharide degrading CAZymes: 49% (GH28, PL1, PL22) targeted pectin, whereas 27% targeted beta-mannin (GH130 and GH113), 15% many polysaccharide bonds (GH140), and 4% beta xylan (GH19 and GH120), with the remaining CAZymes detected only once or a few times.

Together these results add to our baseline understanding of soil viral ecology and implicate viruses in C cycling via infection and directly encoded auxiliary metabolic genes. This knowledge provides a path for soil viruses be incorporated terrestrial ecosystem models to better understand terrestrial C cycling.

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 subcontracts to the Ohio State University and the University of California, Berkeley. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

High-Throughput Determination of a Subcellular Metabolic Network Map of Plants

Kevin Radja[1], Suryatapa Ghosh Jha[1], Tallyta Silva[2], Charles Hawkins[1], Angela Xu[1], Bo Xue[1], Christine Aquino[1], Edward J. Wolfrum[3], Markita P. Landry[4], David W. Ehrhardt[1], Jenny C. Mortimer[2], **Seung Y. Rhee[1]*** (srhee@carnegiescience.edu)

[1]Carnegie Institution for Science, Stanford, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]National Renewable Energy Laboratory, Golden, CO; [4]University of California, Berkeley, CA

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 compartmentation of metabolic reactions through the locations of enzymes is critical to understanding, modeling, and engineering plant metabolism. Yet, the localization of the majority of the predicted enzymes are 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 new project, a trans-disciplinary team with expertise in plant cell biology, genomics, metabolic modeling, algorithm development, synthetic biology, geochemistry, nanotechnology, and analytical chemistry will develop 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. Using the pipeline, the team will create a high-quality subcellular map of small molecule metabolism as well as accurately compartmentalized metabolic network models in Sorghum and Brachypodium. The models will be experimentally validated by measuring a series of outputs in response to environmental challenges, and by knocking out gene expression in somatic tissue using CRISPR technology. Here, we will outline our approach to this project, as well as some initial progress on transient sorghum transformation and data analysis.

The pipeline will rely on the existing metabolic pathway database SorghumBicolorCyc. This database 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 database. The pipeline implements a novel network-based classifier to infer compartmentalization of the Sorghum metabolic pathway database. The network-based classifier utilizes an existing classifier which predicts protein subcellular localization based on sequence information as training on the partially compartmentalized network.

Based on the predictions, target metabolic genes will be chosen to validate the localization in cells. To achieve this, Gateway and In-Fusion molecular cloning technology will be used to fuse these candidates with known and effective fluorescent proteins, and subsequently expressed in the plants using high-throughput nanomaterial-based plant transformation techniques. Utilizing high resolution confocal live imaging (EMCCD spinning disk and Leica HyD point scanning), the locations of these enzymes will be validated *in planta.* The dataset collected from these validations will be used for developing the network maps and refine current models.

Overall, this project aims to holistically decipher the complexity of plant metabolic networks in order to engineer pathways to tackle the impending problems of changing climate, food security and 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

A Thousand Highly Non-Repetitive Promoters for Controlling Transcription Rates in Clostridia during Syngas Fermentation

Nick Fackler[2†], Daniel Cetnar[1†], Ayaan Hossain[1], Steven D. Brown[2], Michael Köpke[2], and **Howard M. Salis**[1*] (salis@psu.edu)

[1]The Pennsylvania State University, University Park, PA; [2]LanzaTech Inc, Skokie, IL.

[†] These authors contributed equally to this work

https://salislab.net/; http://www.lanzatech.com

Project Goals:

New genetic tools are needed to engineer metabolic pathways and networks with many genes (enzymes, transporters, regulators), particularly in non-model organisms growing in conditions with industrial relevance. As a key challenge, it remains difficult to stably express many genes because designers are forced to re-use similar genetic parts, thereby introducing repetitive DNA sequences that trigger homologous recombination and genetic instability. To overcome this challenge, we applied our new algorithm, the Non-Repetitive Parts Calculator, to design over 30000 highly non-repetitive promoter sequences to control transcription rates in *Clostridia*. We experimentally characterized these non-repetitive promoters in *Clostridium autoethanogenum* during gas fermentation using a low-cost syngas feedstock, obtaining over 1000 transcription rate measurements simultaneously. To do this, we combined barcoded oligopool synthesis, library-based cloning into integrative vectors, high-throughput transformations, state-of-the-art gas fermentation facilities, and next-generation sequencing (DNA-Seq, RNA-Seq). With this toolbox, over 1000 genes can be simultaneously expressed in *Clostridia* with tunable transcriptional control across a 1,000,000-fold range, all without introducing more than a 15 bp repeat sequence. This non-repetitive toolbox of promoters enables a breadth of metabolic engineering applications in an important industrial organism.

References

Ayaan Hossain, Eriberto Lopez, Sean M. Halper, Daniel P. Cetnar, Alexander C. Reis, Devin Strickland, Eric Klavins, and Howard M. Salis. "Automated Design of Thousands of Highly Non-Repetitive Genetic Parts for Engineering Evolutionary Robust Genetic Systems", Nature Biotechnology, *in review*.

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.

Multiplex Genome Engineering for Bioproduction of 3-Hydroxypropionic Acid and 1,3-Propanediol from Waste Gases

Fungmin (Eric) Liew (eric.liew@lanzatech.com)[1*], Nick Fackler[1], James Daniell[1], Steven D. Brown[1], Michael Köpke[1], and **Howard M. Salis**[2]

[1]LanzaTech Inc, Skokie, IL; [2]The Pennsylvania State University, University Park, PA.

http://www.lanzatech.com; https://salislab.net/

Project Goals: Gas fermentation is a commercially scalable platform for the sustainable biomanufacturing of valuable chemicals from abundant, low cost C1 feedstocks. We have engineered improved *Clostridia* strains that produce 3-hydroxypropionic acid (3-HP) from biomass syngas or industrial waste gas, characterized during continuous culture with online monitoring. 3-HP is an ideal bio-renewable precursor to acrylates and polymers (acrylonitrile, acrylamide, acrylic acid and acrylate esters) with a global market estimated as 3.63 million tons per year. To do this, we carried out genome-scale modeling and genome engineering to evaluate and introduce 3-HP biosynthesis pathways into an industrial *Clostridium* strain. We also developed new genetic tools that are capable of controlling the expression of many enzymes simultaneously, without introducing repetitive DNA. By combining these genetic tools with system-wide modeling, we are systematically redirecting metabolic flux towards 3-HP biosynthesis and eliminating byproduct formation with overall improved titers, yields, and productivities.

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 (a greenhouse gas) and carbon monoxide (a harmful pollutant that will be oxidized to CO_2 when released in the atmosphere). LanzaTech is a world-wide 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.

In this project, we are engineering industrial *C. autoethanogenum* strains to manufacture 3-hydroxypropionic acid (3-HP) with commercially relevant metrics (volume, titer, yield, productivity) from a syngas feedstock. As a first step, we applied a customized genome-scale metabolic model to evaluate the yield and thermodynamic feasibility of 25 different 3-HP biosynthesis pathways. *In silico* optimization revealed that high 3-HP yields can be achieved, but

require a significant level of strain engineering: from 5 to 30 enzyme expression levels need to be modified to yield significant improvements.

This challenge motivated the development of new genetic tools that can tunably control many enzyme expression levels in *C. autoethanogenum* to avoid slow iterative cycles and growth defects or genetic instabilities observed for knock-out of specific target genes. We developed the first dynamic CRISPRi knock-down system for *C. autoethanogenum* and are currently scaling up the number of enzymes that can be simultaneously targeting by leveraging Extra Long sgRNA Arrays (ELSAs), which are capable of expressing up to 20 CRISPR sgRNAs within a compact, non-repetitive DNA cassette (Reis et al., 2019).

By introducing the malonyl-CoA reductase from *Chloroflexus aurantiacus* we have demonstrated *de novo* biosynthesis of 3-HP in *C. autoethanogenum*. Surprisingly, we also observed considerable production of 1,3-propanediol (1,3-PDO); by itself, 1,3-PDO is an important chemical with market size of $490 million in 2019. We confirmed that the conversion of 3-HP to 1,3-PDO in *C. autoethanogenum* occurs via aldehyde::ferredoxin oxidoreductase (AOR) enzymes (Liew et al., 2017). Knocking-out or knocking-down AOR expression eliminates 1,3-PDO production and increases 3-HP production. The ultimate goal of this project is to introduce model-designed ELSAs into the *C. autoethanogenum* genome, guided by genome-scale modeling as well as techno-economic analysis, to deliver strains that convert syngas into 3-HP with improved titers, yields, and productivities.

References

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/https://doi.org/10.1016/j.ymben.2017.01.007

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

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.

White-Rot Fungi Utilize Lignin-Derived Compounds as a Carbon Source

Carlos del Cerro Sánchez[1], Erika Erickson[1], Kelsey J. Ramirez[1], Tao Dong[1], Jan-Fang Cheng[2], Miranda Harmon-Smith[2], Robert Evans[2], Allison R. Wong[3], Elizabeth K. Eder[3], Hugh D. Mitchell[3], Samuel O. Purvine[3], Meagan C. Burnet[3], Rosalie K. Chu[3], Lye Meng Markillie[3], David W. Hoyt[3], and **Davinia Salvachúa**[1*] (davinia.salvachua@nrel.gov)

[1] National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO.

[2] Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA.

[3] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA.

Project Goals: This study aims to investigate the hypothesis that white-rot fungi can simultaneously depolymerize lignin extracellularly and catabolize depolymerization products intracellularly as carbon and energy sources. Understanding this potential biological activity and identifying the most promising fungal strains for lignin turnover and catabolism will justify future investment in genetic tool development to enable metabolic engineering in white-rot fungi for lignin bioconversion to bioproducts.

Lignin is currently an undervalued aromatic polymer in lignocellulosic biorefineries, mainly due to its heterogeneity and recalcitrance, but it is a key substrate to enable a sustainable plant-based bioeconomy. White-rot fungi (WRF) are recognized as the most efficient lignin-degrading organisms in Nature and are thus potential biocatalysts for lignin bioconversion. However, while lignin *depolymerization* by WRF has been studied for decades and is well accepted, the ability of WRF to *catabolize* lignin remains virtually unknown.

To ascertain if WRF simultaneously depolymerize lignin and utilize lignin-derived compounds, we are employing in *silico*, in *vivo,* and in *vitro* approaches based on hypothesis-driven, systems-biology studies. Recent ^{13}C-labeling analyses with two WRF that exhibit different lignocellulose degradation patterns have revealed that these organisms can indeed funnel carbon from lignin degradation products to central metabolism. Considering the lack of information on aromatic catabolic pathways in WRF, we have initiated pathway and enzyme discovery via in *silico* analyses. Specifically, we are conducting homology searches in two selected fungal genomes using as *template* enzymes already characterized in bacteria and yeast. Based on these results, we have hypothesized a pathway that WRF most likely use for the conversion of 4-hydroxybenzoic acid, an abundant aromatic compound found ester-linked to lignin in poplar. Multi-omic analyses (transcriptomics, proteomics, and metabolomics) based on differential substrate have also facilitated enzyme down-selection and confirmed metabolites from the hypothesized pathway. Lastly, we have initiated enzyme validation via *in vitro* studies to verify the function of the down-selected enzymes (e.g. decarboxylases, hydroxylases, and dioxygenases). Overall, the knowledge gained through this work will serve as a foundation to employ WRF in lignin bioconversion to bioproducts and provide deeper insight into global carbon cycling.

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. 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.

Approaches to the Development of Sustainable Energy Sorghum Biofuel Feedstocks in Drought Prone and Low Nitrogen Environments

Daniel P. Schachtman[1] (Daniel.Schachtman@unl.edu), Stephanie Futrell[1], Jiating Li[1], Amy Sheflin[2], Jessica E. Prenni[2], Yeyin Shi[1], Cody Creech[1], Asaph B. Cousins[3], Ellen L. Marsh[1], Emily Goren[4], Peng Liu[4], Stephen Kresovich[5]

[1]University of Nebraska, Lincoln, NE; [2]Colorado State University, Fort Collins, CO; [3]Washington State University, Pullman, Washington; [4]Iowa State University, Ames, Iowa; [5]Clemson University, SC

Website for project: https://sorghumsysbio.org/

Overall project goals:

- Establish a foundational, systems-level understanding of plant, microbial, and environmental interactions that will lead to translational strategies to enhance growth and sustainability of sorghum through improved genetic and microbial adaptations to water and nutrient limited environments.

Specific objectives addressed by this poster:

- Phenotypic characterizations of a diverse panel of sorghum genotypes across multiple years to define the most productive lines under drought and low nitrogen conditions.
- Associate genotypic and environmental effects with improved sorghum performance using robust statistical approaches.

The overall project involves both plant genetics and studies of the soil microbial communities associated with sorghum. In this poster we highlight the progress made towards identifying stress tolerance energy sorghum germplasm using:

- classical methods of measuring biomass
- more advanced approaches using:
    - metabolomic data to predict biomass
    - hyperspectral radiometry to measure nitrogen, chlorophyll and specific leaf area in the field on large populations
    - UAV to measure spectral qualities and temperature of the crop canopy
- chemical analyses such as wax composition and leaf sucrose content.

All of these traits were chosen because of their potentially importance in abiotic stress tolerance in sorghum. Results will be presented on how we are making these measurements, the prediction method and outcomes based on metabolite data, some preliminary wax composition results and canopy temperature data. Next steps being pursued to understand the genetic mechanisms of the variation in these observed traits will also be presented. Ultimately our goal is to understand the mechanisms that are important for abiotic stress tolerance and the underlying genetic factors that will contribute to enhancing the biomass production of energy sorghum for the biofuel feedstocks.

This project is funded by the DOE BER Sustainable Bioenergy Research Program, Award DE-SC0014395, and was also supported by DOE JGI Community Science Program

Data access, mining and visualization. Tools to accommodate an interdisciplinary project.

Philip Ozersky[1*] (pozersky@danforthcenter.org), Jeff Berry[1], Scott Lee[1,] Todd Mockler[1], and **Daniel Schachtman**[2]

[1]Danforth Plant Science Center, St. Louis, MO; [2]University of Nebraska, Lincoln;

Website for the project: www.sorghumsysbio.org

Overall Project Goals:

- Establish a foundational, systems-level understanding of plant, microbial, and environmental interactions that will lead to translational strategies to enhance growth and sustainability of sorghum through improved genetic and microbial adaptations to water and nutrient-limited environments.

Project Objectives:

- Conduct deep census surveys of root microbiomes concurrent with phenotypic characterizations of a diverse panel of sorghum genotypes across multiple years to define the microbes associated with the most productive lines under drought and low nitrogen conditions.
- Associate systems-level genotypic, microbial, and environmental effects with improved sorghum performance using robust statistical approaches.
- Develop culture collections of sorghum root/leaf associated microbes that recapitulate root-enriched sequences defined in the census.
- Perform controlled environment experiments for in-depth characterization and hypothesis testing of $G_{sorghum}$ x $G_{microbe}$ x E interactions.
- Validate physiological mechanisms, map genetic loci for stress tolerance, and determine the persistence of optimal microbial strains under greenhouse and field conditions.

Our project generates a wide variety of data from field crop measurements to indoor phenotyping systems and analytics of *Sorghum bicolor* samplings such as metabolites and stable isotopes. Additionally, sequences were produced for transcriptomic and metagenomic data. These datasets are disseminated through our project website and interface allowing all members of the project access to foster coordinated efforts and data sharing. Integration of tools like JBrowse and project produced Shiny apps allow project members to visualize, mine and parse project data. We are also leveraging pipelines to utilize *Sorghum bicolor* pan-genome data to help identify the most productive lines under drought and low nitrogen conditions. The poster will display our project website, database structure, interface utility, tools, and pipeline.

This project is funded by the DOE BER Sustainable Bioenergy Research Program, Award DE-SC0014395, and was also supported by DOE JGI Community Science Program.

Title: Leaf Carbon Isotope Composition in Diverse Sorghum Lines

Asaph B. Cousins[1]* (acousins@wsu.edu), Balasaheb V. Sonawane[1], Max Braud[2], Indrajit Kumar[2], Rajiv Parvathaneni[2], Andrea Eveland[2], Kira Veley[2], Jeffrey Berry[2], Rebecca Bart[2], Varsha Pathare[1], Stephanie Futrell[4], Ellen Marsh[4], Yen Ning Chai[4], Peng Liu[3], Cody Creech[4], Ismail Dweikat[4], Stephen Kresovich[5], Allyn Pella[4], Ellen Marsh[4], Peng Wang[4], **Daniel P. Schachtman**[4]

[1]Washington State University; [2]Donald Danforth Plant Science Center; [3]Iowa State University; [4]University of Nebraska-Lincoln; [5]Clemson University

http://sorghumsysbio.org

Project Goals:
The *overall project goal* is to establish a foundational, systems-level understanding of plant, microbial, and environmental interactions that will lead to strategies for enhancing growth and sustainability of sorghum through genetic and microbial adaptations to water and nitrogen limited environments.

The ***specific objective*** of the research presented here is to screen diverse panels of sorghum genotypes to determine the genetic factors influence leaf carbon isotope composition ($\delta^{13}C_{leaf}$). Towards achieving our project goals, we have conducted several phenotyper, greenhouse and field experiments on various populations of sorghum genotypes. Data will be presented on greenhouse experiments used to determine the relationship of whole plant water use efficiency (WUE_{plant}) with leaf intrinsic transpiration efficiency (TE_i) and $\delta^{13}C_{leaf}$ in a select number of sorghum lines. As predicted from theoretical models of C_4 photosynthesis we have demonstrated that $\delta^{13}C_{leaf}$ is related to TE_i, when efficiency of CO_2 concentrating mechanism (leakiness) remains constant. Accordingly, $\delta^{13}C_{leaf}$ has been proposed as a high-throughput phenotyping tool for TE_i in C_4 plants. We have scaled this research up to test the variation in $\delta^{13}C_{leaf}$ in 30 diverse line of sorghum grown under controlled environment growth conditions in the Bellweather Phenotyping System at the Danforth Plant Science Center and in the field near Scottsbluff, Nebraska. We have followed up on these initial experiments with screens of $\delta^{13}C_{leaf}$ across the sorghum biomass association panel (BAP) grown under both field and controlled environment growth conditions. Given the large genetic and phenotypic diversity within the BAP it is likely that variation in $\delta^{13}C$ exists in this population that can be associated with genetic loci.

Future directions
Leaf and whole plant traits will be assessed across sorghum genotypes to help identify and select for genomic traits and potentially elite lines for enhanced water use efficiency.

Funding statement
This project is funded by the DOE BER Sustainable Bioenergy Research Program, Award DE-SC0014395, and was also supported by DOE JGI Community Science Program.

Sorghum root microbiome dynamics under nutrient-limited and drought conditions

Susannah G. Tringe (sgtringe@lbl.gov) [1,2*], Dawn Chiniquy[2], Kyle Hartman[1], Jinglie Zhou[1], and **Daniel Schachtman**[3]

[1]DOE Joint Genome Institute, Berkeley, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]University of Nebraska, Lincoln, NE

http://www.sorghumsysbio.org

Project Goals:

We aim to establish a foundational, systems-level understanding of plant, microbial, and environmental interactions that will lead to translational strategies to enhance growth and sustainability of sorghum through improved genetic and microbial adaptations to water and nutrient limited environments. In working towards this goal, we are conducting deep census surveys of root microbiomes concurrent with phenotypic characterizations of a diverse panel of sorghum genotypes across multiple years to define the microbes associated with the most productive lines under drought and low nitrogen conditions.

Abstract:

Sorghum bicolor is a genetically diverse crop cultivated for a variety of agronomic uses, including grain, sugar, and energy production. However, cultivation of energy sorghum for biofuel production will require the use of marginal lands with potentially low nutrient availability and/or periods of water stress. All plants growing in soil harbor diverse communities of microbes that inhabit the areas in, on, and around their roots. Selected members of these microbial communities can provide benefits to their plant hosts, including direct growth promotion and conferring tolerances to abiotic and biotic stress. To examine the effects of nutrient and water stress on soil and root microbial communities and explore a possible microbial solution to increase the nutrient use efficiency and resilience to water stress in sorghum, we are utilizing 16S rRNA sequencing to survey the bacterial communities in replicate soil, rhizosphere, and root samples collected from ~30 different sorghum genotypes grown under different nitrogen (N; high/low) and water (watered/drought) treatments over multiple growing seasons at two sites in Nebraska.

In a small-scale pilot experiment in 2015, we collected ~200 soil, rhizosphere, and root samples from 10 different sorghum genotypes grown under high or low N conditions. We observed that early rhizosphere samples exhibit a significant reduction in overall diversity attributable to a dramatic increase in the bacterial genus *Pseudomonas* which occurred independent of host genotype in both high and low N fields, and was not observed in the surrounding soil or associated root endosphere samples. Nearly all the *Pseudomonas* reads in this dataset were assigned to a single OTU at 97% identity; however, ASV-level resolution demonstrated that this OTU comprises a large number of distinct *Pseudomonas* lineages. Furthermore, single-molecule long read

sequencing enabled high-resolution taxonomic profiling of the *Pseudomonas* lineages within the dataset. In additional field experiments in 2016 and 2017, we sampled three selected genotypes at four time points throughout the growing season to monitor changes in the bacterial communities over time and again observed decreases in Shannon diversity in rhizosphere samples early in the growth season concomitant with a marked increase in the relative abundance of *Pseudomonas*. Our ongoing work focuses on characterizing the genomic and metabolic features of these Pseudomonas populations and integrating our findings with the metabolomic and phenotypic data generated with project collaborators.

This project was funded by the DOE BER Sustainable Bioenergy Research Program, Award DESC0014395, and was also supported by DOE JGI Community Science Program. 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.

Gene Regulatory Networks Enabling Fungi to Selectively Extract Sugars from Lignocellulose

Claire E. Anderson[1], **Jonathan S. Schilling[1]** (schillin@umn.edu), Jesus Castano[1*], 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:

Certain filamentous fungi are uniquely able to deconstruct lignocellulose, and their poorly understood mechanisms have potential biofuels applications. A key hindrance to harnessing these fungal mechanisms has been their spatial complexity. Our past work has shown that differentiated networks of hyphae that penetrate wood are not metabolically uniform, with critical reactions occurring near the hyphal front. Standard omics analyses of these fungi from artificial media or from colonized wood ground en masse fail to distinguish expression of key gene products occurring in localized regions along growing hyphae.

Our focus for this research is specifically on brown rot fungi, a more recently evolved decay fungal group (relative to white rot) that circumvents the lignin barrier to extract sugars from lignocellulose. The genetic basis for how this capacity evolved away from white rot multiple times remains unknown, despite the modern options to align the compare brown rot and white rot fungal genomes. Our new collaboration aims to focus omics techniques to map and integrate expression over networks of wood-degrading fungal hyphae *in planta*. Wood-degrading fungal genomes are an emerging resource, particularly brown rot functional types. We recently optimized a thin-section wood set-up that can finely resolve reaction zones along an advancing mycelium. Within these zones, we can employ deep omics approaches without the typical noise of whole-sample homogenization. By co-localizing gene expression, secretions, and wood modifications, we can prioritize the genes most useful for application, as well as understanding the underlying regulation of a globally important decomposition mechanism – brown rot.

Our goal is to discover which genes are differentially up-regulated across the mycelia of wood-degrading fungi *in planta*, particularly at the leading edge of wood decomposition where reactive oxygen species (ROS) are deployed. To do this, we need to compare global expression profiles among mycelial regions. To map this wood-fungal interaction, we must match gene expression patterns with the extracellular secretome and with physiochemical wood modifications. Given this potential for substrate-fungus feedback, we are cross-checking genes using separate clade representatives for brown rot fungi alongside their white rot ancestors, harnessing the JGI MycoCosm portal and several key resources and expertise at JGI and the Pacific Northwest National Laboratory.

References:

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.).

A Systems Approach to Enhancing Seedling Establishment for Increased Yields in the Oilseed Crop, *Camelina sativa*

Hesham Abdullah[1], Bibin Paulose[1], Rajneesh Singhal[1], and **Danny J. Schnell[1*]** (schnelld@msu.edu)

[1]Department of Plant Biology, Michigan State University, East Lansing, MI, USA

https://groecamelina.natsci.msu.edu

Project Goals:

Our research plan aims to establish the non-food oilseed crop plant, *Camelina sativa,* as a commercially viable, dedicated biofuel and bioproduct feedstock. We are focused on improving seed and oil yields by employing an integrated genetic and metabolic systems approach to increase the rates of photosynthetic CO_2 capture and conversion to triacylglycerols (TAGs). Camelina has the advantages of low agronomic inputs and natural resistance to biotic and abiotic stresses relative to other oilseed crops, and Camelina oil-based blends have been tested and approved as liquid transportation fuels. Camelina also benefits as a synthetic biology platform from a fully sequenced genome, well-established molecular genetic tools, and numerous resources available from its close relative, *Arabidopsis thaliana*. Despite these advantages, the major limitation in widespread adoption of Camelina as an industrial oilseed crop is its modest oil yield. Our project will address yield directly by employing a tissue-specific and whole-plant systems approach to identify the major regulatory mechanisms that limit 1) **carbon fixation in photosynthetically active source tissues (leaves)**, 2) **the transport of fixed carbon from source to sink tissues (seeds)**, and 3) **the allocation of fixed carbon to TAG production**. The limiting factors identified in our analyses will be individually validated and combined using multi-gene stacking and genome editing technologies to engineer Camelina.

Abstract:

Varying environmental conditions, including changes in temperature, light, water and nutrient availability, result in energy and metabolic imbalances that significantly impact crop productivity and agricultural production. Metabolic interactions between chloroplasts and mitochondria, including photorespiratory metabolism, are known to be essential in mitigating these imbalances, and strategies to enhance the plasticity of these networks have the potential to enhance crop resilience to environmental stress. We previously described improved carbon assimilation, increased water and nitrogen use efficiency, and higher seed yields in Camelina by expressing LIP36, a mitochondrial carrier that promotes oxaloacetate/malate exchange. Metabolic systems analysis demonstrated that LIP36 transport activity (1) enhances the non-cyclic phase of the mitochondrial tricarboxylic acid (TCA) pathway in photosynthetic tissues, and (2) promotes redox exchange between the mitochondria and cytosol. In doing so, LIP36 contributes to cellular redox balance and optimal photorespiratory metabolism, while favoring anabolic over respiratory metabolism. This results in increased stress resilience and improved seed yields under limiting environmental conditions.

In the current study, we took advantage of LIP36 tissue-specific expression to examine the role of mitochondrial metabolic networks in optimizing respiratory, anabolic, and redox metabolism in early seedling development, root growth/architecture, and seed development. We will present combined metabolic, physiological and transcript profiling data that demonstrate a positive impact of LIP36 expression on early seedling establishment. This includes increased shoot growth and root elongation within the first 10 days following germination. Metabolic and transcript analysis demonstrate that the primary impact is on cellular redox balance. Our data suggest that the ability to increase the plasticity of redox exchange between the cytoplasm and mitochondria, may decrease ROS generation and alter the expression of important stress resilience pathways that increase robust seedling establishment. Enhancing seedling establishment is a key agronomic factor for improving the stand and the ultimate yields of crops grown on marginal lands. We anticipate that this trait could contribute to increased adoption of Camelina as an industrial oilseed crop.

This work is funded by grant #DE-SC0018269 from the Department of Energy BER program.

Advancing Field Pennycress as a New Oilseed Biofuels Feedstock that Does Not Require New Land Commitments

John Sedbrook[1]* (jcsedbr@ilstu.edu), Winthrop Phippen,[2] John Ralph,[3] and David Marks[4]
[1]Illinois State University, Normal, Il; [2]Western Illinois University, Macomb, Il; [3]University of Wisconsin-Madison, WI, [4]University of Minnesota, Minneapolis, MN
http://www.wiu.edu/pennycress/ http://cbs.umn.edu/marks-lab/home
https://about.illinoisstate.edu/jcsedbr/ http://www.iprefercap.org/

Project Goals: This project aims to genetically improve the agronomic traits of pennycress (*Thlaspi arvense* L.; Field Pennycress) for its use as an offseason oilseed-producing cash cover crop grown throughout the U.S. Midwest. We have identified and sequence-indexed an array of EMS-induced pennycress mutant lines exhibiting improved phenotypes for all agronomic traits we have assessed. We have also developed and demonstrated the utility of pennycress *Agrobacterium*-mediated plant transformation and CRISPR genome editing, generating pennycress lines with undetectable levels of erucic acid in seed oil, reduced glucosinolate content, reduced seed coat fiber, reduced pod shatter, and reduced seed dormancy. This project aims to identify agronomically-relevant alleles, validated through field testing, and to stack those alleles into top pennycress breeding lines to generate elite varieties having the following traits allowing for commercialization: 1) Harvestable seed yields of at least 1,500 lbs/acre; 2) Reduced sinigrin (glucosinolate) to below the regulatory limit; 3) Reduced seed coat fiber content to improve the seed meal nutritional value 4) Shortened time to maturity to expand the range where pennycress can be harvested in time to plant full-season soybeans.

Pennycress (*Thlaspi arvense*; Field pennycress) is an oilseed plant of the Brassicaceae family closely related to Arabidopsis, carinata, camelina, and rapeseed canola. Pennycress is native to Eurasia and naturalized to North America, growing widespread throughout temperate regions of the world. Domesticated pennycress could be grown as a winter annual oilseed-producing cash cover crop, for example, planted in the fall at or near the time of corn harvest and harvested in the spring in time to plant full-season soybeans throughout the 80 million-acre U.S. Midwest Corn Belt. Elite pennycress varieties could provide additional income to farmers and supply-chain businesses thereby strengthening rural communities. Pennycress could also provide ecosystem services as a cover crop, reducing soil erosion and nutrients runoff and providing habitat and pollinator support on otherwise barren farmland.

Field trials have demonstrated that pennycress can be seeded in upper Midwest cornfields in the late summer/early fall, at which time the plants form ground-hugging rosettes of leaves that overwinter with extreme cold tolerance. Pennycress plants are quick to flower in the spring, producing nectar-generating flowers for pollenating insects and oil- and protein-rich seeds that can be harvested without disrupting soybean planting or yields. As an energy crop adopted throughout the U.S. Midwest Corn Belt, domesticated pennycress could annually produce 3 billion gallons of liquid transportation fuels and 20 million tons of high-protein seed meal. However, to fully realize this potential, more-involved genetic improvements conferring abiotic and biotic stress resilience must be made.

A focus of this USDA NIFA-funded project, in collaboration with participants of the USDA AFRI-funded IPREFER project (Integrated Pennycress Research Enabling Farm and Energy Resilience) and the start-up company CoverCress, Inc., is to develop first-generation pennycress varieties harboring the core domestication traits necessary for commercialization. For this presentation, we will highlight progress we have made in identifying, characterizing, and

field-test-validating pennycress mutations conferring low seed-oil erucic acid content, reduced glucosinolate content, reduced seed coat fiber, reduced pod shatter, and earlier flowering time. Mutations in multiple genes have been generated and stacked in various combinations, providing alternative routes to producing commercially-viable elite pennycress varieties. Trait-improving alleles are being incorporated into our breeding programs and are being evaluated at multiple field sites throughout the Midwest in conjunction with the IPREFER project and CoverCress, Inc., with the aim to establish agronomic practices and supply chains necessary for Covercress™ (domesticated pennycress) commercialization within five years.

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award numbers 2014-67009-22305 and 2018-67009-27374.

Competition Between Methanotrophs for Copper

Christina S. Kang-Yun[1*] (csrkang@umich.edu), Philip Dershowitz[2], 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
Aerobic methanotrophs - microorganisms that oxidize methane by coupling it to dioxygen reduction - play a critical role in the biogeochemical cycling of carbon. More specifically, these intriguing microbes consume substantial amounts of methane generated via methanogenesis, and thus are important "filters" that control environmental emissions of methane. Expression and activity of alternative forms of methane monooxygenase (MMO, responsible for the conversion of methane to methanol), however, is controlled by copper, or the canonical "copper-switch"

Methanotrophs have been found to have multiple mechanisms of copper uptake. The first well-characterized copper binding compound or chalkophore – methanobactin – was found to be expressed by some methanotrophs of the *Methylocystaceae* family within the Alphaproteobacteria, e.g., *Methylosinus trichosporium* OB3b. Methanobactin (MB) is a modified polypeptide containing two heterocyclic rings with associated thioamide groups that collectively are responsible for copper binding with extremely high affinity. Not all methanotrophs, however, can produce MB. Rather, methanotrophs of the *Methylococcaceae* family of the Gammaproteobacteria rely on an outer membrane protein (MopE or CorA) for copper sequestration as well as some *Methylococcaceae* secreting a copper-binding compound akin to MB, but with much weaker affinity for copper.

Given the importance of copper in methanotrophy, this raises several intriguing questions. First, do methanotrophs that express MB have a competitive advantage for copper sequestration? Competition between methanotrophs for copper is likely, with such competition affecting overall methanotrophic community composition, and by extension methanotrophic activity. Second, given that MB is secreted into the environment and then taken up after binding copper, can copper-MB complexes be "stolen" by other microbes? Such phenomena would require non-MB expressing methanotrophs to have the uptake system identified for MB, i.e., the TonB-dependent transporter required for MB uptake (MbnT).

Herein we show that, similar to that found for siderophore theft between certain microbes, "cheating" exists between methanotrophs for MB. Specifically, *Methylomicrobium album* BG8 and *Methylocystis* sp. Rockwell, lacking genes for MB biosynthesis, carry genes for the TonB-dependent transporter required for MB uptake. Growth and gene expression studies show that these methanotrophs are not starved for copper in the presence of MB, but are in the presence of triethylenetetramine, a strong abiotic copper chelating agent. In addition, *Methylocystis* sp. Rockwell more readily uptakes a specific type of MB, indicating presence of a specific MbnT. Such data indicate that copper-MB complexes were taken up by *M. album* BG8 and *Methylocystis* sp. Rockwell. These results also indicate MB may be considered to be a "public good" and provides a new mechanism to manipulate microbial communities for control of greenhouse gas emissions *in situ*.

This work was supported by the Department of Energy, Award # DE-SC0020174.

Genetic tools to optimize lignocellulose conversion in anaerobic fungi and interrogate their genomes

Ethan Hillman[1,2*] (ehillman@purdue.edu), Casey Hooker[1,3*] (hookerc@purdue.edu), and **Kevin Solomon[1,2,3]**

[1]Department of Agricultural and Biological Engineering, Purdue University,West Lafayette, IN; [2]Purdue University Interdisciplinary Life Sciences (PULSe) Program, Purdue University, West Lafayette, IN; [3]Laboratory of Renewable Resources Engineering (LORRE), Purdue University, West Lafayette, IN

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 isolated from giraffes, wildebeest, and donkeys. Initial characterization of these species confirm that they form distinct species from 2 genera of anaerobic fungi that exhibit high enzymatic activity against a range of untreated lignocellulolytic substrates regardless of lignin composition (e.g. GM poplar lines, sorghum, alfalfa, corn stover). Anaerobic fungi extensively tailor the relative abundance of secreted CAZymes to adapt to differences in substrate composition and achieve consistently high-levels of synergistic activity. To better understand this response and identify tools for genetic engineering, in the first phase of the project we are sequencing the transcriptomes of these isolates across various substrates and their genomes in partnership with the DOE-JGI. Using these resources, we intend to identify sequences such as inducible promoters and environmental conditions that regulate them to develop novel tools for gene expression. Our ongoing work with DOE resources such as MycoCosm has already validated constitutive promoters that can express

distinct genes in specified cellular compartments and alter cellular phenotypes of anaerobic fungi for the first time.

As anaerobic fungal genomes have extremely low GC contents (~15%) that impede accurate genome assembly, in parallel we are leveraging Hi-C (chromosomal conformation capture) sequencing technologies to improve sequence assembly and determine genomic structure to better infer gene function. While efforts are currently underway, early results have already identified a handful of assembly artifacts that were erroneously generated by conventional assembly algorithms. Ongoing work aims to generate and annotate the most accurate and high-resolution genomes of anaerobic fungi to date and determine ploidy, enabling more sophisticated genetics studies.

Anaerobic fungi display many classical hallmarks of epigenetic regulation within their genomes. Their genomes contain dozens of annotated histone proteins, and histone- and DNA- modifying proteins that are regulated with growth substrate[1,3,4]. This regulation is also strongly correlated with CAZyme expression in a gene-specific manner[4]. Our preliminary studies with the histone deacetylase inhibitor suberoylanilide hydroxamic acid (SAHA) corroborate this as we can globally modulate H3K4 and H3K27 trimethylation levels with an effective doubling in xylanase activity. To date we have investigated the effect of a panel of epigenetic inhibitors on anaerobic fungi establishing for the first time that anaerobic fungal CAZymes are epigenetically regulated. We are beginning to identify the epigenetic marks that are affected by these inhibitors via Western Blots and are working in partnership with PNNL EMSL and the DOE-JGI to establish the effects of specific epigenetic marks at a given loci on CAZyme expression and activity.

In summary, the ongoing work creates new tools to manipulate anaerobic fungi facilitating new opportunities for study and engineering. These approaches will be used to generate a deeper systems-level understanding of anaerobic fungal physiology while establishing fundamental knowledge about epigenetic 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).

Beneficial Partners: Mycorrhizal Resource Exchange in Bioenergy Cropping Systems

Rachel Hestrin[1*] (hestrin1@llnl.gov), Megan Kan[1], Prasun Ray[2], **Rhona Stuart[1]**, **Christine V. Hawkes[3]**, **Kelly Craven[2]**, **Mary Firestone[4]**, **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]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC; [4]University of California, 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, plant resilience to stress, and carbon (C) allocation belowground. A better understanding of plant-mycorrhizal relationships can inform more sustainable and productive 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 otherwise unsuitable for agricultural production. We are investigating context-dependent resource exchange between *Panicum hallii*—a plant model species closely related to switchgrass—and two mycorrhizal fungi: the arbuscular mycorrhizal fungus (AMF) *Rhizophagus irregularis* and the ericoid mycorrhizal fungus *Serendipita bescii*. Both fungal species 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 fungal species 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' marginal soil harvested from a pasture in Oklahoma. We imposed water limitation in half of the microcosms in order to assess plant and mycorrhizal response to drought. 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 three timepoints over the course of a growing season. Coupled with stable isotope probing (SIP), this design allows us to track plant- and mycorrhizal-derived ^{13}C into other microbial taxa, soil C pools, and C fluxes (CO_2, volatiles, dissolved C). Since each microcosm included root and hyphal exclusion chambers, we expect to be able to examine key biogeochemical fluxes, nutrient exchange, multipartite interactions, microbial community assembly, metabolite production, and gene expression in spatially distinct ecological niches. Our current results indicate that plant performance was slightly enhanced in

the presence of mycorrhizal fungi under both drought and water-replete conditions. Isotopic analyses and NanoSIMS imaging suggest that a substantial quantity of the C that plants allocate to their mycorrhizal partners is further assimilated by other microorganisms. Although much of this C is retained below ground, some returns to the atmosphere as CO_2 and volatile organic compounds. These results shed light upon the complex and dynamic nature of plant-microbe-soil interactions and can be used to evaluate the potential role of mycorrhizal partnerships in 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.

Categorizing metabolic exchange and signaling reveal distinct mechanisms of mutualistic algal-bacterial interactions

Xavier Mayali* (mayali1@llnl.gov)[1], Jeff Kimbrel[1], Christina Ramon[1], Ty Samo[1], **Rhona Stuart[1], Peter Weber[1]**

[1]Physical and Life Sciences, 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.

To gain a mechanistic understanding of autotroph-heterotroph interactions in algae-dominated biofuel-producing systems, we are using the saltwater diatom *Phaeodactylum tricornutum* as a model. We have isolated 15 strains of heterotrophic bacteria from outdoor *P. tricornutum* raceway ponds which span the taxonomic diversity of microbial communities associated with saltwater microalgae, numerically dominated by Gammaproteobacteria, Alphaproteobacteria, and Bacteroidetes. We maintain these bacteria as stable co-cultures with *P. tricornutum* in laboratory batch mode without external inputs of organic material, using only light and inorganic nutrients to sustain growth. We are using microscopy, genomics and transcriptomics to investigate what functional genes are required, and in some cases, expressed for bacterial attachment and interaction with *P. tricornutum*. In addition, stable isotope probing combined with imaging via nano secondary ion mass spectrometry (NanoSIMS) enables us to quantify the cell-specific activities of both algae and bacteria when incubated together under different conditions.

Using ^{13}C and ^{15}N labeling of high molecular weight algal excreted organic matter (collected via solid-phase extraction) added back to algal-bacterial co-cultures, we have quantified bacterial incorporation and remineralization of organic C and N back to the algal cells. Surprisingly, we found that incorporation of algal organic matter is not universal among our isolates, suggesting the algal-associated bacteria can be divided into two categories. The first category comprises bacteria with strong metabolic interactions with their algal host, that incorporate algal-organic C and N and generally (but not always) remineralize N (and sometimes C). The second category

comprises bacteria with weak or no metabolic interactions with their algal host, that do not incorporate much algal-excreted organic matter. Some of these bacteria, in fact, are able to incorporate inorganic nitrogen and unknown sources of C. Particularly noteworthy is that bacterium *Marinobacter* strain 3-2, an algal mutualist, generally belongs to the second category, suggesting that mutualistic bacteria are not necessarily taxa with strong metabolic exchange with their algal host. Another bacterium, *Algoriphagus* strain ARW1R1, exhibits strong metabolic exchange with *P. tricornutum* but is only mutualistic under low light and high nutrient conditions, which is partially mediated by the metabolism of the small molecule nicotinamide (vitamin B3). From profiling these different types of bacteria and comparing with other metrics of influence on algal physiology, a framework has emerged for different grouping of bacterial associates of *P. tricornutum*, which has implications for our understanding of niche partitioning and diversity maintenance of algal-associated microbial communities.

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.

Characterizing algal metabolites and their role in biotic interactions

Vanessa Brisson[1] (brisson2@llnl.gov), Kristen Reese[1], Carolyn Fisher[2], Michael Thelen[1], Xavier Mayali[1], **Trent Northen**[3,4], **Todd Lane** (twlane@sandia.gov)[2], **Rhona Stuart**[1]

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Sandia National Laboratories, Livermore, CA; [3]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [4]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.

It has become increasingly clear that complex metabolic interactions between algae and their associated microbes influence algal physiology and growth. The suite of metabolites produced and exchanged between algae and bacteria are potentially important mediators of these interactions. Through a series of experiments, we set out to profile the intracellular, dissolved, and volatile metabolites produced by a diverse suite of microalgae.

To begin to characterize the diversity of algal metabolite secretion, we used an LCMS metabolomics analysis to compare the extracellular and intracellular metabolite profiles of phylogenetically diverse algal strains: the freshwater algae *Chlamydomonas reinhardtii*, the saltwater algae *Microchloropsis salina*, *Desmodesmus* sp. strain C406, which can grow in either fresh or saltwater, and the diatom *Phaeodactylum tricornutum*. The metabolite profiles of the intracellular and extracellular metabolites were quite distinct, suggesting that the detected exometabolites are unlikely to have originated from lysed cells. Exometabolite profiles also differed significantly between algae, with a few shared metabolites. Interestingly, when *Desmodesmus* was grown in freshwater versus saltwater, the metabolite profiles of both the intracellular and extracellular metabolites changed significantly. This suggests that along with a broad phylogenetic diversity, secreted as well as internal metabolites are modulated in response to external environmental cues and physiological state.

To follow up on the dynamics of exometabolite production, we focused on *P. tricornutum* throughout a growth cycle, and identified metabolites present at various stages of algal growth. The identified metabolites included phytohormones, several B-vitamins, and B-vitamin derivatives. There was a marked progression of metabolite composition and accumulation over time in both the intracellular and extracellular metabolite profiles. Within the extracellular metabolite pool, some metabolites were produced early in growth but did not continue to accumulate in later in growth, while others were at low or undetectable levels in early time points and only accumulated in late log and stationary phase. This suggests that *P. tricornutum* metabolite secretion is both dynamic and regulated to some degree by growth stage. We are currently investigating the potential of these extracellular metabolites from *P. tricornutum* to influence the growth of bacteria from our isolate library.

In parallel, in collaboration with Sandia National Laboratories we are identifying chemical and genetic signatures of microbial communities that are either indicative of pond health or the presence of algal grazers, parasites and pathogens. Towards this end, we have used both solid-phase microextraction (SPME) and thermal desorption (TD) coupled with gas chromatography-mass spectrometry (GC-MS) to survey the production of algal volatile organic compounds (AVOCs) by several strains of microalgae in the presence and absence of algal grazers and other deleterious species and stress conditions. The addition of the grazer, *Brachionus plicatilis,* to healthy cultures of *M. salina* led to emissions of the carotenoid degradation products, trans-β-ionone and β-cyclocitral which increased with rotifer consumption of algae. In addition to carotenoid breakdown products we have identified additional AVOCs that are indicative of grazer activity or cell disruption in algal mass cultures. Our results indicate that specific AVOCs may be early indicators of grazing or other algal disruption providing a useful tool to monitor algal biomass production and prevent pond crashes. Our results also shed light on the complex suite previously uncharacterized volatile organic compounds in produced as a result of trophic interactions in these diverse algal cultures.

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.

Examining the role of physical proximity and diffusion of metabolites in algal-bacterial interactions

Hyungseok Kim* (hskimm@mit.edu)[1], Jeffrey A. Kimbrel[2], Ty Samo[2], **Rhona Stuart[2],** Jessica Wollard[2], Dušan Veličković[3], Christopher Anderton[3]**, Cullen 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; [3]Environmental and Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA.

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.

Symbioses are governed by the exchange of material and information, encompassing a broad diversity of biomolecules with differing rates of diffusion. We are particularly interested in the role of physical association on algal-bacterial interactions, where exchange will be governed in part by diffusive processes. Although metabolic interactions between algae and bacteria are largely driven by diffusion of metabolites, there has been no systematic approach to control the diffusion process in a culture system. Here, we present research on developing approaches to examine the consequences of spatial proximity on these interactions and community assembly of algal-associated microbiomes. We have taken a dual approach to this problem: first in designing and testing a novel platform based on a biocompatible, nanoporous and structured hydrogel that allows metabolic communication between microbial species without any physical contact (Figure 1), and second by applying metabolite imaging to map diffusion of exuded and exchanged metabolites. Using the hydrogel platform, we have cultured *Phaeodactylum tricornutum,* a model and bioenergy-relevant diatom, in the center of a set of microwells, and examined bacterial community development as a function of distance from the algae. We identified specific bacterial taxa whose relative abundances were affected by proximity to the algal cells, indicating that diffusive processes can play a role in structuring algal-associated microbial communities. Using isolates from our *P. tricornutum*-associated isolate library, we further verify that these bacterial

taxa have their own growth characteristics under controlled diffusion of algal exudates. These results indicate that individual bacterial taxa within a community have unique strategies to co-exist with an algal host, allowing us to better understand these complex interactions and their role in structuring the entire community. In parallel, through a partnership with EMSL, we have begun to apply metabolite imaging to characterize metabolites exuded by *P. tricornutum* and a bacterium when grown in close spatial proximity. We have identified a broad range of algal and bacterial metabolites, including some putative plant hormones that may be involved in algal bacterial mutualism. In future research we plan to combine these two approaches to image metabolites in the hydrogel devices in order to identify and test key compounds that play a role in governing these interactions.

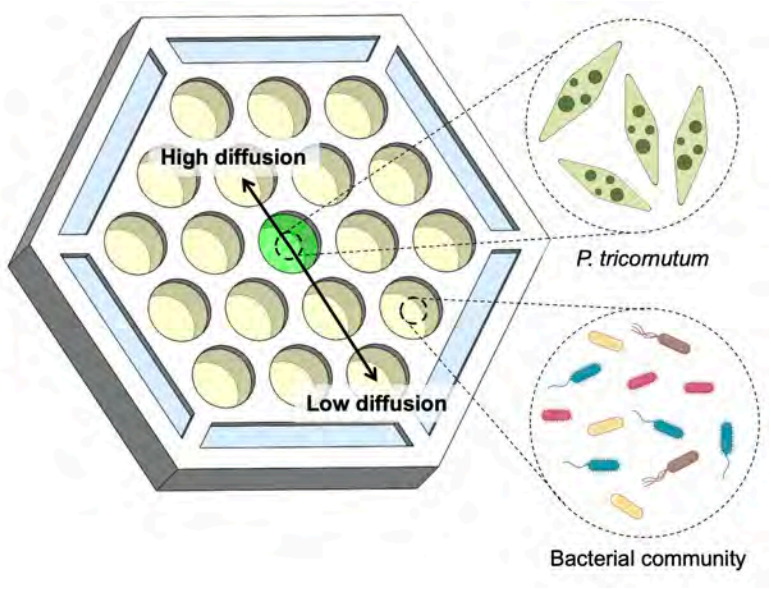


Figure 1. Schematic diagram of co-culture platform for algal and bacterial communities under controlled diffusion of metabolites.

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.

System-level analyses of beneficial interactions in an algal-bacterial co-culture

Ali Navid[1] (navid1@llnl.gov), Miriam Windler[2], Alfred Spormann[2], Patrik D'haeseleer[1], Mary Lipton[3], Sam Purvine[3], **Xavier Mayali**[1] and **Rhona Stuart**[1]

[1]Physical & Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Department of Civil and Environmental Engineering, Stanford University, Stanford, CA; [3]Environmental and Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

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.

As part of our SFA, we have generated global expression data for algal-bacterial co-cultures, for which we are interested in the regulation of their metabolic interactions. We apply metabolic modeling approaches to this proteogenomic data to both constrain our models, interpret our results, and inform follow-up experiments. One illustration of the utility of this approach comes from our investigation of the interaction between the model green algae *Chlamydomonas reinhardtii* and an Actinobacterium, *Arthrobacter sp.* strain P2b. Our experimental results demonstrated increased algal biomass in the presence of P2b cells as well as cell-free P2b spent medium. Moreover, the results showed that the active compounds in the spent medium were heat resistant and smaller than 3 kDa, suggesting that small molecules could be involved. We next examined changes in algal global protein expression in response to both P2b co-culturing and P2b spent medium, to try and determine which molecules may be responsible for the changes in algal physiology observed.

To gain a more systems-level understanding, we analyzed the proteomic expression data using a human-curated genome-scale computational model of metabolism in *C. reinhardtii*[1] with our in-house developed GX-FBA[2] method of constraining flux balance analysis models with transcriptomic and proteomic data. The GX-FBA analyses identified two processes that were consistently differentially regulated between axenic cultures of *C. reinhardtii* and those when it was either grown in a co-culture with P2b or in medium containing P2b spent medium. In all

cases we found that compared to the axenic culture 1) the pathways for production of sulfur containing amino acids (i.e. cysteine and methionine) and 2) the pathway for production of vitamin D3 were upregulated. We hypothesize that the first metabolic response may be due to cysteine being a precursor for production of glutathione, a potent antioxidant. Glutathione may be needed to offset the increased production of harmful reactive oxidative species (ROS) that are produced during photosynthesis under high algal biomass.

We followed up these results by directly testing the addition of vitamin D3 to *C. reinhardtii* and *Chlamydomonas*-P2b co-cultures, grown under two different media. Co-cultures grown in P49 media (with yeast extract and tryptone), which allows P2b growth, exhibited a 100% growth increase when 1 micromolar vitamin D3 was added. Co-cultures grown in Bold's medium, with no organic C and N sources, where P2b cannot grow, exhibited a more modest but still significant 38% increase in algal growth. Vitamin D3 additions to axenic *C. reinhardtii* did not exhibit growth increases, suggesting strain P2b needs to metabolize vitamin D3 into a different component in order to exhibit mutualistic effects. Together, these results demonstrate the utility of applying metabolic models to direct follow-up experiments, and indicate that vitamin D3 may play an important role in *C. reinhardtii* -bacterial interactions.

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.

1 Chang, R. L. *et al.* Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. *Molecular systems biology* **7** (2011).

2 Navid, A. & Almaas, E. Genome-level transcription data of Yersinia pestis analyzed with a New metabolic constraint-based approach. *BMC Syst Biol* **6**, 150, doi:10.1186/1752-0509-6-150 (2012).

Tools for Importing, Comparing and Merging Functional Annotations for Improved Metabolic Modeling in KBase

Patrik D'haeseleer[1] (dhaeseleer2@llnl.gov), Jeffrey A. Kimbrel[1], Sam Brinker[1], Janaka N. Edirisinghe[2], Felipe Liu[2], James Jeffryes[2], Ali Navid[1], Chris Henry[2], and Rhona K. Stuart[1]

[1]Lawrence Livermore National Laboratory, Livermore CA, USA
[2]Argonne National Laboratory, Lemont IL, USA

https://bio-sfa.llnl.gov/
https://narrative.kbase.us/#appcatalog/module//MergeMetabolicAnnotations

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy 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.

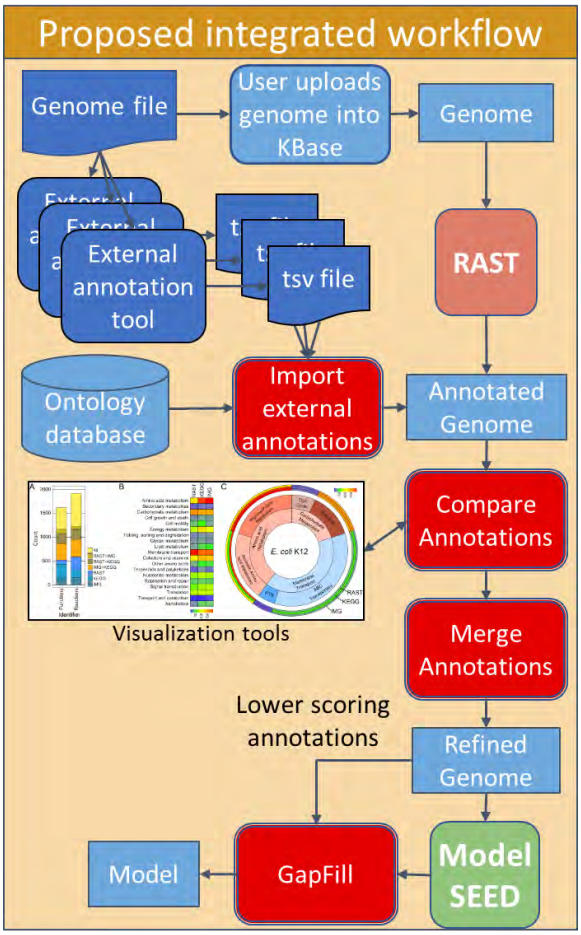
Metabolic pathway analysis, and especially metabolic modeling, is one of the cornerstones of modern Systems Biology, because it allows us to go straight from sequence data to gene functions to an understanding of how whole biological systems function. Accurate metabolic models require well-annotated genomes. Unfortunately, assigning functional annotations to genes is an imperfect science, and annotated genomes typically contain 30-50% of genes with little or no functional annotation, severely limiting our knowledge of the "parts lists" that the organisms have at their disposal.

We have shown that single metabolic annotation tools such as RAST or KEGG tend to be incomplete and inconsistent, and that merging annotation from multiple sources can drastically increase the number of genes and metabolic reactions included in metabolic models (1). Merging annotations added on average 40% more reactions, 3-8 times more substrate-specific transporters, and 37% more metabolic genes, compared to annotation using only a single tool. These results are even more pronounced for pathways outside of the core carbon metabolism, and for bacterial species that are phylogenetically distant from well-studied model organisms.

The DOE Systems Biology Knowledgebase (KBase) contains a suite of powerful Apps for building genome-scale metabolic models. ModelSEED, originally developed by our collaborator Chris Henry, is the central Flux Balance Analysis model building App in KBase, and one of the most popular metabolic modeling tools for generating draft models because of its accessibility,

ease of use and quality. However, it currently only supports metabolic annotations produced by the Annotate Microbial Genome App, based on RAST (Rapid Annotations using Subsystems Technology). This means that so far it has been impossible for researchers who may prefer to use other high quality annotation tools such as KEGG or even JGI's IMG platform to import their annotations into KBase, let alone merge annotations from multiple sources.

We are developing a set of KBase Apps to allow users to upload functional annotations from popular third-party annotation tools, compare and merge them, and use them for metabolic modeling. (1) An Import App allows the user to upload a simple tab-separated file with annotation data in the form of EC numbers, KEGG or MetaCyc reactions identifiers, or GO identifiers. (2) A Compare App allows the user to compare metabolic annotations from different sources, by mapping all of them to the ModelSEED reaction database. (3) The Merge App will provide the user with a simple yet flexible scoring mechanism to select a preferred set of annotations from among the full set of functional identifiers mapped to each gene in the genome. (4) Finally, we will also assist the KBase metabolic modeling team to make modifications in the existing ModelSEED App and the Gap Filling tool, to enable users to build models from the merged highest-confidence annotations, and prioritize the remaining lower-scoring annotations for gapfilling.



The Import and Compare Apps have been completed, and we will do a demo during the KBase Experience Hands-on Session showing users how they can use some common third party annotation tools to generate additional metabolic annotations, import these into their genome of interest in KBase, and use them for metabolic modeling.

Publications

(1) Griesemer M, Kimbrel JA, Zhou CE, Navid A, D'haeseleer P. Combining multiple functional annotation tools increases coverage of metabolic annotation. BMC Genomics. 2018 Dec 19;19(1):948.

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.

Understanding Variation in the Switchgrass Microbiome Across Scales: Evidence for Both Host Filtering and Environmental Control

Marissa R. Lee[1*] (mrlee2@ncsu.edu), Nathaniel Yang[1], Jennifer Pett-Ridge[2] and **Christine V. Hawkes[1]**

[1]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC; [2]Physical and Life Sciences 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.

Switchgrass (*Panicum virgatum*) is a promising biofuel species that can be cultivated across a large geographic range including marginal lands with low inputs. Recent work demonstrates that components of the switchgrass microbiome can substantially modify plant physiology, growth, and stress tolerance, which can impact both plant production and ecosystem function. However, a key remaining gap in understanding the role of the switchgrass microbiome in plant success is how the microbiome varies across sites, and what controls that variation. To address this, we characterized fungal communities in switchgrass soils, roots, and leaves at 14 sites ranging from the coast to the mountains of North Carolina. The sites varied in climate, soil properties, and management (monoculture, mixed-grass, or tree-intercropping). The samples were sequenced (ITS rRNA) on Illumina MiSeq. We analyzed how plant size, site characteristics, and climate factors affected fungal community structure.

There was significant site-level variation in switchgrass-associated soil, root, and leaf fungi, which was largely explained by local conditions. Fungal community composition in all samples was primarily affected by soil texture, which ranged from very sandy (86% sand) to very clayey (37% clay) across sites. In soils, fungi also varied with pH, P, Mg, and MAT. In contrast, root and leaf fungi were most strongly affected by plant size and soil DOC. Despite shared environmental drivers, there was little overlap in fungal identity between roots and leaves, which shared only 17% of fungal taxa. Soils shared <10% of fungal taxa with roots and <1% with leaves, suggesting that the plant is filtering some taxa from the available pool as well as being colonized by taxa from other sources. Only roots and soil fungi were affected by site management, which mainly reflected the 5x greater SOM in sites with intercropped trees.

Our findings highlight that soil and plant characteristics are associated with divergence in switchgrass microbiomes across landscapes, and that switchgrass-associated fungi are not a

simple reflection of the taxa found in soils. Because we also isolated ~2000 fungi representing ~400 taxa from roots and leaves at these sites, this work also facilitates the next step to relating function to composition across the landscape.

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 a subcontract to North Carolina State University.

KBase Science GSP: Towards a Viral Ecogenomics Toolkit at KBase

Benjamin Bolduc[1], Paramvir S. Dehal[2], **Jennifer Pett-Ridge[3]**, and **Matthew B. Sullivan[1*]** (mbsulli@gmail.com)

[1]The Ohio State University, Columbus OH; [2]Lawrence Berkeley National Laboratory, Berkeley CA; [3]Lawrence Livermore National Laboratory, Livermore CA

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.*

The LLNL Soil Microbiome SFA "Microbes Persist: Systems Biology of the Soil Microbiome" seeks to understand how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial resides in soil under changing moisture regimes. These processes are almost certainly shaped by viral constraints, yet we know very little about viral populations, function and activity in soils. Here we seek to establish viral ecogenomic tools at KBase to empower users to explore the role of viruses in environmental microbial communities by identifying, characterizing, and ecologically contextualizing viruses in large-scale sequence datasets. While many microbial analytical tools exist in KBase, no such viral tools were available until our project. To date, we have integrated two tools into KBase; this has required creating new data models and workflows to handle viral data. First, VIRSorter uses database matches and a probabilistic model to identify virus genomic fragments from viral and microbial metagenomes, providing users with a measure of confidence and evaluation of prophages. Second, when thousands of new viruses are discovered, there is the high likelihood that many are not yet known to science. To taxonomically classify these to approximately genus-level assignments (with confidence metrics), we have integrated vConTACT2, a gene-sharing network analysis tool, into the KBase environment. These tools, along with phage annotation tools developed at LLNL (PHANOTATE and PhATE), form a foundation for integrating the emerging viral ecogenomics toolkit into the KBase ecosystem.

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. Work was also funded by KBase, a Genomic Science program within the U.S. Department of Energy, Office of Science, Office of

Biological and Environmental Research, funded by awards DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

The role of viruses in the carbon cycle along a permafrost thaw gradient

Christine L. Sun[1], Lindsey Solden[1], Ahmed Zayed[1], Adjie Pratama[1], Dylan Cronin[1], Ben Bolduc[1], IsoGenie Project Coordinators[1], Virginia Rich[1], Matthew B. Sullivan[1]* (sullivan.948@osu.edu)
[1]The Ohio State University, OH, USA

Project goals: *The activity of microbes in soil profoundly affects global energy and nutrient cycles. In consequence, whatever shapes soil microbial activity shapes the world. Substantial recent work in environmental microbiology has taught us that viruses are a key driver of microbial ecology in other systems, and we expect the same to hold true in soil. But soil is a complex milieu – highly structured, chemically and physically heterogenous, and resistant to extrapolation. Thus, even as new methods have revolutionized our understanding of microbial and viral ecology in other systems, our understanding of soil microbial ecology has lagged, and our understanding of soil viral ecology is still further behind.* ***Our objective is to develop paradigms for understanding the role of viruses in soil ecology, and to build the tools – scalable new methods, new databases, and new model systems – to test these paradigms.***

Permafrost accounts for approximately 30%-50% of global soil carbon (C). Unfortunately, permafrost is thawing due to elevated temperatures resulting from climate change. It is not entirely understood how thawing permafrost will impact the release of greenhouse gasses, which could in turn accelerate climate change. Microbial communities play a critical role the C cycle but viruses are less well studied in permafrost systems. If marine systems are a guide, then viruses are likely to impact C cycling by controlling microbial communities via predation, transferring genes from one host to another, and metabolically reprogramming their host cells via regulatory take-over and directly, encoding auxiliary metabolic genes (AMGs). Here, we examined viruses along a permafrost thaw gradient over eight years (2010-2017) using data from bulk metagenomic sequencing and viral eco-genomic analyses.

Samples were collected from a palsa (dry permafrost), a bog (intermediate thaw), and a fen (wet, fully thawed) in Stordalen Mire, Sweden. In total, we identified 4,132 unique viral populations (≥ 10kb contigs dereplicated at 95% average nucleotide identity and 80% coverage) that clustered into nearly 614 novel viral genera via gene-sharing networks. This dataset increases both known permafrost viral populations and viral genera by 2-fold. Viral populations were site specific, with only 16 viral populations (0.2%) shared across all sites and 336 viral populations (4%) shared across any two sites. To evaluate how viruses might be playing a role C cycling, we *in silico* predicted hosts for our viruses, using metagenome-assembled-genomes (MAGs) that were generated from the same samples. Preliminarily, we found 26 MAGs that serve as hosts for 42 viruses, using CRISPR spacer matching. Of the MAGs identified as hosts, the majority (77%) are putative key C cyclers, implicating viruses in C cycling via infection. Building upon this work, we are optimizing virus-host prediction by using a combination of different computational methods in order to assign hosts to more permafrost viruses. In addition, high throughput methods are being used to detect directly encoded AMGs in viral sequences. Together, these viruses will be used to further our understanding of the role and impact of viruses in permafrost.

This study was funded by the Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research, proposal #0000248445.

Viruses may manipulate the global carbon cycle through carbohydrate active enzymes

Lindsey Solden[1], Sabina Leanti La Rosa[2], Simon Roux[3], Cristina Howard-Varona[1], Christine Sun[1], Rebecca Daly[4], Michael Wilkins[4], Virginia Rich[1], Phillip Pope[2], **Matthew Sullivan[1]**

[1]The Ohio State University, OH, USA; [2]Norwegian University of Life Sciences, NO; [3]Lawrence Berkeley National Laboratory, CA, USA; [4]Colorado State University, CO, USA

Project Goals: *The activity of microbes in soil profoundly affects global energy and nutrient cycles. In consequence, whatever shapes soil microbial activity shapes the world. Substantial recent work in environmental microbiology has taught us that viruses are a key driver of microbial ecology in other systems, and we expect the same to hold true in soil. But soil is a complex milieu – highly structured, chemically and physically heterogenous, and resistant to extrapolation. Thus, even as new methods have revolutionized our understanding of microbial and viral ecology in other systems, our understanding of soil microbial ecology has lagged, and our understanding of soil viral ecology is still further behind.* ***Our objective is to develop paradigms for understanding the role of viruses in soil ecology, and to build the tools – scalable new methods, new databases, and new model systems – to test these paradigms.***

Viruses that infect microorganisms are important players in the global carbon cycle, with relatively extensive evidence in oceans and emerging evidence in soils. Besides lysing key microbial players, viruses also metabolically reprogram their hosts – indirectly through regulatory rewiring and directly through virus-encoded Auxiliary Metabolic Genes (AMGs). In the oceans, key viral AMGs include those that impact central C metabolism, photosynthesis, and N and S cycling. In soils, however, carbohydrate active enzymes (CAZymes) may represent novel types of AMGs, though their study has only recently begun. CAZymes catalyze the breakdown, biosynthesis, or modification of carbohydrates and glycoconjugates, releasing soluble carbon to the environment, and thus representing markers of viral importance in soils.

Here, we surveyed over 1.2 million publicly available viral genomes from diverse environments to identify the frequency and ecological context of virus-encoded CAZymes. The frequency of viral-encoded CAZymes was higher than that of well-studied viral hallmark genes (e.g. terminase), with CAZymes found in ~20% of surveyed viruses. The most common (10% of the total) CAZyme domains were GH24, GH23 and GH19 which are often found in lysozyme, viral baseplate proteins and tail fibers, and likely aid the virus during host cell entry. Further, the CAZyme domains in lysozyme and baseplate proteins are specific to host cell wall type, with viruses that infect gram positive bacteria often containing GH25 and GH73, while viruses that infect gram negative bacteria containing GH23, GH24 and GH19. Using RefSeq viral isolate genomes, we developed a random forest model leveraging the frequency of CAZyme domains to predict host cell wall type. The result is a model that predicts with up to 86% accuracy, whether the virus infects eukaryotes, gram-positive bacteria, or gram-negative bacteria. Such a model can aid current prediction tools in increasing the accuracy of predicting virus-host relationships in the environment.

Other identified CAZyme domains are environment-specific, such as CAZyme domain AA10 which is detected across almost all soil environments (97% of viruses with CAZymes). This cellulose-oxidizing enzyme may aid viral hosts in degrading plant material, which is observed in other environments. Similarly, marine viral contigs contain CAZymes to break down many algal polysaccharides such as alginate, carageenan, and heparin, and viruses from both human and ruminant microbiomes contain CAZymes for degrading mucin. These CAZymes may provide the machinery for their bacterial hosts to utilize different carbon substrates, altering the carbon cycle by silently manipulating infected cells. Together, our global informatics survey of viral CAZymes helps to further understand the role of viruses in soil carbon cycling.

This study was funded by the Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research, proposal #0000248445.

Novel Microbial Routes to Synthesize Industrially Significant Precursor Compounds

William R. Cannon[3]* (William.Cannon@pnnl.gov), Justin A. North[1], Adrienne Narrowe[2], Kelly C. Wrighton[2], and **F. Robert Tabita**[1]*

[1]The Ohio State University, Columbus, OH; [2]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 recently discovered microbial processes is tenable using plentiful lignocellulose and/or CO_2 feedstocks. The overall long-term objective is to develop an industrially compatible microbial process to synthesize ethylene in high yields. The discovery of a novel and genetically regulated anaerobic pathway to produce high levels of ethylene (the DHAP ethylene pathway) impacts the following specific aims:

1. Fully probe the catalytic potential of all enzymes of the DHAP ethylene pathway and determine the regulatory mechanism of DHAP-ethylene pathway gene expression. Model the thermodynamics and kinetics of ethylene synthetic pathways. (Tabita and Cannon)

2. Discover effective and active ethylene enzymes encoded in cultured and uncultured organisms from anoxic environments. (Wrighton)

3. Construct a modular set of optimized genes (from Aims 1 and 2) on a DNA fragment containing specific regulatory elements that will allow high level gene expression in model organisms that have been flux optimized. (Tabita, Wrighton, and Cannon)

Abstract

We have discovered a novel pathway for recycling distinct byproducts of enzymatic processes requiring S-adenosyl-L-methionine (SAM) as a cofactor [1,2]. This pathway, called the DHAP Shunt, consists of a phosphorylase (MtnP), isomerase (MtnA), and a novel class II aldolase (Ald2) for the salvage of 5'-methylthioadenosine (MTA) (**Fig. 1B**) [1]. A gene cluster for this pathway is widespread in nature (~ 10 % of all bacteria) and this pathway has been verified in *R. rubrum, R. palustris*, *B. thuringiensis* and Extraintestinal Pathogenic *E. coli* strains [1-4]**.** Intriguingly, during *R. palustris* aerobic growth, copious amounts of methionine precursor, methanethiol (CH_3-SH), is produced from (2-methylthio)ethanol (**Fig. 1C,** DHAP-Methanethiol Shunt) [3]. Conversely, **during anaerobic growth we discovered that (2-methylthio)ethanol was further metabolized to methionine, with copious amounts of ethylene gas produced in the process** (**Fig. 1C**, DHAP-Ethylene Shunt) [1]. **This is the first reported anaerobic route to ethylene, involving novel genes and enzymes.** Based on the discovery of this novel and genetically regulated anaerobic pathway to produce high levels of ethylene, the following specific aims are addressed to enhance ethylene production:

Specific Aim 1: Fully probe the catalytic potential of all enzymes of the DHAP ethylene pathway (Tabita and Cannon)**.** SAM is also utilized by Radical SAM and methyltransferase enzymes, producing 5'-deoxyadenosine (5'dAdo) and S-adenosylhomocysteine as byproducts, respectively. We determined the native flux of 5'dAdo and MTA through the DHAP shunt under aerobic and anaerobic growth conditions in *R. rubrum*. During anaerobic growth, 5'dAdo and MTA were metabolized at a rate of ~590 and ~220 nmol/h/g dry cell weight, respectively. During aerobic growth this rate was 100-fold less for both compounds. No metabolism of S-adenosyl-L-homocysteine by the DHAP shunt was observed. This is consistent with subsequent kinetic measurements of purified DHAP shunt enzymes. The DHAP shunt phosphorylase, isomerase, and aldolase were specific for both MTA and 5'dAdo with near equal catalytic efficiency, but not active with S-adenosylhomocysteine (**Fig. 1**) [2]. This indicates that competition of SAM usage between Radical SAM enzymes producing 5'dAdo and those enzymes producing MTA for ethylene production via the DHAP Shunt must be considered for ethylene optimization.

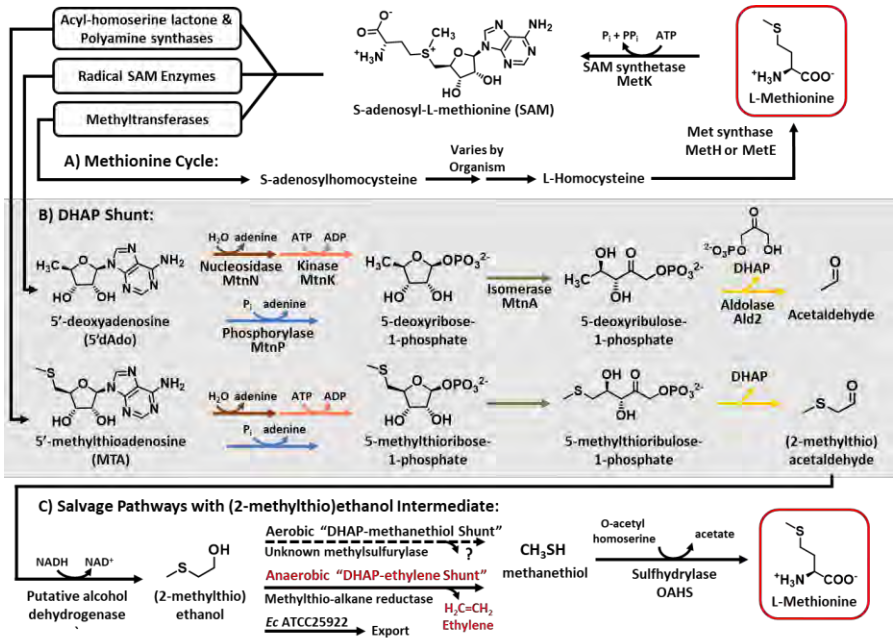


Fig. 1 Salvage of SAM byproducts via the DHAP Shunt Pathways

Understanding catalytic potential through modeling. We are using new physics-based models to evaluate the potential for ethylene production as well. The use of physics-based models are important because whether the end product ethylene is produced in quantity ultimately depends on both catalysis and thermodynamics. The new models are based on the Marcelin-De Donder formulation of mass action kinetics, which differs from the usual kinetic rate laws because it includes both kinetic and thermodynamic terms.

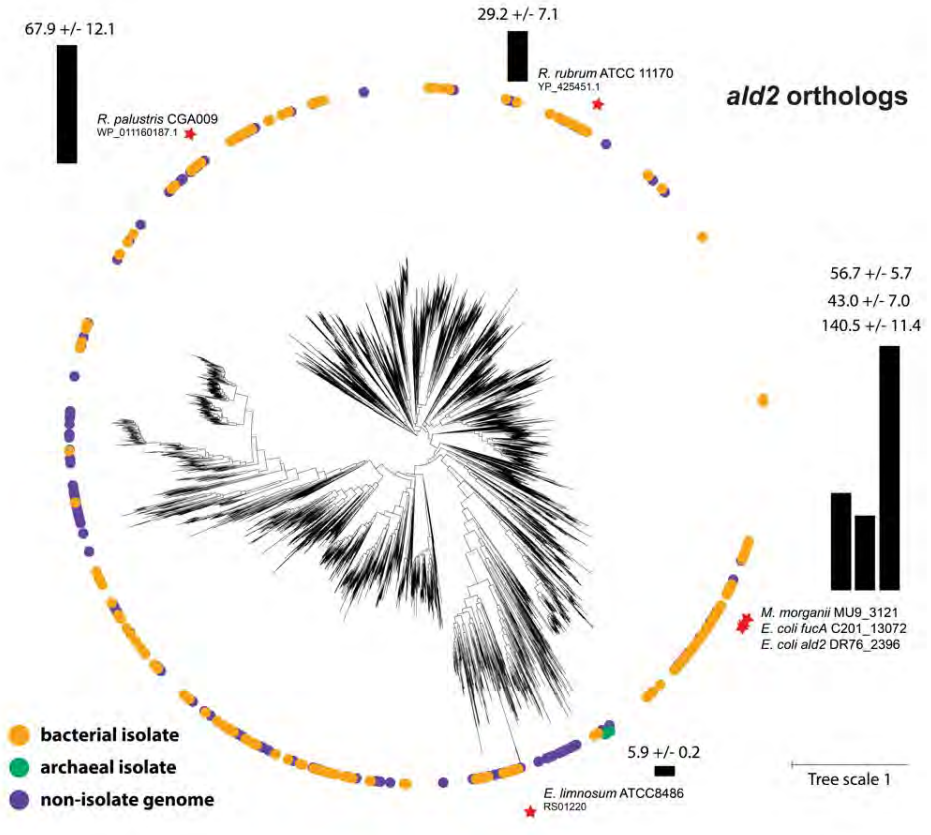
Currently, the metabolic model includes: the Calvin-Benson-Bassham (CBB) cycle, the non-oxidative pentose phosphate pathway, glycolysis, ethanol utilization, the tricarboxylic acid cycle (TCA), the glyoxylate shunt, glutamine and aspartate synthesis, homoserine synthesis, both the S-adenosyl-methionine cycles I and II, methionine synthesis and the DHAP cycle. To date, this is one of the largest thermokinetic models of this kind that have been constructed.

The model correctly fixes CO_2 via the CBB cycle and grows on ethanol in *R. palustris*. Conversion of ethanol to acetaldehyde, acetate and finally acetyl-CoA is highly favored. Acetyl-CoA then feeds into the TCA cycle. The model also predicts that acetate production from *O*-acetyl-L-homoserine via *O*-acetyl-L-homoserine acetate-

lyase is highly favored, despite the fact that this reaction is also predicted to be highly controlled by post-translational regulation.

But acetate is also an end-product of ethylene synthesis in the DHAP pathway. Consequently, because the production of acetate is highly favored from ethanol utilization and O-acetyl-L-homoserine degradation, the driving force for ethylene synthesis is only as strong as the thermodynamic force removing ethylene from the system due to desolvation. It is not clear at this point that this is experimentally the case, as the model is a reduced model of *R. palustris* metabolism. Further improvements to the model are required before any firm conclusions can be drawn regarding the driving force for ethylene production. However, the modeling analysis to-date predicts that the relative amounts of methane, ethane and ethylene are consistent with our experimental data (See abstract and poster *Novel nitrogenase-like C-S lyases link bacterial anaerobic methionine salvage to ethylene and methane production*).

Specific Aim 2: Discover effective and active ethylene enzymes encoded in cultured and uncultured organisms from anoxic environments (Wrighton). Previous studies had indicated that aldolase and isomerase orthologs genes could substantially enhance ethylene levels relative to endogenous genes (1, and unpublished studies). To scale up the search for ethylene-enhancing orthologs we applied targeted functional metagenomics to systematically query genes from the environment. Mining of JGI IMG/M genome and metagenome sequence databases for candidate orthologs to the MTR- 1P isomerase (*mtnA*) and the MTRu-1P aldolase (*ald2*) genes yielded 1,371,813 and 96,049 candidate genes for *mtnA* and *ald2*, respectively. Subsequent filtering based on synteny, phylogeny, and sequence homology with experimentally validated enzymes yielded over 2500 candidate orthologs for each target gene, from multiple metagenomic samples covering a wide variety of environments including wetlands, forest soils, rhizosphere, and bioreactors (**Fig. 2**).



In order to both ensure recovery of active enzymes and maximize sampling of undefined biochemical diversity, we selected two hundred pairs of candidate *mtnA / ald2* genes considering proximity to high-functioning orthologs and phylogenetic breadth. These genes have been synthesized by the JGI DNA Synthesis Science program for screening via our high throughput lysate activity assay. Optimal orthologs will be further validated via physiological complementation to assess enhanced ethylene production.

References:

1. North, J. A., Miller, A. R., Wildenthal, J. A., Young, S. J. & Tabita, F. R. Microbial pathway for anaerobic 5'-methylthioadenosine metabolism coupled to ethylene formation. Proc Natl Acad Sci U S A 114, E10455-E10464 (2017)
2. North, J. A., Wildenthal, J. A., Erb, T. J., Evans, B. E., Byerly, K. M., Gerlt, J. A., & Tabita, F. R. A bifunctional salvage pathway for two distinct S-adenosylmethionine byproducts that is widespread in bacteria, including pathogenic *Escherichia coli*. Mol Microbiol (2020) doi: 10.1111/MMI.14459
3. Miller, A. R., North, J. A., Wildenthal, J. A. & Tabita, F. R. Two distinct arobic methionine salvage pathways generate volatile methanethiol in *Rhodopseudomonas palustris*. MBio 9, e00407-18 (2018)
4. Beaudoin, G. A. W., Li, Q., Folz, J., Fiehn, O., Goodsell, J. L., Angerhofer, A., Bruner, S. D., & Hanson A. D. Salvage of the 5-deoxyribose byproduct of radical SAM enzymes. Nat Commun 9, 3105 (2018)

Funding:

This work is funded by DOE BER grant DE-SC0019338 to Tabita, Wrighton, and Cannon.

Novel nitrogenase-like C-S lyases link bacterial anaerobic methionine salvage to ethylene and methane production

Justin A. North[1]* (north.62@osu.edu), Adrienne B. Narrowe[2], Weili Xiong[3], Kathryn M. Byerly[1], Guanqi Zhao[1], Sarah J. Young[1], Srividya Murali[1], Johnathan A. Wildenthal[1], William R. Cannon[4], Kelly C. Wrighton[2], Robert L. Hettich[3], and **F. Robert Tabita**[1]

[1]Department of Microbiology, The Ohio State University, Columbus, OH; [2]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO; [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [4]Pacific Northwest National Laboratory, Richland, WA.

Project Goals

Ethylene is the most widely employed organic precursor compound in industry. The recently discovered efficient microbial anaerobic ethylene synthetic processes was previously partially characterized, but the terminal enzyme(s) responsible for ethylene production remained unknown. The long-term goal of this project is to identify, isolate, and characterize the specific enzyme(s) that catalyzed anaerobic microbial 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 in elucidating the unknown ethylene-producing enzyme system are as follows:

1. Identify the genes and gene products in *Rhodospirillum rubrum* responsible for anaerobic ethylene synthesis (Tabita, Hettich, and Wrighton)
2. Determine the regulation and control of identified genes required for anaerobic ethylene synthesis (Tabita and Wrighton)
3. Characterize the enzymes and the reaction they catalyze that directly generates anaerobic ethylene (Tabita and Cannon)

Abstract

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 ubiquitous byproduct of polyamine synthesis and homoserine lactone-based quorum sensing in bacteria. The initial steps of the DHAP-ethylene Shunt are catalyzed by a phosphorylase (MtnP), isomerase (MtnA), a novel aldolase (Ald2), and an alcohol dehydrogenase sequentially convert MTA to dihydroxyacetone phosphate (DHAP) and (2-methylthio)ethanol (Fig. 1). **The (2-methylthio)ethanol serves as a direct precursor to ethylene and methionine via an unresolved process involving unknown genes and enzymes** [1]. **This is the first reported anaerobic route to ethylene, involving novel genes and enzymes.** Here we report the identification of the genes and putative gene products specifically responsible for this anaerobic ethylene process and the regulatory system controlling gene expression for the pathway.

1. Genes and gene products in *Rhodospirillum rubrum* responsible for anaerobic ethylene synthesis: We sought to identify the genes and proteins responsible for ethylene production via proteomics and specific gene deletion studies. 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. Surprisingly, proteins with the highest increase in abundance during production of ethylene from (2-methylthio)ethanol corresponded to novel nitrogenase-like proteins of unknown function and previously characterized O-acetyl-L-homoserine sulfhydrylases [2] (Fig. 1; MarBHDK and OASH).

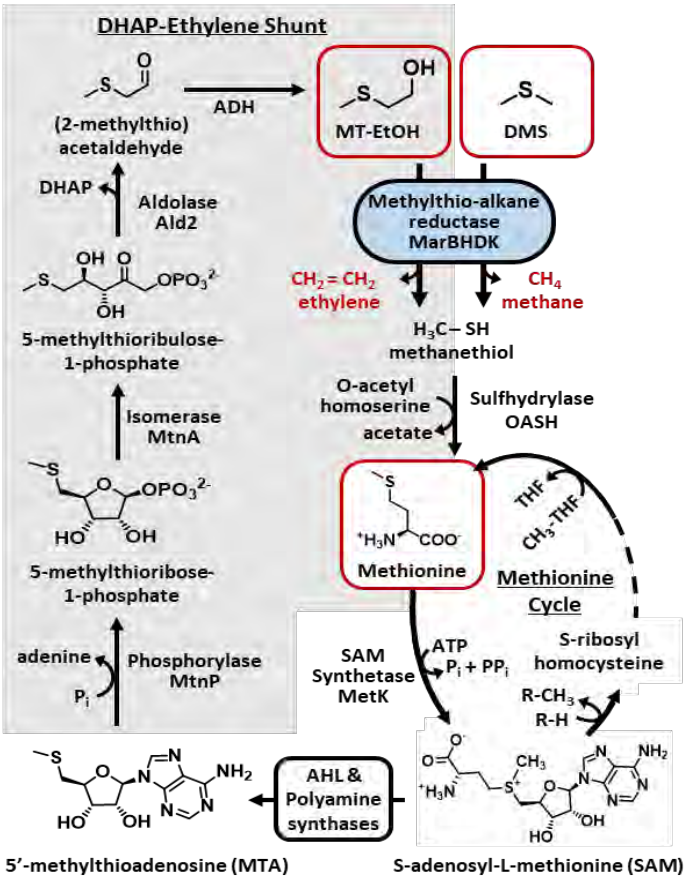


Fig. 1. Full elucidation of the DHAP-ethylene Shunt pathway reveled involvement of novel nitrogenase-like methylthio-alkane reductases (MarBHDK)

Deletion of the genes corresponding to nitrogenase-like proteins (*mar*BHDK), with increased abundance during ethylene inducing conditions, rendered *R. rubrum* incapable of growth or ethylene production utilizing (2-methylethio)ethanol. Reintroduction of these genes expressed *in trans* from a plasmid restored both growth and ethylene production (Fig. 2A). Cells were also grown under inducing conditions for *bona fide* molybdenum nitrogenase (NifHDK) (glutamate as N-source). While maximal acetylene reduction (nitrogenase activity) was observed, nitrogenase was unable to perform the methylthio-alkane reductase activity of MarBHDK to convert (2-methylthio)ethanol to ethylene. **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** (Fig. 1).

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 in order to be utilized as a sulfur source by *R. rubrum* for growth and methionine metabolism (Fig. 2A). Utilization of dimethyl sulfide, the most abundant VOSC in the environment, resulted in stoichiometric production of methane, and ethylmethyl sulfide led to stoichiometric production of ethane gas. Utilization of larger VOSCs such as (3-methylthio)propanol did not require the nitrogenase-like gene products and no gaseous hydrocarbons were detected from growth and methionine metabolism from larger VOSCs. Each putative reaction to produce a gaseous hydrocarbon from its corresponding VOSC was modeled to determine the change in standard Gibbs free energy for the overall reaction in solution (Fig. 2B). Standard free energies were calculated with *NWChem*, a computational chemistry electronic structure software developed under BER funding. The standard free

energies of reaction predict that ethylene formation will be the most favorable. Likewise, the relative difference in gas production observed experimentally between wildtype and Δ*mar*BHDK also suggests, like the standard free energies, that ethylene production from (2-methylthio)ethanol is the most favored catalytic product.

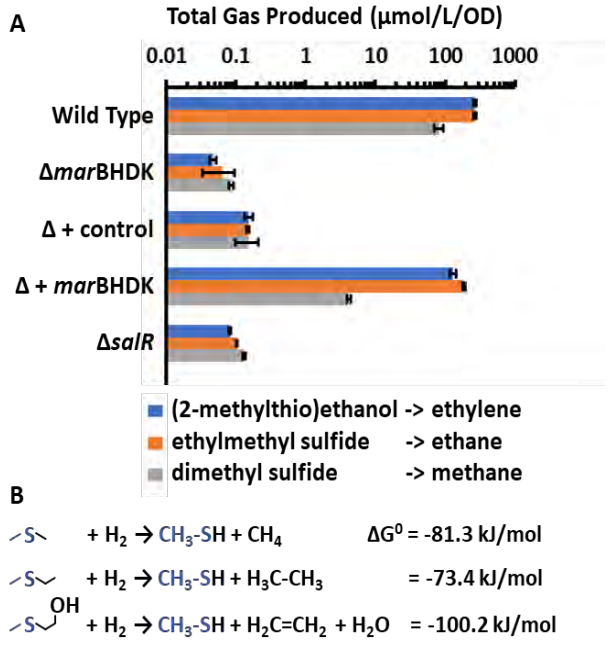


Fig. 2. Proposed reactions catalyzed by nitrogenase-like methylthio-alkane reductases (A) Total ethylene, ethane, or methane produced upon feeding *R. rubrum* with (2-methylthio)ethanol, ethylmethyl sulfide, or dimethyl sulfide, respectively. **(B)** Free energy calculations for each proposed reaction.

2. Regulation of *marBHDK* gene expression: A random mutagenesis screen of *R. rubrum* identified multiple isolates in which an integrated transposon was inserted into a putative LysR family regulator gene *(salR*) located 7 genes away from the *marBHDK* operon. Such mutants were unable to grow or utilize (2-methylthio)ethanol, ethylmethyl sulfide, or diemethyl sulfide as sulfur source for growth or produce any corresponding hydrocarbon gasses (Fig. 2A; Δ*salR*). However, they were still able to grow normally using sulfate as sole sulfur source. Subsequent RNA-seq differential gene expression analysis revealed that the *marBHDK* gene cluster and O-acetylhomoserine sulfhydrylase genes for methionine synthesis (Fig. 1) were upregulated during ethylene inducing conditions (limiting sulfate) only when a functional *salR* gene was present. No changes in expression levels of these genes were observed in the Δ*salR* deletion strains. **Therefore, this LysR-like regulator is termed SalR for Sulfur Salvage Regulator of genes responsible for sulfur salvage from VOSCs and other organic sulfur compounds.**

3. Other Organisms with Methylthio-alkane Reductases:

The *R. rubrum* MarHDK gene products are homologous to nitrogenase proteins NifH (nitrogenase reductase iron protein), and NifDK (nitrogenase catalytic subunits α/β) and are part of the Group IV Nitrogen Fixation-Like clade (Fig. 3). While the *bona fide* nitrogenases (Groups I-III) catalyze nitrogen fixation, the Nitrogen Fixation-Like members of known function catalyze reduction of other distinct compounds containing double bonds such as porphyrin ring reduction during bacteriochlorophyll and F430 cofactor biosynthesis (Fig. 3; Groups V and IV-B). However, save for IV-B and a single group IV member from *Endomicrobium proavitum* (Fig. 3; Group IV-A) [5], which can also catalyze N_2-fixation, the function of the remaining Group IV members is unknown. The homologous methylthio-alkane reductase proteins form their own clade that is distinct from other nitrogenase and nitrogenase-like sequences, suggesting other organisms can utilize environmental VOSCs, producing ethylene and methane in the process (Fig. 1; Group IV-C). Organisms with putative methylthio-alkane reductases include members from alpha-proteobacteria, particularly rhodospirilla and rhizobia, and firmicutes, particularly selenomonads and ruminal clostridia.

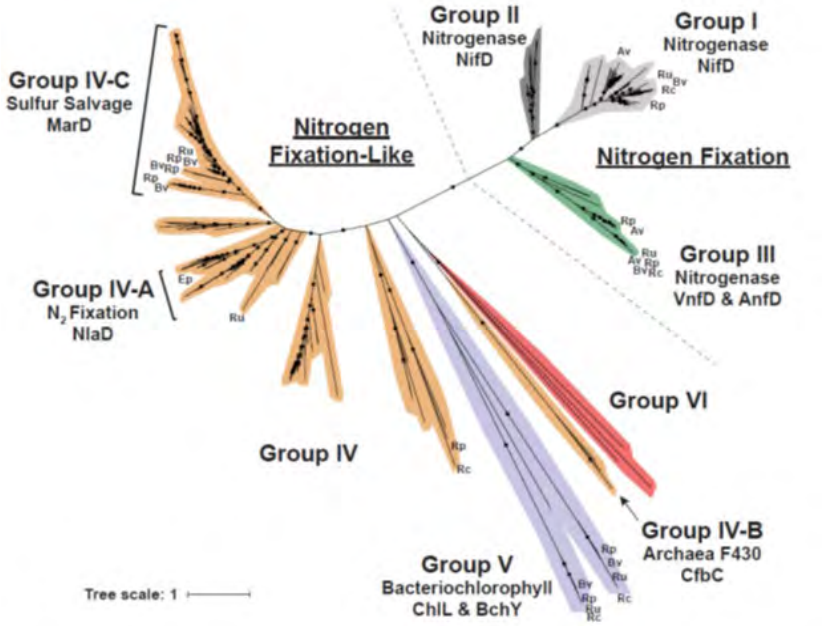


Fig. 3. Phylogenetic tree of NifD superfamily based on an LG+R6 evolution model. The scale bar represents the number of substitutions per site. UFBoot support values of 95% or greater are shown as black circles on branches. Clade coloring and numbering follows Raymond [3] and Méheust [4].

References:

1. North, J. A., Miller, A. R., Wildenthal, J. A., Young, S. J. & Tabita, F. R. Microbial pathway for anaerobic 5'-methylthioadenosine metabolism coupled to ethylene formation. Proc Natl Acad Sci U S A 114, E10455-E10464 (2017)
2. Erb, T. J., Evans, B. S., Cho, K., Warlick, B. P., Sriram, J., Wood, B. M., Imker, H. J., Sweedler, J.V., Tabita, F. R. & Gerlt, J. A. A RuBisCO-like protein links SAM metabolism with isoprenoid biosynthesis. Nat Chem Biol 8, 926–932 (2012)
3. Raymond, J., Siefert J. L., Stales C. R. & Blankenship R. E. The natural history of nitrogen fixation. Mol Biol Evol 21, 541-54 (2004)
4. Méheust, R., Castelle, C. J., Carnevali, P. B. M., Farag, I. F., He, C., Chen, L. X., Amano, Y., Hug, L. A. & Banfield, J. F. Aquatic Elusimicrobia are metabolically diverse compared to gut microbiome Elusimicrobia and some have novel nitrogenase-like gene clusters. bioRxiv 765248; doi: https://doi.org/10.1101/765248
5. Zheng, H., Dietrich, C., Radek, R. & Brune, A. Endomicrobium proavitum, the first isolate of Endomicrobia class. nov. (phylum Elusimicrobia)--an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. Environ Microbiol. 18, 191-204 (2016)

Funding:

This work was supported in part by the University of Colorado Cancer Center's Genomics and Microarray Core Shared Resource funded by NCI grant P30CA046934. This work utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (awards ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. The electronic structure calculations were performed at the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the U.S. DOE OBER and located at PNNL. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under contract DE-AC05-76RLO 1830. This work was supported by an OSU Center for Applied Plant Sciences Seed Grant (to F.R.T) and the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019338 (to F.R.T, K.C.W., and W.R.C.). The proteomics work at ORNL was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program funding.

Corrinoids as model nutrients to probe microbial interactions in a soil ecosystem

Zachary Hallberg,[1] Zoila Alvarez-Aponte,[1] Alexa Nicolas,[1] and **Michi Taga (taga@berkeley.edu)[1]**

[1]Department of Plant and Microbial Biology, University of California, Berkeley

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 will examine 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 gain an unprecedented view of metabolic interactions in a soil microbial community.

Microbial communities in soil are key drivers of biogeochemical cycling and plant growth. Current research methods in soil microbial ecology cannot simultaneously analyze all microbe-microbe interactions in situ, making it difficult to identify keystone interactions that could predictably modulate these communities. We aim to overcome this limitation by focusing on 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. However, in contrast to many other essential cofactors, over a dozen different structural forms of corrinoids exist, and different groups of bacteria require different corrinoids for metabolic needs that play central roles in bioremediation, plant-bacteria symbiosis, organic matter accumulation, and elemental cycling. Here, we test the hypothesis that corrinoids are a key nutrient family in soil microbial ecosystems. We show that a significant number of bacteria are corrinoid-dependent by performing a pilot phenotype-guided isolation of corrinoid-dependent bacteria from soil. Furthermore, we demonstrate that a bioinformatics-to-experiment pipeline can rapidly predict and verify the corrinoid production capacity of soil isolates.

These results are being used to guide corrinoid-based perturbations of soil samples and soil-derived enrichment cultures. We expect that providing excess corrinoid will remove corrinoid

synthesis or availability as a barrier to growth, leading to increased abundance of microbes capable of using the added corrinoid. At the same time, growth of microbes incapable of using the added corrinoid should decrease if the corrinoid interferes with metabolism. These corrinoid-induced changes in the community could potentiate other metabolic networks, leading to differences in the growth of community members that are not directly affected by the corrinoid amendment. Our proposed holistic approach to understanding corrinoid cycling in soil addresses a key issue in soil microbial ecology, namely that soil is too complex to simultaneously understand all microbe-microbe interactions in a whole-community context and the metabolites involved in those interactions. However, by focusing on one key shared, yet structurally diverse, metabolite used by a majority of the community, we can pinpoint key interactions for further study.

This research is funded by the DOE Genomic Sciences Program Grant DE-FOA-00002059 (to MT).

Microbial controls on biogeochemical cycles in permafrost ecosystems

Neslihan Taş[1*], Megan Dillon[1], Yaxin Xue[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]US Geological Service, [3]GFZ German Research Centre for Geosciences Postdam, [4]California State University Northridge

Project Goals: The permafrost carbon reservoirs are currently protected from microbial decomposition by frozen conditions. Upon permafrost thaw microbial metabolism leads to decomposition of soil organic matter, substantially impacting the cycling of nutrients and significantly affecting the arctic landscape. This project use state of the art molecular techniques to resolve complex microbial processes governing the biogeochemical cycles in arctic soils and permafrost to better inform efforts to access uncertainties surrounding ecosystem responses.

Permafrost microbial communities are complex, diverse, and active at subzero temperatures. While carbon turnover at depth is proposed to be slower than surface, especially the fate of carbon in deep permafrost, which is currently protected from the warming climate, is uncertain. Permafrost microbiome is a seed bank of mostly novel organisms that have a diverse and broad metabolic potential. 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 (Alaska, Sweden, Norway, Canada and Russia) showed stark differences in microbial populations and metabolic functions that are not strong driven by environmental conditions (ice content, topography, continuity, active layer depth, and vegetation) or soil chemistry. Recognizing geospatial patterns in soil properties and microbiome characteristics across Arctic permafrost landscapes will allow us to better inform on how permafrost microbes can respond to global climate change.

The microbial response to thaw in arctic environments is not uniform and the relationship between permafrost microbiomes and greenhouse gas (GHG) emissions is not well understood. Following thaw, redistribution of water is a key event that conditions the permafrost for microbial decomposition. We initiated batch-scale permafrost incubation experiments dry, natural, and saturated moisture states and under microaerophilic or anaerobic headspaces. Via metagenomics and metatranscriptomics we dissect the microbial response leading to fermentation and competition between methanogenesis and iron and/or sulfate reduction processes, highlighting the importance of interactions between iron, sulfur and carbon metabolism. We couple omics 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). Analysis showed that variety of organic compounds and metabolites were accumulated in thawed permafrost soils. Especially under saturated conditions while carbohydrates were depleted, soils accumulated aliphatic compounds.

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 in different arctic soils.

This research is supported by DOE Early Career Program by the Office of Biological and Environmental Research in the DOE Office of Science.

Title: Understanding the genetic basis of drought tolerance in bioenergy poplar

Gail Taylor gtaylor@ucdavis.edu [1,2], Jin-Gui Chen[3] and Wellington Muchero[3]

[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. 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: 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 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 will validate our findings on these regulator cis-acting eQTN candidates for drought tolerance in fast transient assays using protoplasts and in longer-term whole tree experiments including accurate deployment of CRISPR-Cas9 gene edits. The innovation in this project comes from the rapid deployment of these approaches that will also benefit from findings in a second *Populus* natural collection from a range of wet and droughted sites across Europe for which early leads on target loci are already available for validation.

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.*

(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.*

(v) Using RNAseq in *P. trichocarpa*, in droughted and control treatments to identify cis- and trans-regulated eQTN.

(vi) Validate up to 50 cis-QTNs, from network hubs using transient protoplast assays.

(vii) To establish Agrobacterium-based gene editing protocols in *Populus*.

(viii) To utilize early leads from previous research to investigate at least 6 candidate genes for drought tolerance in *Populus.*

(ix) To validate up to 20 candidate genes for drought tolerance in *P. trichocarpa* refined from the long-list tested in the transient assays for cis-acting hub gene targets.

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. They will be ready to test at multiple sites at project end. At the same time, the genomic and genetic resources already available in *Populus* will be integrated into high throughput drought phenotyping and a complete atlas of gene expression related to drought stress for this natural population.

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.

Understanding and Eliminating the Detrimental Effect of Thiamine Deficiency on the Oleaginous Yeast *Yarrowia lipolytica*

Caleb Walker[1*] (fyz692@vols.utk.edu), Seunghyun Ryu, [2] Richard J. Giannone, [3] Sergio Garcia, [1] and **Cong T. Trinh**[1]

[1]University of Tennessee, Knoxville; [2] CJ Research Center America, Woburn, Massachusetts; [3]Oak Ridge National Laboratory, Oak Ridge, TN;

Project Goals: To understand thiamine deficiency on cellular metabolism, elucidate genetic basis causing thiamine auxotrophy, and identify thiamine-responsive elements in *Yarrowia lipolytica*.

Abstract text. Thiamine is a vitamin that functions as a cofactor for key enzymes in carbon and energy metabolism for all living cells. While most plants, fungi and bacteria can synthesize thiamine de novo, the oleaginous yeast, *Yarrowia lipolytica*, cannot. In this study, we used proteomics together with physiological characterization to understand key metabolic processes influenced and regulated by thiamine availability and identified the genetic basis of thiamine auxotrophy in *Y. lipolytica*. Specifically, we found thiamine depletion results in decreased protein abundance of the lipid biosynthesis pathways and energy metabolism (i.e., ATP synthase), attributing to the negligible growth and poor sugar assimilation observed in our study. Using comparative genomics, we identified the missing 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase (THI13) in the de novo thiamine biosynthesis of *Y. lipolytica*, and discovered an exceptional promoter, P3, that exhibits strong activation or tight repression by low and high thiamine concentrations, respectively. Capitalizing on the strength of our thiamine-regulated promoter (P3) to express the missing gene from *Saccharomyces cerevisiae* (scTHI13), we engineered the thiamine-prototrophic *Y. lipolytica*. By comparing this engineered strain to the wildtype, we unveiled the tight relationship linking thiamine availability to lipid biosynthesis and demonstrated enhanced lipid production with thiamine supplementation in the engineered thiamine-prototrophic *Y. lipolytica*.(1)

References

1. Walker C, Ryu S, Giannone RJ, Garcia S, Trinh CT. 2020. Understanding and Eliminating the Detrimental Effect of Thiamine Deficiency on the Oleaginous Yeast *Yarrowia lipolytica*. Applied and Environmental Microbiology 86:e02299-19.

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.

Understanding and Harnessing the Exceptional Robustness of *Yarrowia lipolytica* for the Conversion of Biomass Hydrolysate into Designer Bioesters.

Caleb Walker,[1] Seunghyun Ryu,[2] Stephanie Thompson,[3] Richard Giannone,[4] Patricia J. Slininger,[3] Bruce S. Dien,[3] and **Cong T. Trinh**[1*] (ctrinh@utk.edu)

[1]University of Tennessee, Knoxville; [2]CJ Research Center America, Woburn, Massachusetts [3]National Center for Agricultural Utilization Research, USDA-ARS, Peoria, IL; [4]Oak Ridge National Laboratory, Oak Ridge, TN;

Project Goals: To elucidate and harness the exceptional robustness of novel, undomesticated *Y. lipolytica* isolates from genetic diversity screening as a bioenergy-relevant microbial platform for efficient conversion of undetoxified biomass hydrolysates into designer bioesters with continuous recovery using solvent extraction.

Abstract text. Robustness is an important phenotype for bioenergy microbes to acquire but is difficult to engineer. The oleaginous yeast, *Yarrowia lipolytica*, is an exceptionally robust microbe that can tolerate stressful environments,[1] assimilate a wide range of substrates[2,3] and produce high-value chemicals.[4] In this study, we aim to understand and harness these robust characteristics of *Y. lipolytica* for the conversion of biomass hydrolysate into designer bioesters. Specifically, we aim to understand how these *Y. lipolytica* i) tolerate and effectively assimilate inhibitory biomass hydrolysates for superior lipid accumulation, ii) tolerate organic solvents that are required to produce bioproducts in a two-phase fermentation system and iii) endogenously degrade lipids to produce designer bioesters. From a screen of 57 undomesticated *Y. lipolytica* isolates,[5,6] we selected top-performing strains exhibiting robust growth and lipid accumulation in undetoxified biomass hydrolysate for comprehensive growth and multi-omics characterization. Comparative proteomic analysis of these robust *Yarrowia* strains revealed previously uncharacterized proteins that influence robust utilization of biomass hydrolysate sugars and transcription factors regulating lipid accumulation and degradation. We further engineered these strains to improve their robustness for effective conversion of these sugars for lipid production. Next, transcriptomic analysis of an exceptionally solvent-tolerant *Yarrowia* mutant identified genes conferring high solvent tolerance. These foundational insights provide key mechanisms and genetic targets to engineer robustness in *Yarrowia* strains for production of biofuels and bioproducts from lignocellulosic biomass.

References

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 Ryu, S., Hipp, J. & Trinh, C. T. Activating and Elucidating Metabolism of Complex Sugars in *Yarrowia lipolytica*. *Applied and Environmental Microbiology* **82**, 1334-1345, doi:10.1128/aem.03582-15 (2016).

3 Ryu, S. & Trinh, C. T. Understanding Functional Roles of Native Pentose-Specific Transporters for Activating Dormant Pentose Metabolism in *Yarrowia lipolytica*. *Applied and Environmental Microbiology* **84**, doi:10.1128/aem.02146-17 (2018).

4 Ryu, S., Labbé, N. & Trinh, C. T. Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of α-ketoglutaric acid by Yarrowia lipolytica. *Applied microbiology and biotechnology* **99**, 4237-4244, doi:10.1007/s00253-015-6521-5 (2015).

5 Quarterman, J., Slininger, P. J., Kurtzman, C. P., Thompson, S. R. & Dien, B. S. A survey of yeast from the Yarrowia clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate. *Applied microbiology and biotechnology* **101**, 3319-3334, doi:10.1007/s00253-016-8062-y (2017).

6 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.

Employing Bacterial Microcompartments To Create Privileged Redox Pools for Biofuel Production

Svetlana P. Ikonomova[1*] (svetlana.ikonomova@northwestern.edu), Sasha Shirman [1*] (aleksandra.shirman@northwestern.edu), Charlotte H. Abrahamson,[1] Carolyn E. Mills,[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 how we are coupling modeling with experiments to first understand the native function of the 1,2-propanediol utilization microcompartments (MCPs), and particularly the recycling of native cofactors. Using statistical and structural analysis of our model we have determined several guidelines for experimental design, including 1) using variation in the compartment number to disambiguate between the membrane and microcompartment permeability values for future cell-based experiments and 2) optimization of sampling during time-series to capture maximally informative features. We then established an *in vitro* method for analyzing pathway performance of purified MCPs, and detected conversion of 1,2-propanediol to the intermediate propionaldehyde and product 1-propanol. Interestingly, exogenous addition of adenosylcobalamin (AdoB12) was required to activate consumption of 1,2-propanediol. Moreover, we present evidence that cofactors including AdoB12, ATP, and coenzyme A can enter the MCP lumen during the *in vitro* reaction, but also that activity depends on recycling of NAD^+ and NADH. We are currently applying the findings from the native

system towards improving the performance of our first target metabolic pathway: 1,3-propanediol production.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019337.

Building a suite of CRISPR/Cas9 tools for efficient switchgrass gene editing

Eudald Illa-Berenguer[1]* (eillaberenguer@uga.edu) Peter LaFayette[1,2], Wayne Parrott[1,2,3] and **Gerald A. Tuskan**[4]

[1]Center for Applied Genetic Technologies, University of Georgia, Athens, GA; [2] Department of Crop and Soil Sciences, University of Georgia, Athens, GA; [3]Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA, USA; [4]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

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.

CRISPR/Cas9 technology has become the genome editing tool of choice in almost all kingdoms of life and reports of its use in organisms that range from protists, to fungi, animals and plants are widespread in the literature. By capitalizing the increasing knowledge on how this system works, we aim to develop an efficient and predictable gene editing system for switchgrass (*Panicum virgatum L.*). Switchgrass has been recognized as an important biomass resource of fermentable mixed sugars that can yield biofuels and other value-added chemicals and biomaterials. Being able to fine-tune switchgrass metabolic capabilities via CRISPR/Cas9 gene editing would be extremely useful to exploit its potential and well as to validate putative gene functions. However, switchgrass is an allotetraploid, whose two subgenomes show extensive gene duplication. This makes the identification of gene-specific targets very challenging.

Fortunately, since CRISPR/Cas9 system was first described and in parallel to its widespread use, a suite of Cas9/Cas12a proteins with diverse size, activity, recognition target site and trans-activating crRNA (scaffolds) have been characterized. Taking advantage of the wide range of target specificities that the available Cas9/Cas12a proteins offer, we are developing a platform that will allow editing any gene of interest in switchgrass. The ultimate goal is to ensure the editing of all copies in the switchgrass genome. As a first step, we have tested five different Cas9/Cas12a proteins (SpCas9, SaCas9, St1Cas9, Mb3Cas12a, and AsCas12a) in embryogenic rice calli from Taipei-309, aiming to transfer the knowledge acquired in the rice model system into switchgrass. Also, all protein combinations have been tested both at 37°C (optimal temperature for most Cas9/Cas12a proteins) and 27°C (optimal temperature for tissue culture) to assess rate of editing under standard tissue culture conditions. Our results have identified SpCas9 as the best protein in terms of editing efficiency at both 27°C and 37°C. Mb3Cas12a and SaCas9 yielded similar results to SpCas9 but only at 37°C; at 27°C Mb3Cas12a

efficiency dropped by 50% and SaCas9 didn't work at all. This suite of best performing nucleases and working conditions is being tested with gene targets of interest in switchgrass.

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Catalytic Upgrading of n-Butanol to Fully Synthetic Jet Fuel

Zhenglong Li[1]* (liz3@ornl.gov), Brian Davison[1], (davisonbh@ornl.gov), Shiba P. Adhikari[1] and **Gerald A. Tuskan**[1]

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

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.

Upgrading fermentation-based C4 alcohols (n-butanol or isobutanol) into fungible hydrocarbon fuels (e.g., jet, gasoline or diesel) is one promising alcohol-to-jet pathway as the olefin intermediates from C4 alcohols can be readily converted to jet-range hydrocarbons with a high selectivity. Currently there are only very limited reports or patents demonstrating that pure iso-butanol can be upgraded in a multi-step process of dehydration to isobutene followed by oligomerization to form iso-olefins as the dominant hydrocarbon fraction (*1,2*). The hydrocarbons produced from this prior approach have high number of branches, which limits the blending level into jet.

In this work, we target at upgrading of aqueous n-butanol to hydrocarbon fractions that are rich in n-paraffins and isoparaffins with low branches at much milder conditions – this will help to reduce the cost. This project will convert n-butanol to 1-butene rich olefins, which can be readily turned into jet-range hydrocarbons with larger fraction of n-paraffins and lightly branched paraffins (desired fraction in petroleum jet). Meanwhile direct one-step production of aromatic-rich blendstock will be investigated to produce the aromatic component.

We have found that 97% of butenes can be produced with ~100% n-butanol conversion over conventional H-ZSM-5 catalyst at 225 °C and ambient pressure, where 1-butene is the dominating fraction among butene isomers. Increase of space velocity (1 to 8 h^{-1}) and n-butanol concentration (5.5 to 32%) did not show significant impact on n-butanol conversion, C_3-C_6 olefins selectivity and carbon number distributions, which offers potential to significantly reduce the reactor size. The mixed C_3-C_6 olefins were further oligomerized into longer-chain hydrocarbons over Amberlyst-36 catalyst at 150 °C and 15 bar, leading to 70% of jet, 12% of gasoline and 10% of diesel fraction. The liquid hydrocarbons are rich in lightly branched iso-paraffins/olefins and n-paraffins/olefins, which offers great potential for higher blending. Catalyst development with 2D pillared zeolite is planned to further improve the selectivity of 1-butene and then to increase the quality of resulted hydrocarbon blendstock.

References

1. Taylor, J. D., Jenni, M. M., Peters, M. W. (2010). Dehydration of Fermented Isobutanol for the Production of Renewable Chemicals and Fuels. Topics in Catalysis, 53, 1224–1230.
2. Peters, M. W., Taylor, J. D. Renewable Jet Fuel Blendstock from Isobutanol. US Patent : 8,975.461 B2 2015, 23.

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.

Characterization of drought tolerance and water-use efficiency related traits in switchgrass

Yongqin Wang[1,2]* (yqwang@noble.org), Weihong Dong, Marcus Griffiths, Larry York, Michael Udvardi[1,2], Yun Kang[1,2], and **Gerald A. Tuskan**[2]

[1]Noble Research Institute, LLC. Ardmore, Oklahoma; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

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.

In the CBI project, we will characterize drought tolerance and water use efficiency (WUE) related traits in a switchgrass genome-wide association study (GWAS) panel. Causal alleles controlling biomass yield and persistence under drought stress will be identified via GWAS analyses. Key genes and mechanisms underlying important sustainability traits will be further validated to facilitate subsequent genomic selection and maker-assisted breeding programs.

Switchgrass (*Panicum virgatum*) is a promising feedstock for biofuels in the United States. As for most crops, periodic drought often limits its productivity, especially in marginal lands where it is likely to be planted. In this CBI project, we aim to evaluate drought tolerance and WUE related physiological and biochemical traits of a switchgrass GWAS panel comprising 415 sequenced genotypes (Juenger *et al.*, 2016). Due to the outcrossing nature of switchgrass, we optimized a micro-propagation method via node culture (Alexandrova *et al.*, 1996), and have successfully propagated the majority of the switchgrass GWAS panel using this method.

In addition, experimental conditions were optimized by performing a preliminary phenotyping experiment using 12 switchgrass genotypes. Traits characterized included shoot and root biomass, water use efficiency, root architecture, root/shoot ratio, leaf area, leaf thickness, stomatal density, leaf cuticular wax, and leaf osmotic pressure. Among 100 lines that have been characterized so far, large genotypic variations were observed in all the traits. Once the entire GWAS panel has been phenotyped, GWAS analysis will be performed to identify quantitative trait loci (QTL) markers and key genes controlling biomass yield and persistence under drought stress. These results will facilitate subsequent genomic selection, maker-assisted breeding and biotechnology strategies to enhance sustainability in switchgrass production.

References

1. Alexandrova K. S., Denchev P. D., Conger B. V. (1996). Micropropagation of switchgrass by node culture. Crop Science, 36, 1709-1711.
2. Juenger T. E., Schmutz J., Wolke T., Fritschi F., Zare A., Bartley L., Jastrow J., O'Brien S., Matamala R., Watson S., Costich D. (2016). Climate adaptation and sustainability in switchgrass: exploring plant-microbe-soil interactions across continental scale environmental gradients. IN: Genomic Science Contractors–Grantees Meeting XIV and USDA-DOE Plant Feedstock Genomics for Bioenergy Meeting.

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.

Economic Impact of Yield and Composition Variation in Bioenergy Crops: *Populus trichocarpa*

Renee M. Happs[1], Andrew W. Bartling[2], Crissa Doeppke[1], Anne E. Harman-Ware[1], Robin Clark[3], **Erin G. Webb**[3] (mailto:webbeg@ornl.gov), Mary Biddy[2], Jin-Gui Chen[4,5], Mark F. Davis[1], Wellington Muchero[4,5], Brian H. Davison[4,5], and **Gerald A. Tuskan**[4,5]

[1]Biosciences Center and Center for Bioenergy Innovation, National Renewable Energy Laboratory; [2]National Bioenergy Center and Center for Bioenergy Innovation, National Renewable Energy Laboratory; [3]Environmental Sciences Division, Oak Ridge National Laboratory; [4]Biosciences Division, Oak Ridge National Laboratory, [5]Center for Bioenergy Innovation, Oak Ridge National Laboratory.

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.

To achieve a bio-based economy, it is necessary to consider variability within a feedstock population. We must understand the range of key phenotypic characteristics in order to select economically advantageous genotypes for domestication in an optimized supply chain. In this analysis, we measured cell wall composition traits in a large natural variant population of *Populus trichocarpa*. The compositional analysis results were combined with agronomic growth data from the matching genotype to conduct a series of techno-economic analyses utilizing Aspen Plus [1] based on Humbird [2] of a lignocellulisic biomass refinery. These analyses evaluated the impacts of physical and compositional variability and determining the ultimate phenotypic drivers for yield and economic metrics. Here we show that while ethanol yield per land area per year and minimum fuel selling price (MFSP) were most strongly impacted by tree size, when considering the largest 25% of trees, size and carbohydrate content were nearly identical influencers on MFSP, highlighting the need to focus on both size and carbohydrate content in selecting economically optimal feedstocks.

Feedstock cost is one of the greatest operational expenses in a biorefinery, so understanding which feedstocks are advantageous relative to one another in terms of both cost and quality should be an important consideration for any biorefinery [2,3]. No formally accepted standard of lignocellulosic biomass quality has been previously defined, but several metrics may be used. Process yield, defined as amount of fuel produced per unit of feed, is one metric that evaluates the impact of a feedstock on total production. However, growers are typically concerned with achieving the highest biomass yields for a given land area with little consideration given to process yield or quality. To reconcile the gap between grower and biorefinery, both biomass yield and process yield can be combined and normalized to define the metric of fuel yield per land area per year [5]. This can be an important consideration when evaluating the land footprint required to reach a production target, but in conjunction to process yield may also serve as a negotiating point between the biorefinery and grower, or grower and plant breeders, in determining

price. With respect to economics, minimum fuel selling price (MFSP) needed to achieve a net present value of zero for a plant is often used to summarize the entire biorefinery in terms of economic drivers such as capital expense, operating expense, and revenue, where feedstocks that lead to the lowest MFSP can be regarded as the best performers. While not a direct measure of feedstock quality, MFSP can serve as a bottom-line economic indicator of the value of quality and quantity and through multiple analyses, determine which feedstock characteristics are most impactful.

As interest in poplar as a lignocellulosic feedstock grows, so too will the need to understand its variability across a population. There is a need to comprehend this variation not only as the range of physical and compositional phenotypes, but also how these phenotypes impact overall feedstock quality and biorefinery economics. Determination of advantageous processing phenotypes can provide insight guiding plant breeders into focusing on attributes which are the most influential to improving yields, while simultaneously reducing production costs. In this paper, we present the larger economic impact a broad range of natural variation in plant growth and cell wall composition for *Populus trichocarpa* has on the bioenergy landscape as MFSP, process ethanol yield, and field ethanol yield.

References

1. NREL Aspen Plus Model (2011 Design Report) https://www.nrel.gov/extranet/biorefinery/aspen-models/downloads/dw1102a/dw1102a.bkp
2. Humbird et al. (2011) NREL/TP-5100-47764. doi: 10.2172/1013269
3. Dutta, A. et al. Process Design and Economics for Conversion of Lignocellulosic Biomass to Ethanol: Thermochemical Pathway by Indirect Gasification and Mixed Alcohol Synthesis. (National Renewable Energy Laboratory, Golden, CO (United States), 2011).
4. Dou, C., Gustafson, R. & Bura, R. Bridging the gap between feedstock growers and users: the study of a coppice poplar-based biorefinery. Biotechnology for Biofuels 11, 77, doi:10.1186/s13068-018-1079-y (2018).

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 CRISPR-Cas Systems for Genome Editing in *Pseudomonas putida* KT2440 and *Clostridium thermocellum*

Julie Walker[1,4], Jacob Fenster[1,4], Sean Stettner[1,4], Anthony Lanahan[2,4], Tianyong Zheng[2], Camilo Toruno[2], Emily Freed[1,4], Andrew Levitt[1], Audrey Watson[1], Jeff Cameron[1,4], Lee Lynd[2,4], Dan Olson[2,4], Carrie Eckert[1,3,4]* (carrie.eckert@nrel.gov), and **Gerald A. Tuskan**[4]

[1]University of Colorado, Boulder; [2]Dartmouth College, Hanover; [3]National Renewable Energy Laboratory, Golden, CO; [4]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, 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.

Efficient microbial conversion of lignocellulose to fuels and chemicals is key to an economically viable bioproduction platform, particularly if coupled with feedstocks designed for optimal microbial performance. At their core, current approaches for accelerated domestication of model organisms such as *E. coli* mirror the cycle of "Design, Build, Test," revealing underlying design principles that inspire creative solutions to modern engineering challenges. The programmability of CRISPR-Cas machinery enables library based high-throughput and multiplexed experiments that have allowed for rapid genotype-phenotype mapping (1), protein engineering (2), strain engineering (3), and gene discovery (4). Our goal is to develop efficient CRISPR-mediated genome editing systems to enable these and other cutting-edge genome editing technologies in CBI microbes and feedstocks.

Both CRISPR-Cas editing and interference using the *Streptomyces pyogenes* Cas9 system have been demonstrated in the bacterium *Pseudomonas putida* KT2440, a promising candidate for the industrial production of renewable chemicals from lignin. CRISPR-interference (CRISPRi) was optimized for *P. putida* by screening inducible promoter systems that express catalytically-dead spCas9. The arabinose inducible promoter system performed best and was used to image the repression of the essential division protein *ftsZ* in real time. In addition, we have quantified the dynamic range of repression by targeting a genomically-integrated fluorescent reporter as well as key metabolic genes that compete with pathways to target products to increase titers. To optimize CRISPR-Cas gene editing, the transformation protocol from Sun et al. 2018 (5) was improved to increase the colony forming units by 100-fold while keeping the editing efficiency at 100%. The minimum homology arm (HA) length requirements for gene deletion and introduction of single codon mutations were determined. We are utilizing this system to recapitulate mutants from adaptive laboratory evolution (ALE) experiments and generate target mutagenesis libraries to evaluate protein function. In addition, we are working with JGI to generate a genome wide guide RNA library (237,000 guides) to identify functional guides for gene editing and repression via CRISPRi to

expand this system for genome scale studies.

Although the *Streptomyces pyogenes* Cas9 system has been utilized across a number of microbial and eukaryotic platforms, unfortunately it is not active in the growth temperature range of *Clostridium thermocellum* (55-60ºC), a thermophilic bacterium capable of directly converting cellulose to sugars, bypassing the need for chemical processing of lignocellulosic feedstocks. To enable CRISPR editing, we adapted and validated both the Type I-B native CRISPR system in *Clostridium thermocellum,* as well as the Cas9 system from *Geobacillus stearothermophilus* (6,7). Multiple attempts for CRISPR/Cas homology directed genome repair were largely unsuccessful for both systems, most likely due to the low efficiency of homologous recombination in *C. thermocellum.* To overcome this limitation, recombineering machinery was isolated from the thermophilic organism *Acidithiobacillus caldus* and expressed in *C. thermocellum*. An increase in homologous directed repair was observed in *C. thermocellum* strains expressing recombineering machinery when compared to parental strains, enabling a more rapid and efficient genetic engineering system in *C. thermocellum*. We are currently working on the evaluation of promoter systems to enable inducible CRISPRi in *C. thermocellum.*

Ultimately, we aim to utilize these CRISPR/Cas systems for rapid, HTP methods for phenotype-to-genotype discovery in both *P. putida* and *C. thermocellum* such as: 1) rational protein engineering, 2) complete residue substitution libraries, 3) pathway optimization, and 4) discovery of new gene functions by genome-wide targeting strategies. These tools will expand and accelerate the canonical the "Design, Build, Test" cycle in support of CBI research needs.

References

1. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, et al. A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. Cell. 2016 Jun 2;165(6):1493–506.

2. Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat Biotechnol. 2017 Jan;35(1):48–55.

3. Tarasava K, Liu R, Garst A, Gill RT. Combinatorial pathway engineering using type I-E CRISPR interference. Biotechnol Bioeng. 2018 Jul;115(7):1878–83.

4. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science (80-). 2014 Jan 3;343(6166):84–7.

5. 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.

6. Harrington LB, Paez-Espino D, Staahl BT, Chen JS, Ma E, Kyrpides NC, et al. A thermostable Cas9 with increased lifetime in human plasma. Nat Commun. 2017 Nov 10;8(1):1424.

7. Walker JE, Lanahan AA, Zheng T, Toruno C, Lynd LR, Cameron JC, et al. Development of both type I-B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium Clostridium thermocellum. Metab Eng Commun. 2020 Jun;10:e00116.

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

High-throughput functional characterization of *Populus trichocarpa* UDP-glycosyltransferases

Stephanie Galanie[1]*, Anna Furches[1], David Kainer[1], Jessa Murphy[1], Piet Jones[1], Erica T. Prates[1], Nancy Engle[1], Wellington Muchero[1], Daniel A. Jacobson[1], Timothy J. Tschaplinski[1] and **Gerald A. Tuskan[1]**

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI addresses strategic barriers to the current bioeconomy in the areas of 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to 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 is a fast-growing woody crop currently used for paper, pulp, and plywood with potential as a domestic feedstock for bioenergy and bioproducts. In order to breed or engineer poplar genotypes that are high-yielding, robust, and can be converted into value-added chemicals, we need to understand poplar secondary metabolism and its genetic basis. Secondary metabolism is important in plant response to stressors and influences resource allocation to and away from the cell wall. For example, lignin shares phenylpropanoid precursors with poplar phenolic glycosides involved in defense against insect herbivores, pathogens, UV radiation, and drought.[1, 2] UDP-dependent glycosyltransferases (UGTs) are responsible for catalyzing sugar transfer to phenolics and other metabolites, which influences their localization within cells and tissues, physiological role, and metabolic fate. UGTs are a large multi-functional enzyme class, so determining which ones have roles in secondary metabolism and which substrates they act upon from sequence homology alone is not currently possible.

We used a *P. trichocarpa* metabolite-genome wide association study to identify UGT encoding genes with single nucleotide polymorphisms (SNPs) associated with increased and decreased levels of metabolites.[3-5] Next, we profiled the activities of the enzymes against a substrate panel of phenolics and precursors; we identified enzymes with activities unique from those previously observed in *A. thaliana*.[6] Several of these enzymes with unique activity profiles are undergoing *in vivo* untargeted validation. This work will advance understanding of poplar secondary metabolism and aid in developing a robust, high-yielding feedstock.

References

1. C. J. Tsai, S. A. Harding, T. J. Tschaplinski, R. L. Lindroth and Y. N. Yuan, *New Phytologist*, 2006, **172**, 47-62.
2. T. J. Tschaplinski, P. E. Abraham, S. S. Jawdy, L. E. Gunter, M. Z. Martin, N. L. Engle, X. Yang and G. A. Tuskan, *Annals of Botany*, 2019, **124**, 617-626.

3. J. Zhang, Y. Yang, K. Zheng, M. Xie, K. Feng, S. S. Jawdy, L. E. Gunter, P. Ranjan, V. R. Singan, N. Engle, E. Lindquist, K. Barry, J. Schmutz, N. Zhao, T. J. Tschaplinski, J. LeBoldus, G. A. Tuskan, J.-G. Chen and W. Muchero, *New Phytologist*, 2018, **220**, 502-516.
4. L. M. Evans, G. T. Slavov, E. Rodgers-Melnick, J. Martin, P. Ranjan, W. Muchero, A. M. Brunner, W. Schackwitz, L. Gunter, J.-G. Chen, G. A. Tuskan and S. P. DiFazio, *Nature Genetics*, 2014, **46**, 1089.
5. H. B. Chhetri, D. Macaya-Sanz, D. Kainer, A. K. Biswal, L. M. Evans, J.-G. Chen, C. Collins, K. Hunt, S. S. Mohanty, T. Rosenstiel, D. Ryno, K. Winkeler, X. Yang, D. Jacobson, D. Mohnen, W. Muchero, S. H. Strauss, T. J. Tschaplinski, G. A. Tuskan and S. P. DiFazio, *New Phytologist*, 2019, **223**, 293-309.
6. M. Yang, C. Fehl, K. V. Lees, E.-K. Lim, W. A. Offen, G. J. Davies, D. J. Bowles, M. G. Davidson, S. J. Roberts and B. G. Davis, *Nature Chemical Biology*, 2018, **14**, 1109-1117.

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.

AkiraProt: An Ensemble Workflow for Proteome-Wide Structural Analysis

Erica T. Prates[1,2*] (teixeiraprae@ornl.gov), Manesh Shah[1,2], B Kirtley Amos[2,3], Allison Werner[1,4] Gregg T. Beckham[1,4], Daniel Jacobson[1,2] and **Gerald A. Tuskan[1]**

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Computational Systems Biology Lab, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [3]Department of Horticulture, University of Kentucky, Lexington, KY; [4]National Bioenergy Center, National Renewable Energy Laboratory, Golden CO

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.

The design of plants and microbes for cost-effective biomass conversion involves understanding the complex genetic basis of targeted traits. Functional screening techniques and statistical methods for genotype-phenotype mapping are continuously evolving, but their results are often limited to data correlations without capturing the specific roles of genes. Knowledge of intermediate omics layers (e.g., transcriptomics, proteomics and metabolomics), in turn, can help elucidate the molecular pathways of the information propagation from genes to phenotypes. The structural characterization of proteins is of paramount importance as they are the functional manifestation of genes.

Here, we present AkiraProt, a flexible and robust pipeline for high-throughput protein structure prediction, and its use for integration of structural proteomics in our Systems Biology studies. The workflow finely parses protein sequences and determines the optimum combination of state-of-the-art methods to be employed for protein structure modeling. Case-by-case generated protocols are selected to provide both the highest possible local resolution and maximum information about the overall shape of each protein. AkiraProt has been effectively used for structural analysis of eukaryotic and prokaryotic proteomes across several studies conducted within different focus areas of the Center for Bioenergy Innovation. Here we show its application leveraging the supercomputers at the Oak Ridge Leadership Computing Facility to solve the structural proteome of *Pseudomonas putida* KT2440, a widely studied host for biomass lignin valorization. Detailed analysis of specific proteins of interest contributed, for example, to the functional characterization of *P. putida* transport proteins involved in the optimum intake and the tolerance towards lignin-derived compounds. Other current applications of models generated with AkiraProt include prediction of key sites in high-throughput targeted mutagenesis for trait selection and ligand screening for protein functional mapping.

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.

SNPeffect: Identifying Functional Roles for SNPs using Metabolic Networks

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; and [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee

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.

Genetic sources of phenotypic variation have been a focus of plant studies aimed at improving agricultural yield and understanding adaptive processes. Genome-wide association studies (GWAS) identify the genetic background behind a trait by examining associations between phenotypes and single-nucleotide polymorphisms (SNPs). Although such studies are common, biological interpretation of the results remains a challenge; especially due to the confounding nature of GWAS population structure and the systematic biases thus introduced.

Here, we propose a complementary analysis ('**SNPeffect**') that offers putative genotype-to-phenotype mechanistic interpretations by integrating biochemical knowledge encoded in metabolic models. SNPeffect was used to explain differential growth rate and metabolite accumulation in 914 *P. trichocarpa* accessions as the outcome of SNPs in enzyme-coding genes. To this end, we also constructed a genome-scale metabolic model for *Populus trichocarpa*, the first for a perennial woody tree and further extracted a leaf-specific subnetwork using transcriptomics data collected from 390 poplar trees. As expected, our results indicate that growth is a complex polygenic trait governed by carbon and energy partitioning. The predicted set of functional SNPs are associated with experimentally-characterized growth-determining genes and also suggest functions for putative genes. Functional SNPs were found in pathways such as amino-acid metabolism, nucleotide biosynthesis, and cellulose and lignin biosynthesis, in line with breeding strategies that target pathways governing carbon and energy partition. A follow-up optimization procedure 'SNP-rank' was used to rank order the functional SNP set to provide specific targets for functional characterization and genomic selection priors. 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.

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.

QTL mapping and candidate gene discovery of metabolomic signatures in switchgrass

Thomas H. Pendergast IV[1,2*](thpiv@uga.edu), Nancy Engle[1,3], Audrey Labbé[1,3], Madhavi Martin[1,3], Zamin Yang[1,3], Soyeon Choi[1,2], Stephanie Galanie[1,3], Tim Tschaplinski[1,3], Katrien M. Devos[1,2] and **Gerald A. Tuskan**[1,3]

[1]Center for Bioenergy Innovation; [2]University of Georgia, Athens, Georgia; [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee

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.

Switchgrass (*Panicum virgatum* L.) is an emerging bioenergy crop undergoing rapid domestication efforts, and thus there is significant interest in understanding its genomic and metabolic pathways regulating important agronomic traits. Metabolomics reveal the chemical footprint of organisms, providing insight into the pathways that govern physiology, stress response, digestibility and sustainability. Here, we combine high-throughput metabolic analysis with genetic and trait mapping to identify the genomic regions, and ultimately genes, that underlie variation in crucial traits.

We developed a F2 switchgrass population, constructed a high-density genetic map, and generated metabolic profiles for a subset of 152 switchgrass F2 progeny using gas chromatography-mass spectrometry (GCMS). Quantitative Trait Loci (QTL) analysis associated these metabolomic signatures with recombination patterns in our genetic map, with 134 significant QTL for 101 metabolomic compounds, including compounds involved in lignin and stress response pathways. In particular, we found 9 lignan compounds with overlapping peaks on Chromosome 2N, explaining up to 77% of metabolite concentration, and identified 4 monolignol diversification candidate genes. Additionally, 4-O-coumaroyl- and 4-O-caffeoyl quinic acids shared a significant peak on Chromosome 3N. A candidate gene underlying this QTL varies with ecotype, with upland genotypes having low expression and numerous mutations and deletions. *In vitro* profiling of its biochemical activity confirmed that quinic acid is the preferred substrate. Similar analyses, combining genomic analysis, RNASeq, sequencing, CRISPR-Cas9 knockouts, and gene substrate affinity analysis, are being carried out to identify candidate genes underlying metabolomic variation for other mQTL. Understanding the genomic regulation of diverse primary and secondary metabolites provide insight into switchgrass recalcitrance, herbivore and disease resistance, drought tolerance, and other emergent sustainability and biomass traits.

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

High-throughput Reductive Catalytic Fractionation for Lignin Characterization in the Genome Wide Association Study of Poplar

Michael Stone[1,4*](mailto:mlstone@mit.edu), David Brandner[2], William Mounfield[1,4], Renee Happs[2], Richard Giannone[3,4], Suresh Poudel[3,4], Robert Hettich[3,4], Daniel Jacobson[3,4], Wellington Muchero[3,4], Gerald Tuskan[3,4], Gregg Beckham[2,4], Yuriy Román-Leshkov[1,4] , and **Gerald A. Tuskan**[3,4]

[1]Massachusetts Institute of Technology, Cambridge, Massachusetts, [2]National Renewable Energy Laboratory, Golden, Colorado, [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee, [4]The Center for Bioenergy Innovation, Oak Ridge, Tennessee

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.

Reductive catalytic fractionation (RCF) has emerged as a promising technique to extract and depolymerize lignin, an oxygenated aromatic polymer making up 15-30% of all biomass and the largest natural source of aromatic hydrocarbons.[1] RCF is a 2-step process, extracting lignin from whole biomass with a polar-protic solvent, and then selectively cleaving C-O ether bonds using a hydrogen donor and a heterogeneous catalyst.[2] When operating at complete ether bond cleavage at the catalyst, the lignin yields and product distributions are determined by the distribution of C-O and C-C linkages between monolignols present in native lignin in the plant.[3] Thus, to improve overall lignin utilization we must optimize the lignin structure in the plant to decrease heterogeneity in the product distribution. Unfortunately, targeted genetic improvements are not yet possible as the causal relationships between genes and lignin structure are not well understood at present.

Genome wide association mapping studies (GWAS) are a powerful tool to leverage the species-wide variability contained in a collection of individual organisms to draw links between naturally occurring genetic mutations and phenotypes.[4] We have designed and tested a 24-well reactor plate along with work-up protocols capable of performing high-throughput RCF – which we will use to analyze over 1000 poplar trees (*Populus trichocarpa*) with unique genotypes. Using LC-MS/MS, NMR and GC-FID, we can obtain high-quality quantitative data on lignin structure at a fine chemical scale to elucidate 1) which genotypes lead to exceptional lignin extraction and depolymerization and 2) whether certain genotypes contain unique and advantageous lignin chemistry that has yet to be discovered. Applying the GWAS analysis technique, we can relate these unique traits to the specific genes that drive favorable lignin biosynthesis. This presentation will demonstrate the proof-of-concept experimental design for our high throughput system, described above, and the reliability of the data which can be generated and fed into the GWAS pipeline.

References

1. Schutyser, W.; Renders, T.; Van den Bosch, S.; Koelewijn, S. F.; Beckham, G. T.; Sels, B. F., Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading. *Chemical Society Reviews* **2018,** *47* (3), 852-908.

2. Anderson, E. M.; Stone, M. L.; Katahira, R.; Reed, M.; Beckham, G. T.; Román-Leshkov, Y., Flowthrough Reductive Catalytic Fractionation of Biomass. *Joule* 2017.
3. Anderson, E. M.; Stone, M. L.; Katahira, R.; Reed, M.; Muchero, W.; Ramirez, K. J.; Beckham, G. T.; Román-Leshkov, Y., Differences in S/G ratio in natural poplar variants do not predict catalytic depolymerization monomer yields. *Nat. Commun.* 2019**,** *10* (1), 2033.
4. 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* 2019**,** *56*, 250-257.

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.

Functional genomic and cross-species studies uncover novel regulators of phenylpropanoid biosynthesis

Meng Xie[1,2,3*], Jay Chen[1,2,*] (chenj@ornl.gov), Jin Zhang[1,2], Vasanth R. Singan[4], Melissa J. McGranahan[5], Peter R. LaFayette[5], Sara S. Jawdy[1,2], Nancy Engle[1,2], Crissa Doeppke[1,6], Timothy J. Tschaplinski[1,2], Mark F. Davis[1,6], Erika Lindquist[4], Kerrie Barry[4], Jeremy Schmutz[4,7], Wayne A. Parrott[5], Feng Chen[3], Wellington Muchero[1,2], and **Gerald A. Tuskan[1,2]**

[1]Center for Bioenergy Innovation, Oak Ridge, TN; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]University of Tennessee, Knoxville, TN; [4]Joint Genome Institute, Walnut Creek, CA; [5]University of Georgia, Athens, GA; [6]National Renewable Energy Laboratory, Golden, CO; [7]HudsonAlpha Institute for Biotechnology, Huntsville, AL

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.

The phenylpropanoid pathway is responsible for the synthesis of a wide variety of bioactive chemicals, including lignin, p-coumaric acid, and flavonoids. However, genetic mechanisms regulating the carbon flux towards desirable molecules remain largely unknown. In order to maximize valuable bioproducts yield in biomass feedstocks, we leveraged the genomic and genetic resources of *Populus trichocarpa* genome-wide association studies (GWAS) population to identify and characterize genes and specific alleles underlying the regulation of phenylpropanoid biosynthesis. Furthermore, we extended *Populus* discoveries to other bioenergy feedstocks through cross-species studies.

By taking a GWAS approach using ~1,000 *P. trichocarpa* natural variants, we have linked one *Populus* 5-enolpyruvylshikimate 3-phosphate (EPSP) synthase gene (named *PtrEPSP-TF*) to lignin biosynthesis. Subsequently, our Omics analyses of *Populus* transgenic lines overexpressing *PtrEPSP-TF* and biochemical characterization of PtrEPSP-TF protein revealed that PtrEPSP-TF carries a DNA-binding helix-turn-helix motif and functions as a transcriptional regulator of the phenylpropanoid pathway upstream of MYB46 (Xie et al., 2018). Meanwhile, our GWAS analysis led to the identification of high-impact single nucleotide polymorphisms (SNPs) in the *PtrEPSP-TF* gene. Biochemical and transactivation analyses demonstrated that one of these high-impact SNPs resulted in the substitution of 142nd amino acid and dramatically impairs the DNA-binding and transcriptional activity of PtrEPSP-TF (Xie et al., 2020). These discoveries provide molecular targets for genetic engineering and genome-editing to regulate phenylpropanoid biosynthesis to improve biomass

feedstocks characteristics. We continue to utilize our discoveries to identify new regulators of the phenylpropanoid pathway and to gain mechanistic understanding of their regulatory actions.

References

1. Xie, M., Muchero, W., Bryan, A., Yee, K., Guo, H., Zhang, J., Tschaplinski, T., Singan, V., Lindquist, E., Payyavula, R., Barros-Rios, J., Dixon, R., Engle, N., Sykes, R., Davis, M., Jawdy, S., Gunter, L., Thompson, O., DiFazio, S., Evans, L., Winkler, K., Collins, C., Schmutz, J., Guo, H., Kalluri, U., Rodriguez, M., Feng, K., Chen, J.G, Tuskan, G. (2018). A 5-enolpyruvylshikimate 3-phosphate synthase functions as a transcriptional repressor in *Populus*. *Plant Cell* 30: 1645-1660.
2. Xie, M., Zhang, J., Singan, V., McGranahan, M., LaFayette, P., Jawdy, S., Engle, N., Doeppke, C., Tschaplinski, T., Davis, M., Lindquist, E., Barry, K., Schmutz, J., Parrott, W., Chen, F., Tuskan, G., Chen, J.G, Muchero, W. (2020). Identification of functional single nucleotide polymorphism of *Populus trichocarpa PtrEPSP-TF* and determination of its transcriptional effect *Plant Direct* 4: e00178.

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.

Genetic engineering to produce C-lignin deposition in plant stems

Jaime Barros[1,2]* (Jaime.Barros-Rios@unt.edu), Chan Man Ha[1,2], Chunliu Zhuo[1,2], Xiaolan Rao[2], Xirong Xiao[1,2], Xiaoqiang Wang[2], Fang Chen[1,2], Richard A. Dixon[1,2], and **Gerald A. Tuskan**[1]

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; and [2] University of North Texas, Denton

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.

Catechyl (C) lignin, discovered by our team in 2012 in the seed coats of diverse non-crop species [1,2], holds promise as a high-value co-product in biorefining [3,4] due to its more linear polymer structure. A CBI goal is to understand the biosynthetic pathway to C-lignin and develop strategies to introduce the polymer into cell walls of CBI target species (poplar and switchgrass) through genetic engineering without compromising plant growth performance. Lignin biosynthesis switches from formation of classical guaiacyl (G) lignin to C-lignin during development of the seed coat of the plant *Cleome hassleriana*. We are using molecular genetics to better understand C-lignin biosynthesis in Cleome seeds [5] and construct this pathway in vascular tissues of bioenergy crops. Our results suggest that: i) the transcriptional repression of monolignol *O*-methyltransferases coupled with expression of specific forms of cinnamyl alcohol dehydrogenase and laccase underlie the switch to C-lignin, ii) young seedlings of *Medicago truncatula* naturally produce C-lignin and its level was doubled by co-disruption of the genes COMT and CCoAOMT, and iii) C-lignin accumulation is limited by both the provision of the caffeyl alcohol monomer and the initiation of its polymerization by specific laccases. Future work with transgene constructs and isotopic experiments with labeled caffeyl alcohol have been designed to test the role of specific laccases on C-lignin biosynthesis.

References

1. Chen et al. (2012) PNAS 109:1772-1777. doi: 10.1073/pnas.1120992109
2. Tobimatsu et al. (2013) Plant Cell 25:2587-2600. doi: 10.1105/tpc.113.113142
3. Ragauskas et al. (2013) Science 311:484-489. doi: 10.1126/science.1114736
4. Berstis et al. (2016). ACS Sustainable Chemistry & Engineering 31:5327-35. doi: 10.1021/acssuschemeng.6b00520
5. Zhuo et al. (2019). Plant Journal 99:506-520. doi: 10.1111/tpj.14340

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Genome shuffling and bacterial quantitative trail locus (QTL) mapping in *Pseudomonas putida*

Julie E. Chaves[1]* (chavesje@ornl.gov), Delyana P. Vasileva[1], Jared C. Streich[1], Leah H. Burdick[1], Dawn M. Klingeman[1], Daniel A. Jacobson[1], Joshua K. Michener[1], and **Gerald A. Tuskan[1]**

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

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.

An ideal bacterium for industrial bioconversions would rapidly assimilate varied potential feedstocks, withstand extreme environmental conditions, and efficiently generate a useful product. Each of these desirable traits is highly polygenic, with multiple alleles making quantitative contributions to the overall phenotype. The pangenome of a bacterial species is expected to have more beneficial alleles than would be present in any individual isolate; for example, two strains evolve independent mechanisms (and alleles) that increase inhibitor tolerance. Therefore, identifying and recombining alleles between strains can improve performance.

Pseudomonas putida is an attractive host organism for the conversion of lignocellulose to specialty biofuels due to its natural ability to tolerate and metabolize lignin-derived aromatic monomers. We are optimizing methods for genome shuffling of *Pseudomonas putida* strains displaying desirable traits. For selectable phenotypes of interest, such as tolerance to inhibitory chemicals found in depolymerized substrates (e.g., biomass hydrolyzates), we can directly select improved progeny and identify the causal genetic changes. Additionally, measuring genotypes and phenotypes across a panel of shuffled progeny allows the mapping of genetic determinants for non-growth-associated phenotypes such as product formation. These genetic loci can then be used as targets for rapid future rational engineering efforts.

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.

Improving *Clostridium thermocellum* Product Titers by Increasing the Thermodynamic Driving Force of its Glycolytic Pathway

Shuen Hon[1,2], Satyakam Dash[1,3], Tyler Jacobson[1,4], Jingxuan Cui[1,2], David M. Stevenson[1,4], Teun Kuil[5], Antonius van Maris[5], Daniel Amador-Noguez[1,4], Costas D. Maranas[1,3], Daniel G. Olson[1,2], Lee R. Lynd (Lee.R.Lynd@dartmouth.edu)[1,2], and **Gerald A. Tuskan**[1]

[1]Center for Bioenergy Innovation, Oak Ridge, Tennessee, USA; [2]Dartmouth College, Hanover, New Hampshire, USA; [3]Pennsylvania State University, University Park, USA; [4]University of Wisconsin-Madison, Madison, Wisconsin, USA; [5]KTH Stockholm, Stockholm, Sweden

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.

Clostridium thermocellum is a cellulolytic, thermophilic anaerobic bacterium and is a candidate organism for the consolidated bioprocessing of lignocellulosic biomass into biofuels. Metabolic engineering of *C. thermocellum* is necessary to improve yields and titers of biofuels products such as ethanol and butanol. With respect to ethanol, *C. thermocellum* has already been engineered to produce ethanol at yields upwards of 80% of theoretical maximum. However, maximum ethanol titers from different engineered strains of *C. thermocellum* tend to be limited to ~30 g/L. Our present challenge therefore is to engineer for higher ethanol titer.

Past metabolic engineering attempts in *C. thermocellum* have most focused on altering the metabolic pathways between pyruvate and ethanol. In this work, we first present recent experimental evidence and thermodynamic analyses of *C. thermocellum* metabolism, which show that *C. thermocellum*'s glycolytic pathway appears to have less thermodynamic driving force that of other ethanol-producing microorganisms, such as yeast, engineered *Escherichia coli* and *Thermoanaerobacterium saccharolyticum*, and *Zymomonas mobilis*. These findings lead us to the hypothesis that thermodynamic limitations, and not biophysical inhibition, are limiting further improvements to improving ethanol titer. In particular, these analyses have identified that the pyrophosphate-dependent 6-phosphofructokinase (PPi-PFK) and phosphoenolpyruvate to pyruvate reactions in *C. thermocellum* – represented by the pyruvate:phosphate dikinase and malate shunts – as high priority targets to be replaced with more thermodynamically favorable reactions, such as the ATP-dependent 6-phosphofructokinase (ATP-PFK) and pyruvate kinase (PYK) reactions, respectively

Acting upon the aforementioned conclusions, we also present here our metabolic engineering efforts to improve thermodynamic driving force of *C. thermocellum* glycolysis. A major hurdle that had to be overcome was that pyrophosphate (PPi) was both an essential cofactor in glycolysis, and also a potent inhibitor of biosynthetic reactions. Achieving the ATP-PFK and PYK replacements necessitated not only identifying and eliminating the non-biosynthetic source of PPi from *C. thermocellum*, but also required us to completely uncouple PPi turnover from sugar metabolism. We have been successful at removing pyrophosphate's role from *C. thermocellum* glycolysis and have demonstrated that the modifications do in fact lead to a less reversible glycolysis. This engineered strain of *C. thermocellum* is one significant step closer to possessing the metabolism of ethanol producers, while still retaining its cellulolytic capabilities.

References

1. Tian L, Papanek B, Olson DG, Rydzak T, Holwerda EK, Zheng T, Zhou J, Maloney M, Jiang N, Giannone R, Hettich R, Guss A, Lynd L. 2016. Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum*. Biotechnol Biofuels 9:116.
2. Hon S, Holwerda EK, Worthen RS, Maloney MI, Tian L, Cui J, Lin PP, Lynd LR, Olson DG. 2018. Expressing the *Thermoanaerobacterium saccharolyticum pforA* in engineered *Clostridium thermocellum* improves ethanol production. Biotechnol Biofuels 11:242.
3. Dash S, Olson DG, Joshua Chan SH, Amador-Noguez D, Lynd LR, Maranas CD. 2019. Thermodynamic analysis of the pathway for ethanol production from cellobiose in *Clostridium thermocellum*. Metab Eng 55:161–169.

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.

Enzymatic Synthesis of Xylan Microstructures

Peter Smith[1,2*] (smithm@uga.edu), William York[1,2], Maria Peña[1,2], Breeanna Urbanowicz[1,2], and **Gerald A. Tuskan**[1,3]

[1]Center for Bioenergy Innovation; [2]The Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia; [3]Oak Ridge National Laboratory, Oak Ridge, Tennessee

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.

Xylans are a family of hemicellulosic polysaccharides found within the cell walls of many plants and algal species and consist of a linear backbone of β 1,4 linked xylosyl residues that are often substituted with other glycosyl or acetyl substituents. Xylans are known to interact with cellulose and lignins in the plant cell wall and are a known contributor to biomass recalcitrance to enzymatic conversion for fuels and products.

Here we demonstrate the ability of a recombinantly expressed xylan synthase to produce xylan microcrystalline aggregates with a well-defined morphology *in vitro,* using UDP-xylose as a glycosyl donor and short xylan oligosaccharides as acceptors. We have investigated the conditions under which these aggregates form and the kinetics of these aggregation events with varied acceptor concentrations. Microcrystals produced from the enzymatic polymerization and aggregation of xylan chains appear to form hexagonal structures with an average diameter of 6-8 μM; these microcrystals exhibit birefringent properties when subjected to cross-polarized light. Furthermore, xylan microcrystal crystallinity, morphology, and properties are influenced by fine structural elements, such as the structure of sidechain branching on the xylan acceptor. Together, these results suggest a route to new biobased materials and help to inform our understanding of polysaccharide dynamics and interactions.

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.

Investigating the spatial organization of aromatic catabolism in *P. putida* KT2440

Allison Z. Werner[1,2*] (Allison.Werner@nrel.gov), Davinia Salvachúa[1,2], Isabel Pardo[1], Martyna Michalska[3], Brenna A. Black[2], Bryon S. Donohoe[1,2], Stefan J. Haugen[2], Rui Katahira[2], Sandra Notonier[1,2], Kelsey J. Ramirez[2], Antonella Amore[2], Samuel O. Purvine[4], Erika M. Zink[4], Paul E. Abraham[1,5], Richard J. Giannone[1,5], Suresh Poudel[1,5], Erica T. Prates[1,5], Daniel Jacobson[1,5], Elsayed Mohamed[6], Adam M. Feist[6,7], Philip Laible[3], Robert L. Hettich[1,5], Gregg T. Beckham[1,2], and **Gerald A. Tuskan**[1,5]

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; [2]National Renewable Energy Laboratory (NREL), Golden, CO, USA; [3]Argonne National Laboratory, Lemont, IL, USA; [4]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA; [5]Oak Ridge National Laboratory, Oak Ridge, TN, USA; [6]Technical University of Denmark, Lyngby, Denmark; [7]Department of Bioengineering, University of California, San Diego, La Jolla, USA

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.

The valorization of lignin—an abundant, recalcitrant, and heterogeneous polymer in plant cell walls—is critical to enable the bioeconomy. Biological funneling of lignin-derived aromatic compound mixtures to single value-added products via engineered *Pseudomonas putida* KT2440 has emerged as a means to overcome heterogeneity. However, the spatial organization of lignin catabolism remains unclear, yet has implications for metabolic engineering strategies as well as our collective knowledge of nutrient acquisition by soil bacteria. Here, we first explore spatiotemporal dynamics of the *P. putida* exoproteome during cultivation on lignin-rich media. We observe that many enzymes with known and putative roles in aromatic catabolism are selectively packaged into outer membrane vesicles (OMVs) from early to late stationary phase, corresponding to the shift from bioavailable carbon to oligomeric lignin as a carbon source. Functional assays demonstrate that enzymes contained in the OMVs are active and catabolize aromatic compounds, which supports OMV-mediated extracellular breakdown of lignin-derived aromatics as a strategy for nutrient acquisition by soil bacteria. Second, we present the discovery that two undescribed proteins in *P. putida* are important to the uptake of *p*-coumarate and ferulate, which are abundant hydroxycinnamic acids in plant cell walls. Together, these works improve our understanding of the spatial distribution of lignin-derived aromatic catabolism and seek to support improved efficiency of microbial lignin conversion.

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.

Leveraging Super High Optical Resolution Microscopy to Probe the Interaction Zone Between *Clostridium thermocellum* and Biomass

John Yarbrough[1,3*] (John.Yarbrough@nrel.gov), Dominik Stich[2], Daehwan Chung[1,3], Neal Hengge[1,3], Todd Vinzant[1,3], Shu Huang[1], Yining Zeng[1,3], Michael Himmel[1,3], Yannick Bomble[1,3], and **Gerald A. Tuskan**[3]

[1]National Renewable Energy Laboratory, Golden, CO, USA, [2]University of Colorado Denver - Anschutz Medical Campus, Aurora, CO, USA; [3]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN

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.

Clostridium thermocellum is one of the most efficient microorganisms for the deconstruction of biomass. To achieve this high level of cellulolytic activity, *C. thermocellum* uses large multienzyme complexes known as cellulosomes to breakdown polysaccharides found in plant cell walls. The attachment of *C. thermocellum* bacterial cells to the nearby substrate via the cellulosome has been hypothesized to be the reason for this high efficiency. The region lying between the cell and the substrate can show great variation and dynamics that is affected by the growth stage of cells and the substrate used for growth. Many aspects of plant cell wall deconstruction by cellulolytic bacteria that directly bind to solid substrates remain unknown and resolving this knowledge gap is crucial for consolidated bioprocessing (CBP) applications. It is imperative to obtain a better fundamental understanding of the interactions that exist between the cellulosomes, bacteria, and the substrate. To address this question, we are utilizing unlimited diffraction microscopy (super resolution microscopy) to probe the distribution of cellulosomes at the microbial substrate interface. Using this technique in conjunction with density-based spatial clustering of applications with noise (DBSCAN), initial results demonstrate an increase in concentration of cellulosomes at the interface between the bacterium and the biomass substrate suggesting this increased concentration is an anchoring point between the bacterium and the substrate allowing other cellulosomes to be shuttled onto the biomass substrate.

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Lignocellulose-Fermenting Microbiomes: A "Compass" for Biofuel Process Development.

Evert K. Holwerda [1,2]* (evert.k.holwerda@dartmouth.edu), Chris Ellis,[1] Payal Chirania,[1] Xiaoyu Liang,[2] Suresh Poudel,[1] Richard J. Giannone, [1] Robert L. Hettich,[1] Lee R. Lynd[1,2] and **Gerald A. Tuskan[1]**

[1]Center for Bioenergy Innovation; Oak Ridge National Laboratory, Oak Ridge, TN; [2]Dartmouth College, Hanover, NH

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.

The question **"What biocatalysts are most effective at cellulosic biomass deconstruction?"** is of considerable fundamental interest, foundational for developing processes for biological conversion of cellulosic biomass, and surprisingly unresolved given the decades of research devoted to the field. Recently, substantial differences in solubilization capability have been demonstrated[1,2] and *Clostridium thermocellum* has been shown to be several-fold more effective than a commercial fungal cellulase preparation at solubilizing several feedstocks under a broad range of conditions.[1,2,3]

Lignocellulose-fermenting enrichments (aka microbiomes) contain a diversity of cellulolytic microbes and cellulase genes.[4] While product formation in such microbiomes is difficult to control, their performance and composition provide a "compass" for developing biofuel production systems featuring defined cultures.

We report here a multi-level investigation of lignocellulose-fermenting microbiomes, featuring:

- A stable, semi-continuous, thermophilic, anaerobic, methanogenic microbiome was established and operated at a 10-day residence time and increasing solids loading. Carbohydrate solubilization was undiminished with increasing solids loading up to 150 g/L, providing a value proof of biological capability.
- A new framework was developed, validated and applied that provides unprecedented insights into the comparative reactivity of lignocellulose substrates and the comparative effectiveness of biocatalysts at mediating lignocellulose solubilization.
- Using this framework, it was demonstrated that both a lab strain of *Clostridium thermocellum* and a prominent cellulolytic isolate obtained from the microbiome have equal ability to access lignocellulose and do so over 2-fold faster than the complex microbiome. This contrasts sharply with conventional wisdom that microbiomes will solubilize complex biomass more rapidly and

completely due to an increased diversity of potential degradative mechanisms; this observation awaits explanation.

- For the first time a comprehensive microbial gene catalog with 1 million genes has been created to give a more complete picture of the functional potential of the microbiome in the digestion of switchgrass
- This study revealed that the functional activity of an anaerobic microbiome, both in terms of protein abundances as well as taxonomic activity, varies significantly during solubilization of the same lignocellulosic biomass across various substrate loadings.

References

1. Paye, J. M. D., Guseva, A., Hammer, S. K., Gjersing, E., Davis, M. F., Davison, B. H., Olstad, J., Donohoe, B. S., Nguyen, T. Y., Wyman, C. E., Pattathil, S., Hahn, M. G. & Lynd, L. R. (2016). Biological lignocellulose solubilization: comparative evaluation of biocatalysts and enhancement via cotreatment. Biotechnology for Biofuels, 9:8.
2. Holwerda, E.K, Worthen, R. S., Kothari, N., Lasky, R. C., Davison, B. H., Fu, C., Wang, Z-Y., Dixon, R. A., Biswal, A. K., Mohnen, D., Nelson, R., S,, Baxter, H. L., Mazarei, M., Stewart Jr., C. N., Muchero, W., Tuskan, G. A., Cai, C. M., Gjersing, E., Davis, M. F., Himmel M. E., Wyman, C. E., Gilna, P. & Lynd, L. R. (2019). Multiple levers for overcoming the recalcitrance of lignocellulosic biomass. Biotechnology for Biofuels, 12:15.
3. Lynd, L. R., Guss A. M., Himmel, M. E., Beri, D., Herring, C., Holwerda, E. K., Murphy, S. J., Olson, D. O., Paye, J., Rydzak, T., Shao, S., Tian, L. & Worthen, R. (2016) Advances in Consolidated Bioprocessing using *Clostridium thermocellum* and *Thermoanaerobacter saccharolyticum*. Industrial Biotechnology: Microorganisms, Chapter 10.
4. Liang, X., Whitham, J. M., Holwerda, E. K., Shao, X., Tian, L., Wu, Y-W., Lombard, V., Henrissat, B., Klingeman, D. M., Yang, Z. K., Podar, M., Richard, T. L., Elkins, J. G., Steven D. Brown, S. D. & Lynd, L. R. (2018). Development and characterization of stable anaerobic thermophilic methanogenic microbiomes fermenting switchgrass at decreasing residence times. Biotechnology for Biofuels, 11:243.

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Rapid domestication of poplar using genomic selection and machine learning

David Kainer[1*] (kainerm@ornl.gov), Jonathon Romero[1], Hari Chhetri[1,2], David Macaya-Sanz[1,2], Stephen DiFazio[1,2], Daniel Jacobson[1], and **Gerald A. Tuskan**[1]

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, TN; [2]Center for Bioenergy Innovation, West Virginia University, WV

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.

Populus trichocarpa is an undomesticated woody species with high potential as a bioenergy feedstock, however, its long breeding cycle slows our ability to develop improved varietals. To rapidly improve biomass yield per hectare we need to optimize gains per generation. Sustainable and uniform yield is a highly complex conglomerate of traits, so genetically modifying or selecting for a few key loci is unlikely to achieve our goals. Genomic Selection[1] however, models the genetic (breeding) value of genome-wide variants on target traits in a training population. Parents from a breeding population are then selected solely on their aggregate genetic value for those traits. Furthermore, the same model can be used to predict progeny phenotypes well before they reach maturity, allowing early culling and rapid cycle times based on the genotype.

To this end, we are implementing an annual genomic selection to select high performing parents based on a multi-trait genomic best linear unbiased prediction (GBLUP) index that captures the goals of improved biofuel feedstocks. As a compliment to GBLUP, the CBI computational biology group is applying explainable machine learning models in order to incorporate non-linear relationships between genetic variants (epistasis) and between genetics and environment (GxE) into predictive models. These predictions are being utilized and tested in a cross of poplar parentals. Additionally, we have developed a computational pipeline that generates virtual recombinant progeny from a given cross, thus allowing a machine-learning prediction of progeny performance as a result of prospective parental selections and crosses.

References:

1. Hayes, B. J., and M. E. Goddard. "Prediction of total genetic value using genome-wide dense marker maps." *Genetics* 157.4 (2001): 1819-1829.

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.

The role of beneficial microbes in stress management of *Populus*

Jessy Labbe[1,2]* (labbejj@ornl.gov), Mang Chang[2], Devanshi Khokhani[3], Jean-Michell Ané[3], Paul Abraham[1,2,4], and **Gerald A. Tuskan[1]**

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; and [2]Graduate School of Genome Science Technology, University of Tennessee, Knoxville, TN, [3] Department of Agronomy, University of Wisconsin–Madison, Madison, WI; [4] Chemical Sciences Division at Oak Ridge National Laboratory

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.

Populus is an important commercial crop and one of the leading candidates for bioenergy production; however, the commercial use of *Populus* is easily impacted by abiotic and biotic stresses such as drought, nitrogen limitation as well as major pathogens (e.g. rust fungi, *Melampsora*). In order to maximize sustainable *Populus* biomass production, we proposed to engineer microbial community to alleviate stresses and nutrient deficiency limiting growth and productivity using native microbial collection. We assembled a fungal-bacterial community that benefit each other and are symbionts of *Populus*. At field scale, we demonstrated these beneficial microbes are stable in soil and *Populus* roots (after two years) while they increase plant yield of 30%. We also observed that this microbial treatment significantly reduces the severity of drought and infection of rust pathogens (*Melampsora spp*.).

As a consequence of this result, we further investigated the effect of the beneficial microbes during drought and *Melampsora* infection in *Populus*. There is little known on the role of beneficial root microbes in the modulation of drought in poplar and in poplar immune system to decrease susceptibility to leaf pathogens. We performed greenhouse experiments under well-watered and drought conditions and identified how beneficial microbes enhanced plant tolerance under water stress. In parallel, we employ closed microcosm system with beneficial microbial inoculants and sterile *Populus* plant treated with or without the rust pathogen. Thus, coupling transcriptomic and metabolomic approaches, we aim to characterize the defense-related metabolites and genes primed in poplar by beneficial microbes and to identify specific-defense microbial metabolite produced or transferred to the plant partner. This pioneer work contributes to the understanding of the molecular mechanisms to alleviate impactful stresses in woody plants during beneficial associations to further exploit them and develop a sustainable bioenergy feedstock.

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.

Prospecting thiamine diphosphate-dependent carboligases and characterizing their promiscuity to create novel metabolic pathways from primary metabolites

Bradley W. Biggs[1]* (bradleybiggs2014@u.northwestern.edu), Tracey Dinh[1], Matthew T. Robey,[2] Catherine Majors[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 characterize a library (>100) of thiamine-diphosphate dependent carboligase enzymes against a diversity of α-ketoacid substrates to determine the reaction landscape of this family of enzymes using machine learning, to identify ideals candidate enzymes from this family for biosynthesis applications, and then to use this information to assembly favorable enzymatic pathways to target bioproducts.

Abstract. Recent studies have demonstrated that enzyme promiscuity, the ability of an enzyme to accept non-native substrates and perform non-native chemistries, is widespread in nature. This provides an opportunity for biological engineers to both leverage this capacity for valuable chemical transformations and to hone desired activities. One particularly interesting family of enzymes to this end is thiamine-diphosphate dependent carboligases, which condense two α-ketoacids (or aldehydes) to form new carbon-carbon bonds. Because of an abundance of α-ketoacids in the central metabolism of common metabolic engineering hosts like *Escherichia coli*, this allows for possible assembly of new and favorable biochemical pathways to targets of interest. One-step condensations could generate more efficient routes to desired targets and access to novel molecules, including chiral compounds. Our goal is to characterize a library (>100) of carboligases, map their reactivity on a diversity of α-ketoacid substrates using machine learning, and then utilize promising enzyme candidates for biosynthesis applications. Here, we demonstrate ability to characterize this class of enzymes against a library α-ketoacids on our way to mapping the catalytic promiscuity of this enzyme family.

This work is supported by DOE grant DE-SC0019339.

Using machine learning to model promiscuous activity of thiamine diphosphate-dependent carboligases and side reactions in the *E. coli* metabolome

Tracey Dinh[1]*(tracey.dinh@u.northwestern.edu), Bradley W. Biggs[1], Matthew T. Robey[2], Catherine Majors[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 develop a predictive machine learning model that will elucidate the catalytic landscape of thiamine-diphosphate dependent carboligase enzymes across the chemical space of α-ketoacid substrates. This model will be used to identify enzymes for biosynthetic applications and 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 predict the resulting metabolomic consequences. Specifically, we focus on the biological activity of a family of thiamine-diphosphate dependent enzymes capable of catalyzing the condensation of α-ketoacids and aldehydes. Molecular fingerprints allow us to characterize the chemical space of α-ketoacid substrates, from which we sample and screen for enzyme activity in high-throughput. Using an active learning approach, we plan to train a support vector machine learning model on this bioactivity data. Robust models for each carboligase enzyme will not only allow us to identify novel biosynthetic reactions but also apply predicted promiscuous activity to genome-scale models of host organisms such as *E. coli*. We plan to carry out flux balance analysis to characterize effects of side reactions and select enzymes with minimal cell burden.

This work is supported by DOE grant DE-SC0019339.

Deep Green: Structural and Functional Genomic Characterization of Conserved Unannotated Green Lineage Proteins

Jianlin Cheng[1], Eric Knoshaug[2], Vladimir Lunin[2], Ambarish Nag[2], Ru Zhang[3], and **James Umen**[3]*
(jumen@danforthcenter.org)
*presenting author.

[1]University of Missouri, Columbia, MO; [2]National Renewable Energy Laboratory, Golden, CO; [3]Donald Danforth Plant Science Center, St. Louis, MO.

Project Goals and Overview

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). These unknown conserved proteins are likely to play important roles in core plant biological processes. Through this project, we will establish a pipeline and associated databases for structural prediction and functional characterization of 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 can perform high throughput functional genomics screening. 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.

The Deep Green project is divided into five major objectives to characterize and annotate unknown plant proteins, which include **1)** Assembly and ongoing curation of the Deep Green candidate protein set; **2)** *in silico* structural predictions and network analyses to assign structures and predict function; **3)** Assembly and curation of reverse genetic resources in Chlamydomonas; **4)** functional genomics characterization and prioritization in Chlamydomonas; and **5)** structural validation of selected candidates and functional validation in two important reference plants, Arabidopsis and Setaria.

The rich new data resources produced under the Deep Green project will be curated in one or more public databases, including DOE-supported KBase. These data will help guide researchers in investigating the contribution of conserved unknown proteins to diverse aspects of 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.

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-SC0020400.

Infernet: Gene Function Inference By Leveraging Large, Organ-Specific Expression Datasets And Validation Of Non-Redundant Regulators

Kranthi Varala[1]* (kvarala@purdue.edu), Ying Li[1], and Karen Hudson[2]
[1]Purdue University, West Lafayette, IN; [2] USDA-ARS, West Lafayette, IN

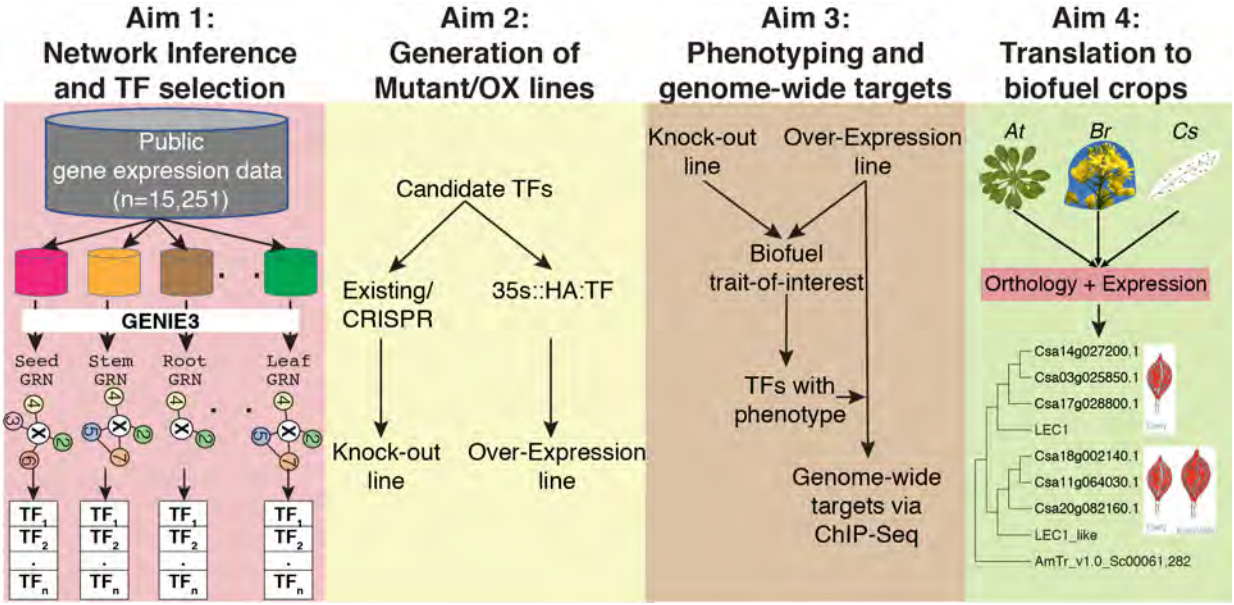
https://www.purdue.edu/hla/sites/varalalab/infernet

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).

Project Goals

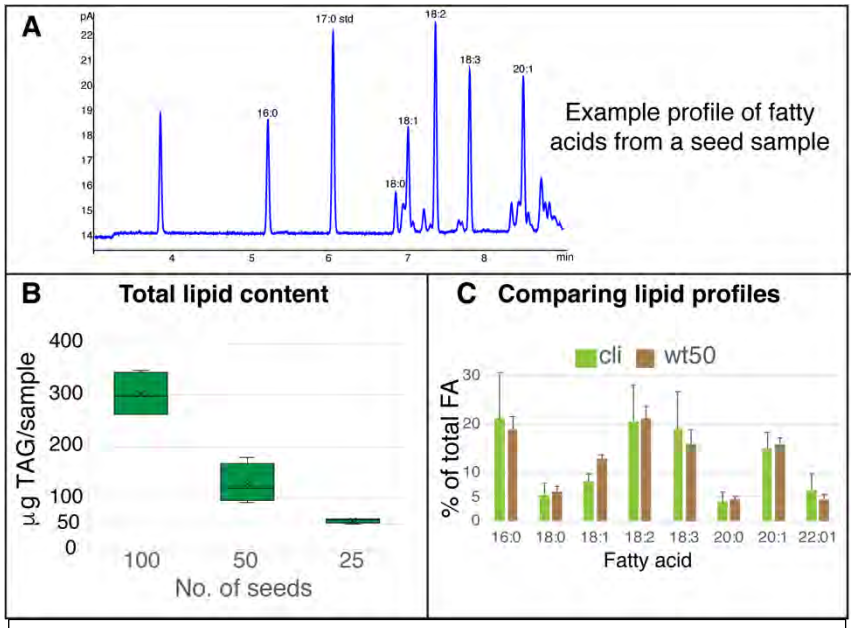
This DOE project focuses on the seed lipid biosynthesis trait as a proof-of-concept and aims to identify multiple novel regulators of this trait. For a given trait of interest, we predict novel TF regulators by leveraging the power of large expression data sets in a model species and then translate that knowledge to a target biofuel species for immediate benefit to the DOE mission. Two of the principal challenges in determining gene function are: 1. Subtle phenotypes that occur in



organ and/or developmental stage specific manner and 2. Functional redundancy among gene family members masks phenotypic effects via genetic or dosage compensation. The data and algorithms developed in this project should prove valuable to any research group interested in discovering novel TF regulators of any plant process. This project provides the data and tools to discover such regulators in the model plant Arabidopsis as well as a computational pipeline to translate the knowledge gained in Arabidopsis to a biofuel species.

Progress

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 novel TFs and are in the process of phenotyping them to detect changes in seed lipid profile. In addition efforts to generate over-expression lines for each of these candidate TFs is ongoing.



Lipid profiling provides quantitative and qualitative assays of seed oil content. **A**. Sample trace of fatty acid methyl esters from 100 seeds. **B**. Total triaclglycerides detected in Col-0 seed samples of different sizes shows that extraction of fatty acids is quantitative. **C**. Fatty acid composition profiles of wild type (wt50) and a line carrying a mutation in a histone modifier (cli).

In parallel, we have built node-vicinity-networks (NVNs) for each of these TFs from the full gene expression data set (n=15,251). These NVNs are being used to identify non-redundant TFs and prioritize those for phenotyping assays.

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.

Dynamic Genome-Scale Metabolic Network Modeling for a Novel Methanotroph-Cyanobacteria Coculture

Kiumars Badr[1*] (kzb0054@auburn.edu), Matthew Hilliard[1], William Whalen[1], Q. Peter He[1], Alexander S Beliaev[2], Marina Kalyuzhnaya[3] and **Jin Wang**[1]

[1]Auburn University, Auburn, AL; [2]Pacific Northwest National Laboratory, Richland, WA; [3]San Diego State University, San Diego, CA

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 interaction feedbacks 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 research objective is to help address those gaps through developing experimental/computational tools to characterize a synthetic photoautotroph-methanotroph 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 fuels and commodity chemicals. However, utilization of biogas represents a significant challenge due to its low pressure and presence of contaminants. In our previous research, we have clearly demonstrated that methanotroph-photoautotroph cocultures offer a flexible platform for efficient biological CH_4-CO_2 coutilization. However, development of multi-organism platforms for commercial biogas conversion present significant challenges which center around our ability to control function and composition of species in the coculture. It has been well recognized that an essential tool for the optimization, design and analysis of the coculture based biogas conversion is the development and validation of kinetics models that can accurately describe and predict the co-culture growth under different conditions [1]. To this end, using *Methylomicrobium buryatense* - *Arthrosipira platensis* as the model coculture system, we have developed an unstructured model to capture the growth dynamics. Specifically, Monod-like models were developed to capture coculture growth. Two sources of substrate were considered in the model: gas transferred from gas phase and gas produced *in situ*. In addition, we rely on the fitted maximum cell growth rate for both strains to capture other potential interactions. Using designed experiments and the developed model, we clearly demonstrated that the synergistic effect within the coculture cannot be fully explained by the *in situ* substrate exchange, and there must be other "metabolic links" to explain the significantly enhanced cell growth of both strains in the coculture.

Dynamic Genome-Scale Metabolic Modeling of the methanotroph-cyanobacteria coculture

In the era of omics big data, computational models are instrumental for turning different sources of data into valuable knowledge. Genome-scale metabolic models (GEMs) represent extensive knowledge bases that provide a platform for model simulations and integrative analysis of omics data. More importantly, it offers a convenient and powerful tool to test various hypothesis regarding "metabolic links" within the methanotroph-photoautotroph coculture. The GEMs,

especially the dynamic GEMs, offer a comprehensive picture of cellular metabolism and serve as a bridge that can better link the work of microbiologists and engineers in understanding and optimizing complex cellular metabolism. By integrating the available knowledge on each strain with data obtained in our own experiments, we use DFBAlab to implement the dynamic GEM for the model coculture. In the dynamic coculture GEM, besides the GEMs for each individual strain, the key inputs to the DFBA model are the uptake kinetics for different substrates. In this work, the substrate update rates are provided from the unstructured dynamic model we already developed.

Model validation was done through comparing model predicted product secretion rates and cell growth rates with experimental measurements over a course of batch coculture growth on biogas. It should be noted that obtaining accurate measurement of each species within the coculture is a highly challenging task. To address this challenges, we have developed an experimental-computational protocol for easy, fast and accurate quantitative characterization of the synthetic coculture [2]. Besides determining the individual biomass concentration of each organism in the coculture, the developed protocol also obtains the individual consumption and production rates of O2 and CO2 for the methanotroph and cyanobacteria. Besides these measurements, the overall dynamic trajectory over time, both measured low-frequency samples and unstructured model predicted high-frequency samples, offers significantly power in validating the model. Figure 1 compares the model predicted population ratio within the coculture for different inoculum ratios with measurements, with (a) from the unstructured kinetic model; (b) from the dynamic GEM.

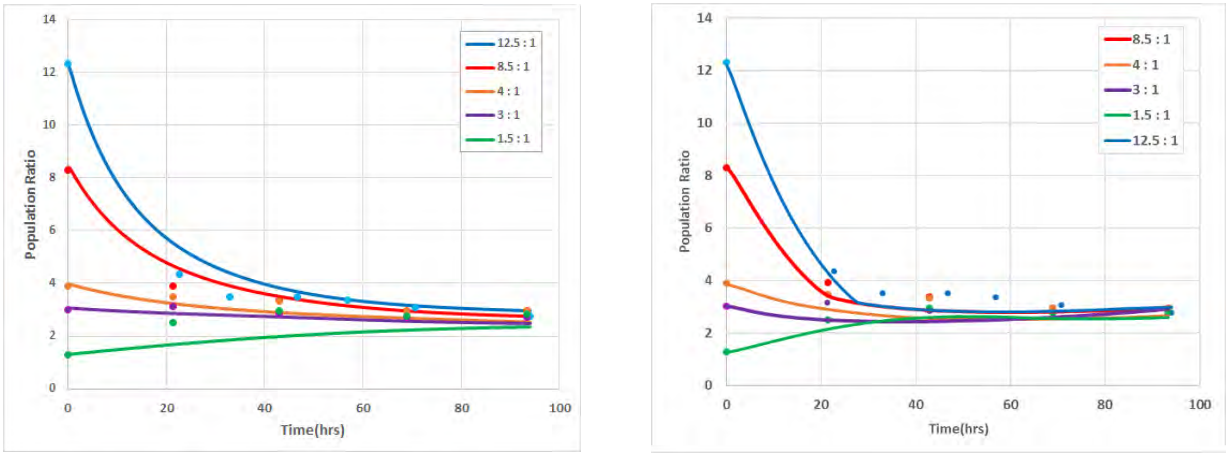


Fig. 1. Comparison of the population ratio within the coculture (cyanobacterium : methanotroph) predicted by the model (lines) with measurements (dots). A. (left) from unstructured kinetic model; B. (Right) from the structured dynamic GEM. Both models correctly predicted that despite different inoculum ratios, after reaching steady state, the population ratio of the two strains converge to the same value.

Publications:

1. 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;

2. Stone K., He Q.P., & **Wang J.** (2019), Two Experimental Protocols for Accurate Measurement of Gas Component Uptake and Production Rates in Bioconversion Processes, *Nature - Scientific Report*, 9 (1), 5899

Funding statement: This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019181; and by the U.S. Department of Education, Graduate Assistance in Areas of National Need Program under Award Number P200A150074.

Tuning C_1-metabolism for efficient utilization of biogas in synthetic photoautotroph-methanotroph binary consortium.

Marcus Bray[1] * (mbray@sdsu.edu), Kalyuzhnaya Marina G. [1], **Jin Wang[2]**

[1] San Diego State University, San Diego; [2] Auburn University, Auburn

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 interaction feedbacks 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 research objective is to help address those gaps through developing experimental/computational tools to characterize a synthetic photoautotroph-methanotroph 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: Paramount to the successful engineering of a synthetic photoautotroph-methanotroph consortia is a deep understanding of the physiology of native phototrophic and methane-converting bacteria. Such knowledge can then be used to achieve optimal growth and carbon conversion of robust co-cultures of engineered cells. *Methylomicrobium alcaliphilum 20Z^R*, is a methanotrophic bacterium able to utilize a range of C1 compounds whose physiology we are currently seeking to both catalog and optimize. This organism will then be used as a methanotrophic partner for construction of a robust synthetic photoautotroph-methanotroph binary consortium. Through a holistic study of *M. alcaliphilum* physiology, we aim to engineer a tunable system for construction of stable binary consortium for a multitude of applications in industrial biotechnology. Previous examples of this included the construction of a comprehensive metabolic framework of *M. alcaliphilum*[1], allowing us to better understand the movement of carbon and energy through the cell[1]. Two current areas of focus are electron/energy flow during methane oxidation and multistep gene expression from transcription through translation.

The first step of CH_4 utilization by *Methylomicrobium alcaliphilum* employs the membrane bound particulate methane monooxygenase (pMMO) which uses O_2 to oxidize CH_4 to CH_3OH for further processing. Also required by this enzyme is an electron pair, the source of which has been a point of contention among scientists. Our previous, metabolic modeling has led us to assert that electrons moving against their voltage gradient, or "uphill", from cytochrome C through the bc_1 complex are being supplied to pMMO for CH_4 oxidation. To test this, we have measured the electron transfer activity of *M. alcaliphilum* intracytoplasmic membranes (ICMs) by monitoring the oxidation of cytochrome c in the presence of ICMs. Our preliminary results show that in the presence of a proton gradient, cytochrome c is oxidized by ICMs, even when cytochrome c oxidase is completely inhibited by KCN (**Figure 1**). These results seem to suggest that a proton gradient can be used to drive electrons uphill in ICMs. We are currently performing further experiments to implicate the bc_1 complex in this observation.

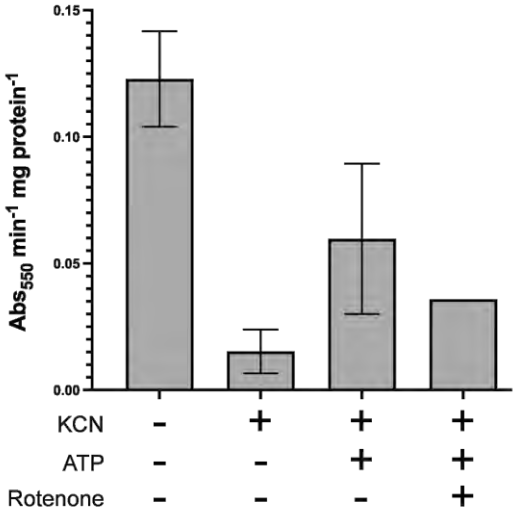


Figure 1. Oxidation of reduced cytochrome c in the presence of isolated *M. alcaliphilum* ICMs. Cytochrome c oxidation was monitored by measuring the decrease in Abs_{550} over time.

In addition to these studies we have set about expanding our ability to control the relative expression of genes within *M. alcaliphilum*. Many regulatory elements exist within bacteria at both the transcription and translational levels. We have currently been investigating the effect that different promoters, ribosome binding sites, and codon usage biases have on gene regulation in *M. alcaliphilum* by measuring the expression of reporter genes containing these elements. Preliminary results have identified representative regulatory elements which can be used in combination with one another to build a tunable gene expression repertoire in *M. alcaliphilum*. The combination of multiple elements will allow for not only turning genes of interest on and off, but fine tuning the expression of genes over a spectrum gene product outputs.

Our research activities will help us to increase carbon conversion efficiency in the methanotrophic partner as well as tune the expression of both endogenous and exogenous genes in relevant pathways. Using our engineered methanotroph in co-culture with optimize the efficiency of biogas conversion to value-added products.

References

1. Akberdin, I. R., Thompson, M., Hamilton, R., Desai, N., Alexander, D., Henard, C. A., … & Kalyuzhnaya, M. G. (2018). Methane utilization in *Methylomicrobium* alcaliphilum *20Z R*: a systems approach. *Scientific Reports*, *8*(1), 2512.

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.

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], Sara Jawdy[1], Dale Pelletier[1], Jon Shaw[2], Trent Northen[3], Adam Healey[3], Jeremy Schmutz[3], Chris Anderton[4], Dusan Veličković[4]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]Duke University, Durham NC, [3]Joint Genome Institute, Walnut Creek, CA; [4]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland WA

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 two overarching hypotheses, 1) 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), and 2) the surrounding microbiome can further adjust the tipping point through facilitation, competition, and antagonism.

To address the first hypothesis, we developed a synthetic community approach where strains of symbiotic cyanobacteria and moss genotypes from our sequenced pedigree population of 186 genotypes can be cultured together in symbiosis or separately and then cross-fed spent medium containing characterized exometabolites. Additionally, moss or cyanobacteria metabolite, metatranscriptomic and metaproteomic data were obtained. Using this approach, we were able to identify putative metabolites involved in nutrient and carbon transfer between community members. Specifically, we discovered the exchange of nitrogen- and sulfur-rich compounds between host and symbiotic cyanobacteria.

In an effort to link our findings from synthetic communities to the broader *Sphagnum* phylogeny, we sequenced 15 additional *Sphagnum* species and applied phylogenetic analyses. We discovered three putative whole genome duplications, two large chromosomal inversions, and

identified a potential hybridization event between hummock and hollow clades. Furthermore, we used gene family analysis to identify 15 genes under positive selection in hummock forming species including a sulfur transporter that was repressed in the cross-feeding study.

To address the second hypothesis, we developed a microbiome transfer approach where field collected *Sphagnum* microbiomes, conditioned to three-years of elevated temperature (ambient + 9 ˚C) or ambient temperature, were isolated and applied to germ-free tissue culture *Sphagnum* and exposed to temperature manipulations. Consistent over two consecutive years, we found that *Sphagnum* grows better at elevated temperatures when inoculated with a warming-conditioned microbiome than when inoculated with an ambient microbiome or no microbes at all. Metatranscriptome data revealed that changes in microbiome nitrogen, carbon, sulfur metabolism and heat shock response differed as a result of microbiome origin (i.e., conditioned to warming or ambient field temperature). On the plant side, expression of stress related genes, including those encoding heat shock genes, were reduced at elevated temperatures when in symbiosis with a microbiome originating from warm field conditions. Future experiments will apply these communities across sequenced moss pedigrees for identification of plant genes mediating beneficial microbial interactions.

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

Adam M. Guss[1] and **Janet Westpheling** (janwest@uga.edu)[2]
[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]University of Georgia, Athens, GA

The native ability of the Gram-positive bacterium, *Megasphaera elsdenii*, to efficiently generate C4 to C8 compounds via the condensation of acetyl-CoA makes it a compelling platform for the production of fuels and chemicals from lactate and plant carbohydrates. Our overall objective is to develop *M. elsdenii* for the conversion of lignocellulosic biomass sugars and organic acids into hexanol and other valuable chemicals. It produces organic acids including butyric (four carbon), hexanoic (six carbon), and in some cases octanoic (eight carbon) acids as major fermentation products when grown on lactate or glucose, likely via a chain elongation pathway using acetyl-CoA units. As carbon chain length increases, fuel properties improve, with the energy density increasing and hygroscopicity decreasing. We used DOE KBase to generate a metabolic reconstruction of the conversion of glucose and lactate to hexanoic acid in *M. elsdenii* and designed a strategy for engineering the strain. We developed the first methods for DNA transformation of two strains and have continued to develop genetic tools to enable more rapid and complex strain construction. Using a combination of heterologous genes for aldehyde dehydrogenase and butanol dehydrogenase in *M. elsdenii*, we have demonstrated production of butanol. We continue to engineer the strain to produce long chain carbon molecules at high yield and high titer and will present a progress report on those efforts.

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-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], Nancy Da Silva[2]

[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, *S. cerevisiae*, there is a clear lack of genome editing tools and standardized genetic parts for biosynthetic pathway construction. In this work, we expand the synthetic biology toolbox by identifying and characterizing a set of 25 different promoters and apply these new genetic parts to engineer the overproduction of a native metabolite, 2-phenylethanol, and a heterologous product, triacetate lactone (TAL). We first developed a one-step markerless multigene integration system that can effectively integrate three unique expression cassettes in a single round of strain engineering. We used this new technique to rapidly create a 27-member strain library that varied the expression of Shikimate pathway genes ARO4, ARO7, and PHA2. This refactoring experiment identified an engineered strain with a five-fold increase in 2-phenylethanol production and demonstrated new capabilities in the rapid engineering of *K. marxianus*. We have also developed a new high-efficiency CRISPR system for our toolbox. Using this system, new pathway knowledge, and computer predictions (using the OptKnock algorithm/*K. marxianus* genome-scale model), we can rapidly engineer metabolic pathways for increased synthesis of TAL from various carbon sources. Initial gene knockouts or heterologous gene integrations have resulted in up to four-fold increases in TAL production from xylose or glycerol. Taken together, the genetic engineering tools and metabolic engineering presented here demonstrate significant advancement in *K. marxianus* as a viable host of biochemical production.

This work was supported by DOE DE-SC0019093

Dissection of Carbon and Nitrogen Cycling in Post-Fire Soil Environments using a Genome-Informed Experimental Community

Thea Whitman,[1,*](twhitman@wisc.edu), Timothy D. Berry,[1] Nayela Zeba,[1] Monica Fischer,[2] Neem Patel,[2] Akiko Carver,[2] Sara Calhoun,[3] Andrei Steindorff,[3] Matthew Traxler,[2] Igor Grigoriev,[2,3] and Thomas D. Bruns[2]

[1]Department of Soil Science, University of Wisconsin-Madison, Madison, WI; [2]Department of Plant & Microbial Biology, University of California-Berkeley, Berkeley, CA; and [3]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

https://whitmanlab.soils.wisc.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.

Wildfires are a natural part of many forest ecosystems, with globally important carbon (C) storage and nutrient cycling consequences, and they are increasing in frequency and severity in western North America [1]. Forest fires affect soil C stocks in complex ways. Some C is released into the atmosphere through combustion, while a large percentage of the C is added to the soil in the form of pyrogenic (fire-affected) organic matter (PyOM). Worldwide, it is estimated that 16% of soil organic matter (SOM) is pyrogenic, while in areas with frequent fires, this number may be as high as 80% [2]. Understanding how wildfires affect SOM cycling requires understanding how microbes respond to PyOM and other post-fire soil conditions. However, our understanding of microbial interactions with post-fire soil is in its infancy. Outstanding questions include: Which microbes are capable of degrading PyOM? What are the relevant genes and metabolites associated with this degradation? How does the presence of nutrients such as nitrogen (N) affect post-fire PyOM and SOM degradation? What are the key interactions among post-fire microbes? We have developed resources during our previous DOE grant (DE-SC0016365) that enable us to dissect the post-fire environment. These include: a $^{13}CO_2$ labeling growth chamber, a "charcoalator" for producing PyOM, "pyrocosms" for simulating depth-resolved effects of fire, and a collection of genomic resources including 12 genome-sequenced pyrophilous fungi and more than 80 pyrophilous bacterial isolates. We are now using these in combination with genomics, transcriptomics, metabolomics, and stable isotope techniques to address the following specific objectives: 1) Determine how dominant post-fire soil microbes

affect the fate of PyOM; 2) Assess the interaction between N availability and PyOM mineralization by post-fire microbial communities and individual pyrophilous microbes; 3) Define the network of microbial interactions that facilitate PyOM breakdown over time and the key genes involved.

For Objective 1, we are using cultured isolates, defined post-fire substrates, genomic tools, and mass spectrometry to dissect the metabolic roles of these pyrophilous microbes. This objective ties in directly to Objective 2 (influence of N on PyOM degradation) and Objective 3 (defining the network of microbial interactions), as the realized metabolic activities of the dominant microbes determined in this part of the work will inform our models of these higher order processes. For Objective 2, we are targeting the following questions: (1) How available to microbes are the N-containing compounds that are retained by PyOM and soils during fire? (2) How does the addition of these N compounds affect SOM and PyOM decomposition rates and the microbes responsible for this decomposition? (3) What are the molecular mechanisms by which post-fire increases in mineral N affect PyOM degradation by microbes? We are approaching these questions first by adding N amendments to the series of experiments described in Objective 1 using our collection of isolated fungi and bacteria, and then using natural microbial communities with laboratory incubations. For Objective 3, our goal is to understand the fungal/bacterial interactions that underlie the observed patterns in post-fire nutrient cycling. At the conclusion of this goal, we will have (1) generated a map of interactions between the major fungal/bacterial taxa, (2) categorized the mechanisms underlying these interactions, and (3) leveraged genomics data to identify genes involved in successional transitions. This information will enable us to construct a conceptual model that accounts for the dynamics of this microbial community, and specifically how these interactions influence metabolism and resource utilization in post-fire soils. Together, our findings will be relevant for ecosystem management, post-fire recovery, and fundamental microbial genomics, ecology, and biogeochemistry.

References

1. Westerling, A.L., Hidalgo, H.G., Cayan, D.R., and Swetnam, T.W. 2006. Warming and earlier spring increase western US forest wildfire activity. Science 313, 940-943.
2. Lehmann, J., Skjemstad, J., Sohi, S., Carter, J., Barson, M., Falloon, P., Coleman, K., Woodbury, P., and Krull, E. 2008. Australian climate-carbon cycle feedback reduced by soil black carbon. Nature Geoscience 1, 832-835.

Funding Statement

This work was funded by the Department of Energy, Systems Biology Enabled Research on the Roles of Microbiomes in Nutrient Cycling Processes program, grant DE-SC0020351 to Thea Whitman, Thomas D. Bruns, Matthew Traxler, and Igor Grigoriev. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Integrating single-cell wetland microbiome structure, function, and activity to ecosystem-scale biogeochemical fluxes

S Bryson[1], L. Chistoserdova[1], D. Stahl[1], C. Pan[2], X. Mayali[3], and **M. Winkler**[1*] (mwinkler@uw.edu)

[1] University of Washington, Seattle, Washington, USA
[2] University of Oklahoma, Norman, Oklahoma. USA
[3] Lawrence Livermore National Laboratory, Livermore, California, USA

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

We propose to enhance understanding of the impact of climate change stressors on microbial wetland communities by reverse engineering natural wetland sediments, using *hydrogel* particles (concept A in Fig. 1) of varying sizes to trap sediment microbiota in stratified *bioreactor* setups (B in Fig. 1). These experimental systems are fully controlled and monitored analogs of the wetland soil column, and their microbial populations will be investigated with molecular tools such as *proteomic-SIP* (C in Fig. 1) and *NanoSIMS coupled with FISH* (D in Fig. 1). The results will be incorporated into mathematical wetland models (E in Fig. 1) to eventually run simulations that can predict how climate change stressors impact carbon and nitrogen fluxes across spatial and temporal scale. In the text below, we further explain concepts A-E which are also summarized visually in Figure 1. This proposal bridges different biological scales, linking single-cell scale measurements (with NanoSIMS and FISH) to population scale (with metagenomics and Proteomic SIP), to the ecosystem scale (with biogeochemical flux measurements) to inform global scale mathematical models.

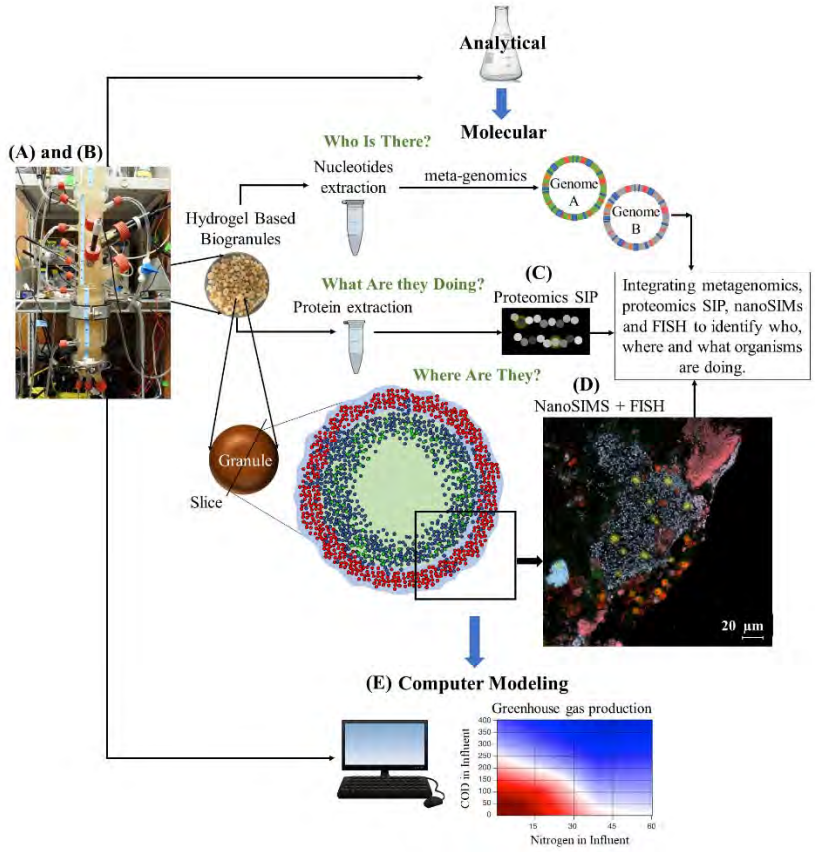


Figure 1: *Overview of the hydrogel based bioreactor setup (A, B), and analysis of the microbial populations with proteomic-SIP (C), NanoSIMS and FISH (D) and mathematical modeling (E).*

Abstract:
Wetland habitats serve essential functions in an ecosystem, including acting as water filters, providing flood and erosion control, and furnishing food and habitats for fish and wildlife. Wetlands cover 6% of the world's land surface but contain about 12% of the global carbon pool, playing an important role in the global carbon cycle (*1, 2*). Wetlands are responsible for roughly one third of global methane (CH_4) emissions. They help regulate atmospheric composition by absorbing and releasing greenhouse gases (GHG), such as CH_4 and

carbon dioxide (CO_2). However, these critical ecosystems are under increasing stress due to anthropogenic activity and climate change, and it is not clear how the accompanying environmental disturbances will impact the underlying microbial communities that drive these processes. The overall effects on microbially mediated functions that produce GHG emissions will likely unfold in distinct ways, depending on the type of wetland and the specific environmental impact. For this research proposal, we have identified two wetland habitats expected to be impacted differently by climate change.

- **Estuarine wetlands** are expected to be altered by changes in temperature, the intensity and duration of precipitation, and by sea level rise (*4*). The resulting intrusion of seawater into coastal wetlands will substantially alter the sediment chemistry, introducing abundant sulfate (SO_4^-) ions that may promote sulfate reduction over methanogenesis for the decomposition of organic matter and also facilitate the oxidation of methane (Fig. 2 left).
- **Lacustrine wetlands** are expected to be impacted by reduced snowpack, earlier snowmelt, and drought events, which will result in increased exposure of soil organic matter to oxygen. This may promote the aerobic microbial respiration of C stocks that have accumulated over centuries to millennia, with CO_2 as the primary product (*3*) (Fig. 2 right).

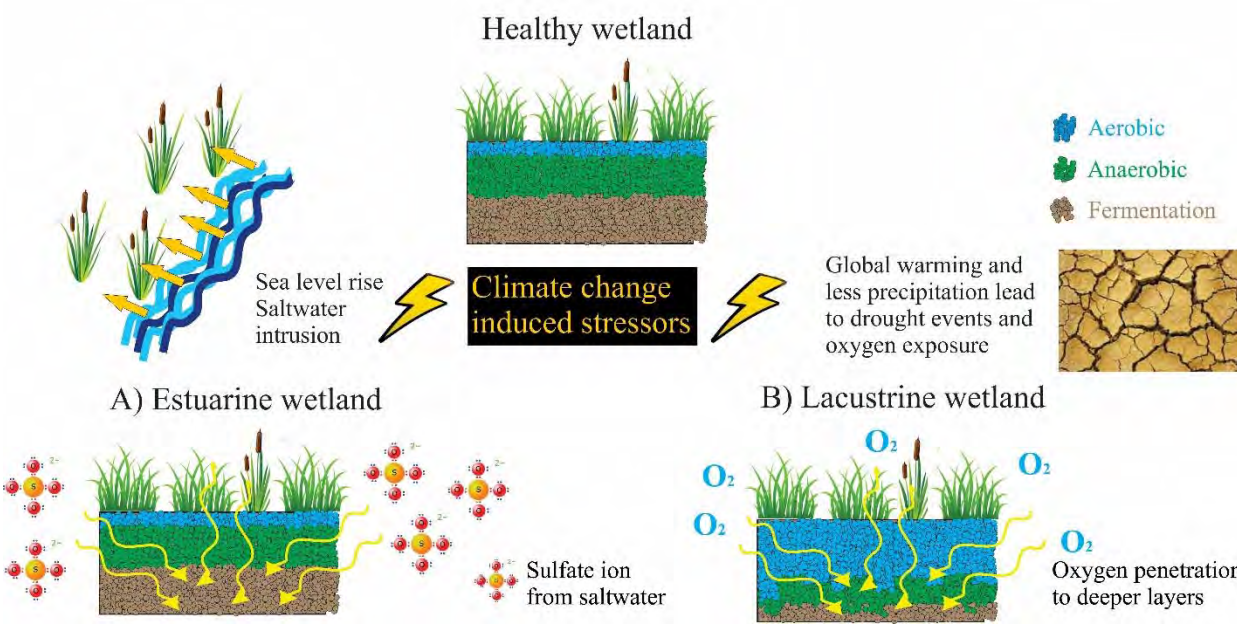


Figure 2: *Schematic overview of the scope of work. Environmental samples will be obtained from A) an estuarine and B) a lacustrine wetland to assess the impact of climate change stressors (sea-level rise for A and drought events for B) on changes in microbial community composition and the herewith associated C, N, P, S cycling including the release and capture of greenhouse gases (N_2O, NO, CO_2, and CH_4).*

This collaborative project takes advantage of expertise in four disciplines (biological process engineering, mathematical modeling, sediment microbiology, and omics driven science) to implement a game-changing technological approach to link the physiology of individual wetland microbial cells to larger mesoscale ecosystem processes. The **encapsulation of naturally resident microbiota in hydrogel particles** (called biogranules) of different sizes provides a format that closely replicates natural diffusion-limited processes, and the entrapment allows the system to experience extremes of imposed environmental conditions without community washout. Native communities entrapped in a novel biogranule format (i.e., biogranule column mesocosms) will be used to examine how anticipated climate-related changes in wetland habitats will impact biogeochemical activities in different redox zones at different scales of resolution, from the functional roles of individual cells, to interacting populations, and to system-level processes. Specifically, the proposed research will **quantify the impact of climate change stressors on model wetland microbial communities and the ecosystem processes** they sustain through the integration of mathematical modeling

and highly controlled hydrogel column experiments. This experimental design provides for both a system-level analysis (by continuous sampling of net fixation and release of gaseous microbial excretion products) and a coincident fine scale analysis of the supporting microbiology using **NanoSIMS paired with FISH, metagenomics, and Proteomic Stable Isotope Probing**. This research will use highly automated column mesocosms to mimic natural wetland conditions by closely simulating their chemical, thermal and physical gradients.

We will specifically investigate how predicted climate change-associated stressors impact the microbial populations of two wetland systems. ***Lacustrine wetlands*** are known to be impacted by lower rainfall and warmer temperature leading to more frequent draught events and therefore to lower above-ground productivity and higher below-ground oxygen exposure. In our experiments we will mimic this scenario by changing the availability of oxygen to test the impact on anaerobic respiratory processes, alterations of the microbial community structure, net metabolic rates, and transformations / partitioning of carbon and nitrogen including the stepwise conversion of intermediates (NO_2-NO-N_2O-N_2) in the denitrification pathway. ***Estuarine wetlands*** are expected to be impacted by saltwater intrusion due to sea level rise, which will shift the microbial community from methanogenesis towards sulfate reduction at a specific salt concentration (which we will experimentally alter). In both wetlands we will trace how carbon mineralization rates will vary as a function of the different redox zones, local microbial functional guild structures, and the forms of available organic carbon. We will infuse the hydrogel biogranule community with **a mixture of organic carbon polymers designed to approximate plant fiber composition** (cellulose, xylan, lignin, and pectin), providing for the controlled and stable microbial contact required for hydrolysis and metabolism of these complex carbon sources. Stable isotope labeling with ^{15}N-labeled ammonium in combination with NanoSIMS will be used to assess changing activities with depth in the individual hydrogel grains, and in relation to varying redox zones of the mesocosm system. Organism-specific activity profiles will be inferred by combined metagenomic and proteomic SIP analyses, and validated by more selective NanoSIMS analysis of carbon mineralization using ^{13}C-labeled intermediates of polymer degradation (coniferyl alcohol, glucose, acetate, CH_4, and CO_2) to resolve patterns of activity distribution and possible metabolic interactions among individual organisms and populations. This experimental format offers a unique way to **bridge different biological scales,** linking **single-cell scale** measurements (with NanoSIMS and FISH) to **population scale** (with metagenomics and Proteomic SIP), to the **ecosystem scale** (with biogeochemical flux measurements) to inform **global scale** mathematical models. The system response to imposed perturbations will be used for iterative modeling. Particle-level transformations will serve to both model and predict net C and N conversions, which should be consistent with measured net capture and release of CO_2/CH_4 from the column system and activities identified by proteomic SIP. These microscale models may contribute to more robust large-scale models and to project future climate-related changes in wetland processes.

Funding statement:
This work is funded by the Department of Energy (grant ##DE-SC0020356).

References:
1. T. R. Christensen *et al.*, *Geophysical Research Letters* **30**, (2003).
2. M. Cao, K. Gregson, S. Marshall, *Atmospheric Environment* **32**, 3293-3299 (1998).
3. C. Potter *et al.*, *Scientific Reports* **7**, 11314 (2017).
4. B. Blankespoor, S. Dasgupta, B. Laplante, *Ambio* **43**, 996-1005 (2014).

Coupling KBASE with PFLOTRAN

Roelof Versteeg[1*](roelof.versteeg@subsurfaceinsights.com), Rebecca Rubinstein[1], Chris Henry,[2] Kelly Wrighton,[3]

[1]Subsurface Insights, Hanover, NH, [2]Argonne National Laboratory, Chicago,IL,; [3]Colorado State University, Fort Collins, CO.

https://www.subsurfaceinsights.com/microbiological-with-pflotran

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

The goal of this project is to develop a high throughput, semi-automated data and analytical pipeline which will allow the use of the PFLOTRAN reactive transport modeling system in conjunction with KBase to model microbiological processes within an ecosystem.

Abstract

Watersheds are complex hydrobiogeochemical systems which provide multiple ecosystem services which are essential for US energy, food and water security. The manner in which watersheds provide these services is through a complex combination of the behavior of different watershed components (e.g. vegetation, soil, surface water) across a range of spatially nested subunits.

A predictive, actionable and science-based understanding of watershed behavior is critical for managing watershed resources and services. Achieving this understanding is extremely challenging due to various complex interactions within a watershed between plants, microorganisms, organic matter, soil, and water as well as the wide range of spatial and temporal scales across which these interactions take place. This challenge is exacerbated as these multi-scale interactions vary over time as a function of landscape position, elevation, human activity, environmental and biogeochemical gradients, and multiple feedback loops. While historically efforts to develop this understanding have focused on physical and chemical processes, there is a growing recognition that we need to develop a better understanding of hydrobiogeochemical processes. This will require the integration of physical, chemical and microbiological watershed data and coupling and integration of analytical tools and models across multiple domains, and should be done through a pipeline which has both data and analysis components. The data component should bring together the datasets required for such an understanding which are then used by the analysis component for the outcome of actionable information on hydrobiogeochemical processes.

Such a pipeline should be robust, semi-automated, high throughput and be useable by scientists from multiple domains.

Our approach in implementing this pipeline has been to develop cloud based software which provides a coupling between KBase (https://kbase.us/) and the open source reactive flow and transport modeling code PFLOTRAN (https://www.pflotran.org/). Our approach leverages and builds on the existing capabilities in these software packages, and specifically uses the Flux Balance Analysis capabilities in KBase to generate the input reactions which can be used by PFLOTRAN to simulate the geochemical processes driven by the microbiological species present in the subsurface.

Acknowledgement:

This effort was supported under DOE SBIR Phase I Award # DE-SC0019619

Innovations in Enzyme and Pathway Engineering for Cell-Free Production of Biofuels and High-Value Chemicals

Saken Sherkhanov* (saken@chem.ucla.edu), Hongjiang Liu, Tyler Korman, Scott McConnell, Kevin Cannon, Kyle Meador, **Robert Clubb[1,2]**, **Todd Yeates[1,2]**, **James Bowie[1,2]**

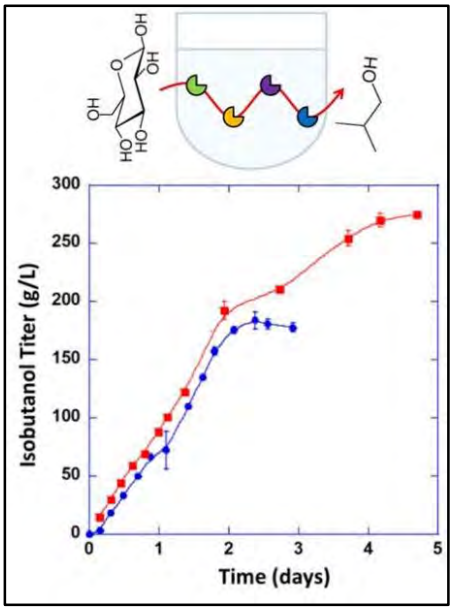
[1]UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA; [2]UCLA Department of Chemistry and Biochemistry, Los Angeles, CA

https://www.doe-mbi.ucla.edu/

Project Goals: Research in the **UCLA-DOE Institute for Genomics and Proteomics** includes major efforts in engineering novel metabolic pathways involving complex combinations of enzymes and enzymatic materials. Cell-free systems are being developed, comprised in some cases of more than 20 enzymes, to produce metabolites ranging from biofuels to high value chemicals. New innovations in cofactor recycling have enabled high turnover processes. Future directions aim to drive carbon-negative pathways using electrical/redox inputs, and to develop novel protein engineering approaches for stabilizing key enzymes in the form of designed materials.

Abstract:

Metabolic engineering enables the production of diverse biochemical materials from microbial platforms. However, living cellular systems operate within a limited range of physical conditions, which can constrain the feasible space for pathway engineering. In addition, the exceptional complexity of native metabolic networks can confound optimization and complicate attempts to direct metabolic flux exclusively to desired products in the cell. In vitro, cell-free, systems of enzymes present opportunities for metabolic design that have only been partially explored. We have pursued cell-free or 'synthetic biochemistry' approaches to develop diverse production platforms in the laboratory. Exploiting the ability to systematically test enzymes and enzyme combinations from divergent origins, and to readily control and optimize concentrations, we have succeeded in producing systems that show high yield, titer and productivity. In one system, we have developed a 16 enzyme pathway that produces the biofuel isobutanol at >90% yield, productivity of 4 g/L/hr and a titer of 275 g/L, production parameters that exceed highly optimized cell-based fermentation parameters. In another we have created a pathway to make cannabinoids at titers that exceed published microbial production by almost 3 orders of magnitude. A further feature of the in vitro approach is high modularity. With multiple pathways now in production, the ability to



modify and combine different modules is accelerating the production of diverse products, including in the broad terpenoid space.

The simplifying advantages of synthetic biochemistry approaches are countered by separate and unique challenges, which our project team has been addressing through innovative methods. In recent work we have developed systems to deal with difficult problems of cofactor balance, through judicious combinations of cofactor-recycling enzymes and by re-engineering cofactor specificities. Another key challenge, in the absence of de-novo protein synthesis in vitro, is enzyme stability. Cost-effective systems require high enzyme activity and longevity, sometimes in solution conditions outside the scope of evolutionary pressures. We are meeting this important challenge through novel enzyme engineering strategies. In addition to traditional approaches aimed at stabilizing individual enzymes through systematic mutation, we are developing new schemes for stabilizing enzymes, alone and in functionally important combinations, through various forms of conformational confinement. Confinement of enzymes has been demonstrated to provide generally stabilizing effects on proteins, especially in harsh environments, and sequestration or co-assembly of multiple sequentially-acting enzymes offers special opportunities for improving the flux through engineered pathways. Our team is pioneering new protein material engineering and chemical biology tools to further advance those studies.

The long term vision of the project is to conceive and create ready-made enzyme combinations and enzymatic materials for biofuel and high value chemical production. We also expect our engineering explorations to reveal new insights into natural and unnatural enzymes systems, including their scope and limits.

Publications/preprints

Opgenorth PH, Korman TP, Bowie JU, A molecular rheostat design that maintains ATP levels needed to drive cell-free synthetic biochemistry systems, *Nature Chem. Biol*., **13**(9):938-942 (2017).

Valliere MA, Korman TP, Woodall NB, Khitrov GA, Taylor RE, Baker D, Bowie JU, A cell-free platform for the prenylation of natural products and application to cannabinoid production, *Nature Comm.*, **10**(1):565 (2019).

McConnell SA, Cannon KA, Morgan C, McAllister R, Amer BR, Clubb RT, Yeates TO. Designed Protein Cages as Scaffolds for Building Multi-Enzyme Materials. *ACS Synth Biol*. (2020). doi: 10.1021/acssynbio.9b00407

This work was funded by the DOE BER Office of Science award # DE-FC02-02ER63421.

Microbial Metabolism, Chemistry, and Communities under Study at the UCLA-DOE Institute for Genomics and Proteomics

Rachel R. Ogorzalek Loo* (rloo@mednet.ucla.edu), John Muroski, Brendan Mahoney, Orlando Martinez, Janine Fu, **Joseph A. Loo, Robert Clubb, Robert Gunsalus,** and **Todd Yeates**

[1]UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA

https://www.doe-mbi.ucla.edu/

Project Goals: Research in the UCLA-DOE Institute for Genomics and Proteomics includes major efforts to elucidate critical microbial processes that decompose and recycle plant, animal and microbial biomass. Towards this end, we seek to decipher the metabolism of syntrophic microbial communities and examine how anaerobic microbes assemble complex cellulosome structures that degrade lignocellulose.

Biomass decomposition and recycling occur in essentially all anaerobic habitats on Earth, as well as in industrial/municipal waste treatment applications. Unfortunately, the current understanding of these critical processes is insufficient to enable modeling and prediction of environmental carbon flow. The benefits from increasing our knowledge of anaerobic decomposition include optimizing the attack and release of plant wall-derived molecules destined for biofuel and industrial feedstock production and improving biogas/sewage and waste stream processing plant design and operation.

Our DOE sponsored research seeks to advance the understanding of syntrophic-based microbial metabolism at molecular- and systems-levels and its role in biomass recycling/remediation. Exploring the pathways and key enzyme reactions of syntrophy begins by mining the genomes of previously unstudied syntrophic bacteria, such as those that metabolize model aliphatic fatty acid and amino acid substrates. That only minimal experimental data pertaining to these organisms is available severely limits the ability to draw conclusions from genome sequence alone, and even adding transcriptomic data may not suffice. For example, *Syntrophomonas wolfei* subspecies methylbutyratica possesses 9-11 paralogs for each reaction in the proposed beta-oxidation pathway along with distinct proteins for metabolizing branched chain fatty acids. Hence, quantitative proteomics is an important contributor, exposing which paralogue levels increase/decrease in response to substrate availability.

Unique to the UCLA efforts are the additional insights reaped by mining the proteomic mass spectrometry data, including how the pathways and key enzyme reactions of syntrophy may be modulated by post-translational modifications, an especially important consideration given that certain reactions possess equilibrium constants near 1. For example, fatty acid acylation of lysines in *Syntrophomonas wolfei* and *Syntrophus aciditrophicus* reflect substrate and metabolic intermediate levels.

In a second research project we are studying *Clostridium thermocellum,* which exhibits the highest level of cellulolytic activity of any microbe thus far characterized and is actively being developed for use in the consolidated bioprocessing of plant biomass into biofuels and chemicals. Its impressive ability to degrade lignocellulose is derived from the activity of a huge surface displayed enzyme complex called a cellulosome, which coordinates the binding of an array of cellulases. Only a few bacterial species within

the order Clostridiales are known to have evolved the capacity to display cellulosomes. Understanding how these microbes display these complex structures is of fundamental interest and could facilitate the construction of recombinant cellulolytic bacteria that have useful industrial applications. We are using an interdisciplinary approach to identify key components of the cellulosome biogenesis machinery. The aims project are to 1) Identify core components of the protein export machinery that produces the cellulosome. 2) Discover secretion stress and quality control systems involved in cellulosome biogenesis. 3) Elucidate the molecular basis of cellulosome tethering to the cell wall. Here we present our recent research progress, with particular focus on our studies RsgI-9, a novel member of a group of clostridial anti-sigma factors that regulate gene transcription of specific cellulosome components by sensing extracellular polysaccharide biomass components.

This work was funded by the Department of Energy BER Office of Science award # DE-FC02-02ER63421.

New Atomic Imaging Technology Development at the UCLA-DOE Institute

Jose A. Rodriguez [1]* (jrodriguez@mbi.ucla.edu), David Eisenberg,[1] and **Todd 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: Short statement of goals. (Limit to 1000 characters)

Research in the UCLA-DOE Institute for Genomics and Proteomics includes major efforts in the area of imaging technologies where we are advancing new methods that exploit electron microscopy. Our team is pioneering the new method of micro-electron diffraction (microED) using electron microscopes for diffraction. We are coupling this with scanning mode experiments that reveal microscale substructure within protein crystals at unprecedented levels of detail. The emerging microED technique presents challenges in phasing diffraction data that our team is tackling while pushing the boundaries of protein size and resolution. Our team has also made critical advances on one of the most outstanding problems in cryo-EM. Owing to low signal-to-noise, modern EM methods cannot resolve protein structures smaller than about 50 kDa in size. We have developed the first example of a working molecular scaffold that can image such small proteins.

References

1. Gallagher-Jones M., Ophus C., Bustillo KC., Boyer DR., Panova O., Glynn C., Zee C., Ciston J., Mancia KC., Minor AM., and Rodriguez, J.A. (2019) Nanoscale mosaicity revealed in peptide microcrystals by scanning electron nanodiffraction. Nat. Comms. Biol. 2, 26.
2. Hattne, J., Shi, D., Glynn, C., Zee, C., Gallagher-Jones, M., Martynowycz, M.W., Rodriguez, J.A. and Gonen, T. (2018) Analysis of global and site-specific radiation damage in cryo-EM. Structure. 26(5) 759-766.
3. Liu Y, Huynh DT, Yeates TO. (2019) A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. Nat. Communications 10, 1864. doi: 10.1038/s41467-019-09836-0.
4. Liu Y, Gonen S, Gonen S, Yeates TO. (2018) Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. PNAS 115: 3362-3367.
5. Thompson MC, Cascio D, Yeates TO. (2018) Microfocus diffraction from different regions of a protein crystal: structural variations and unit-cell polymorphism. Acta Crystallographica D Struct Biol 74: 411-421.

This work is supported by DOE Grant DE-FC02-02ER63421. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Transcriptomic analyses of bulk and single cell *Chalmydomonas* RNA-seq data reveal new gene functions and cell state heterogeneity

Patrice Salome, Sabeeha Merchant, Matteo Pellegrini

UCLA-DOE Institute for Genomics and Proteomics

Project Goals: Research in the UCLA-DOE Institute for Genomics and Proteomics includes major efforts in the elucidation of algal biology using genomics approaches.

The unicellular green alga *Chlamydomonas reinhardtii* is a choice reference system for the study of photosynthesis, lipid and starch metabolism and metal homeostasis. It is also a valuable model for the study of algal biofuel production. Despite decades of research, the function of thousands of genes remains largely unknown, and new approaches are needed to categorically assign genes to cellular pathways. Growing collections of transcriptome and proteome data now allow a systematic approach based on integrative co-expression analysis. We have used a dataset comprising 518 RNAseq samples derived from 58 independent experiments to identify potential co-expression relationships. While on a global scale random gene-pairs are not co-expressed, the determination of the co-expression profile of gene lists (manually curated from the literature) revealed high-confidence candidate genes with roles in cell cycle control, photosynthesis and respiration. Another striking observation was the clustering of nuclear genes largely as a function of their diurnal phase, even after the removal of all RNAseq samples collected over a diurnal cycle. A closer look at the remaining samples uncovered partial but frequent diurnal synchronization, although these samples had been collected from cultures maintained in constant light. We are now also exploring the potential heterogeneity of Chlamydomonas cultures grown in constant light by single-cell RNAseq: we have sequenced ~73,000 cells from two independent experiments representing two genotypes and three biological conditions, which we can identify clearly during visualization of this high-dimension dataset. We are now investigating the sources of heterogeneity within each culture by overlaying informative gene sets from our bulk RNA-seq data, to attempt to characterize diverse cellular states within large populations of cells. In the future, understanding the sources of variability could allow us, for example, to select states that are optimal for biofuel production.

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421

Metabolic engineering of cyanobacteria for enhanced production of ethylene and free fatty acids

Bo Wang[1*](bo.wang.2@vanderbilt.edu), Yao Xu[2], Josh Abraham[3], Dylan Courtney[3], **Brian Pfleger**[3], **Carl Johnson**[2], **Jamey Young**[1, 4]

[1] Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN. [2] Department of Biological Sciences, Vanderbilt University, Nashville, TN. [3] Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI. [4] Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN.

We are developing cyanobacteria as optimized cell factories for producing biofuels and other renewable biochemicals. One approach takes advantage of the global control of gene expression by the circadian "biological clock" to enhance photosynthetic production of ethylene and free fatty acids (FFAs). The second approach involves stabilizing expression of the ethylene forming enzyme (EFE) through removal of the inhibitory byproduct guanidine. The third approach involves metabolic engineering to circumvent native regulation of fatty acid biosynthesis.

In the first approach, we are investigating the ability of circadian clock reprogramming[1] to enhance product formation in three different species or strains of cyanobacteria: *Synechococcus* sp. PCC 7942 and PCC 7002, and *Synechocystis* sp. PCC 6803. We are using both PCC 6803 and PCC 7942 to produce the economically important chemical feedstock ethylene. Ethylene can be produced from CO_2 in PCC 6803 by overexpressing a single ethylene-forming enzyme (EFE) that derives from a plant pathogen.[2] Preliminary studies showed that ethylene production in this engineered *Synechocystis* is under regulation of the circadian clock, but attempts to manipulate the clock in PCC 6803 have been challenging due to the complexity and redundancy of circadian clock regulation in this species.

Ethylene can also be produced by expressing EFE in PCC 7942[3], which has a well characterized circadian clock, but the transgenic strains exhibited instability which is at least partially due to the toxicity of guanidine that is generated concomitantly with ethylene from the EFE reaction[4]. Therefore, we overexpressed a guanidine-degrading (GD) enzyme from PCC 6803 in PCC 7942, after which the cells showed much higher tolerance to exogenous guanidine as well as to the expression of EFE. Recently we have implemented two inducible systems to regulate the expression of EFE in PCC 7942, in order to further enhance strain stability. These stabilized strains of PCC 7942 will enable circadian clock reprogramming to be tested in a more predictable ethylene-producing host species.

Finally, we are developing strains of cyanobacteria that are optimized for production of FFAs, which are precursors to diesel fuels and other high-value oleochemicals including fatty alcohols, methyl ketones, and olefins[5]. We are engineering PCC 7002 to produce medium chain FFAs by overexpressing thioesterases that target the corresponding acyl-ACP intermediates made via fatty acid biosynthesis. Genes are integrated at the site of key acyl-ACP synthetases, thereby eliminating futile cycles. Early strains are producing titers comparable to the best presented in the literature, albeit in shorter production times given the faster growth of PCC 7002. Ongoing work is examining other rate limiting steps in fatty acid biosynthesis as well as product export and toxicity.

Unlike PCC 6803 and PCC 7942, it has not been reported whether a circadian system

exists in PCC 7002. To test if PCC 7002 shows circadian rhythmicity, we have also generated two kinds of luminescence reporters using *Vibrio harveyi* luciferase encoded by *luxA* and *luxB* (*luxAB*) genes. Both reporter constructs exhibited robust circadian rhythms in PCC 7002 under constant light conditions. Moreover, the circadian rhythms of PCC 7002 showed longer circadian periods and different phase angles at 30°C relative to those in PCC 7942. The PCC 7002 rhythms also displayed two other diagnostic circadian properties; they were well entrained by light/dark cycles and compensated over a broad range of temperatures. Therefore, we have discovered that a potent circadian clock exists in PCC 7002, and our recent data have demonstrated that reprogramming of the PCC7002 circadian clock can also promote expression levels of foreign reporter genes. In future studies we will attempt to manipulate it to further enhance production of FFA and other renewable biochemicals.

References

1. Xu, *et al*. Circadian yin-yang regulation and its manipulation to globally reprogram gene expression. *Curr Biol* 2013, 23:2365-74.
2. Wang, *et al*. A genetic toolbox for modulating the expression of heterologous genes in the cyanobacterium *Synechocystis* sp. PCC 6803. *ACS Synth Biol* 2018, 7:276-286.
3. Takahama, *et al*. Construction and analysis of a recombinant cyanobacterium expressing a chromosomally inserted gene for an ethylene-forming enzyme at the *psbAI* locus. *J Biosci Bioeng* 2003, 95:302-5.
4. Wang, *et al*. Photosynthetic production of nitrogen-rich compound guanidine. *Green Chem* 2019, 21:2928-2937.
5. Pfleger, *et al*. Gene construct encoding mutant thioesterase, mutant thioesterase encoded thereby, transformed host cell containing the gene construct, and method of using them to produce medium-chain fatty acids. US Patent: US20190390180A1.

This work was supported by Department of Energy award DE-SC0019404.

Rapid flux phenotyping to accelerate metabolic engineering of cyanobacteria

Piyoosh Babele[1]* (piyoosh.babele@vanderbilt.edu), Berkley Ellis[2*], Jody C. May[2], John A. McLean[2], **Jamey D. Young**[1,3]

[1]Chemical & Biomolecular Engineering; [2]Center for Innovative Technology, Department of Chemistry; [3]Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN USA

Project Goals: The objective of this project is to develop a systematic approach to identify bottlenecks within metabolic pathways that limit carbon flux towards enhanced lysine (LYS) and free fatty acid (FFA) production in genetically engineered cyanobacteria. We are developing methods based on desorption electrospray ionization-ion mobility-mass spectrometry imaging (DESI-IM-MSI) and stable isotope (^{13}C) based nonstationary metabolic flux analysis (INST-MFA) to rapidly quantify metabolite abundance and metabolic flux in bacterial cultures.

We have developed DESI-IM-MSI methods to rapidly sample and annotate biosynthetically produced molecules from microorganisms[1]. DESI-IM-MSI methods to sample microbial metabolites are of particular interest for screening purposes due to rapid sample acquisition and minimal sample preparation associated with ambient sampling. Using the imaging capabilities of this technology, we have associated the composition and amounts of biosynthetic products to different mutations, samples, or colonies within a MSI experiment. Our current work in the development of this DESI-IM-MSI platform has shown success in sampling FFAs from engineered *E. coli* grown on surfaces, LYS from modified cyanobacteria under solid phase growth, and various excreted metabolites within cellular media.

We also developed and successfully utilized INST-MFA to delineate the photoautotrophic metabolism of cyanobacteria[2,3,4]. This work describes our current efforts at refactoring previously engineered lysine-producing *Synechococcus* PCC 7002 strains[5] to identify flux bottlenecks that can be targeted to enhance strain performance. We compared the flux differences between high (TK.032) and low (AM.319) LYS producing strains with a wild-type (WT) control PCC7002 strain. Fluxes through enolase (ENO), pyruvate kinase (PK), and pyruvate dehydrogenase (PDH) were higher, while fluxes through malic enzyme (ME) and phosphoenolpyruvate carboxylase (PEPC) were lower in TK.032 as compared to AM.319 and WT strains. We hypothesize that by perturbing the expression of enzymes around these network nodes, we can remove potential flux bottlenecks and enhance LYS productivity.

Our studies demonstrate the promise of DESI-IM-MSI and INST-MFA to provide rapid analytical readouts on metabolic production and efficiency, which has broad implications in the field of synthetic biology. Further, the union of DESI-IM-MSI and INST-MFA presents a technology capable of rapid annotation of metabolism for optimizing cyanobacterial and algal strains for biotechnological/industrial applications.

References

1. Ellis, B.M. *et al.* Spatiochemically Profiling Microbial Interactions with Membrane Scaffolded Desorption Electrospray Ionization -Ion Mobility- Imaging Mass Spectrometry and Unsupervised Segmentation. *Anal. Chem.* ***91****, 13703-13711.*

2. Jazmin LJ, Xu Y, Cheah YE, Adebiyi AO, Johnson CH, Young JD. Isotopically nonstationary13C flux analysis of cyanobacterial isobutyraldehyde production. *Metab Eng*. 2017;42(October 2016):9-18. doi:10.1016/j.ymben.2017.05.001

3. Young JD. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics*. 2014;30(9):1333-1335. doi:10.1093/bioinformatics/btu015

4. Babele PK, Young JD. Applications of stable isotope-based metabolomics and fluxomics toward synthetic

biology of cyanobacteria. *Wiley Interdiscip Rev Syst Biol Med*. December 2019:e1472. doi:10.1002/wsbm.1472

5. Korosh TC, Markley AL, Clark RL, McGinley LL, McMahon KD, Pfleger BF. Engineering photosynthetic production of L-lysine. *Metab Eng*. 2017;44:273-283. doi:10.1016/j.ymben.2017.10.010

This work was supported by Department of Energy award DE-SC0019404.

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

Kira Tiedge[1]*(kjtiedge@ucdavis.edu), Andrew Muchlinski[1], Lisette Arce-Rodriguez[1], Dorothea Tholl[2], **Philipp Zerbe**[1]

[1]University of California, Davis; [2]Virginia Polytechnic Institute and State University, Blacksburg

Project Goals:

Of the myriad specialized metabolites that plants form to adapt to environmental challenges, terpenes form the largest group. In many major crops, unique terpene blends serve as key stress defenses that directly impact plant fitness and yield. In addition, select terpenes are used for biofuel manufacture. Thus, engineering of terpene 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 terpene blends. 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*) terpene metabolism. These insights would be combined with the development of genome editing tools to design plants with desirable terpene blends for improved biofuel production on marginal lands.

Abstract:

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 stress-related diterpene synthases (diTPSs) and cytochrome P450 monooxygeases (P450s) in the bioenergy crop switchgrass (*Panicum virgatum*). Mining of the allotetraploid switchgrass genome identified the largest thus far known diTPS family in plants, comprising 31 members. Biochemical analysis of 11 diTPSs revealed a modular metabolic network producing a diverse array of diterpenoid metabolites. In addition to *ent*-copalyl diphosphate (CPP) and *ent*-kaurene synthases likely involved in gibberellin biosynthesis, we identified *syn*-CPP and *ent*-labda-13-en-8-ol diphosphate (LPP) synthases as well as two diTPSs forming (+)-labda-8,13E-dienyl diphosphate (8,13-CPP) and *ent-neo-cis-trans*-clerodienyl diphosphate (CT-CLPP) scaffolds not previously observed in plants. Structure-guided protein mutagenesis of the (+)-8,13-CPP and *ent-neo*-CT-CLPP synthases revealed active site determinants that may resemble neo-functionalization events that occurred during diversification of the switchgrass diTPS family. In addition, we identified an unusual subfamily of P450s that are capable of directly converting the CPP and CLPP predicts into corresponding furanoditerpenoid structures, this bypassing the paradigmatic pathway organization of labdane diterpenoid metabolism.

Formation of several diTPS and P450 products, alongside the elicited expression of the corresponding biosynthetic genes, was induced in switchgrass roots and leaves in response to drought and UV irradiation, indicating their possible roles in abiotic stress adaptation. Together, these findings expand the known chemical space of diterpenoid metabolism in monocot crops toward systematically investigating and ultimately improving stress resilience traits in bioenergy crop species.

Funding statement.

This work is supported by the Department of Energy (DOE) Early Career Research Program (DE-SC0019178, to P.Z.), and a DOE Joint Genome Institute Community Science Program grant (CSP2568, to P.Z.).

Evidence for Metabolic Channeling of Glucose into the Oxidative Pentose Phosphate Pathway to Drive NADPH Production in *Rhodosporidium toruloides*

Tianxia Xiao[1*](tianxiax@princeton.edu), Yihui Shen[1], Carl Scultz[2], Mingfeng Cao[2], Hoang V. Dinh[3], John I. Hendry[3], Anshu Deewan[2], Sujit Jagtap[2], Costas Maranas[3], Christopher Rao[2], Huimin Zhao[2], and **Joshua Rabinowitz**[1]

[1]Princeton University, Princeton, New Jersey; [2]University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois; [3]The Pennsylvania State University, University Park, Pennsylvania

https://www.igb.illinois.edu/DOEcenter

Project Goal: Understand the metabolic pathways in R.toruloides.

Fatty acid synthesis enables production of biodiesel from renewable sources. However, efficiently supplying enough NADPH for fatty acid synthesis is challenging. How is this achieved in natural systems? To explore this, here we utilize stable isotope tracing in R.toruloides IFO0880, a yeast strain that engages in copious fatty acid and lipid production. Metabolic flux analysis (MFA) based on 1,2-^{13}C tracing reveals high metabolic flux through the oxidative pentose phosphate pathway (oxPPP) in both nitrogen-rich and nitrogen-scarce environments. ^{2}H-glucose tracing confirms substantial NADPH and fat labeling mediated by transfer of ^{2}H to NADPH at the G6PD step of the oxPPP. Strikingly, quantitative calculations show fat labeling *in excess* of that expected if all NADPH was made via the oxPPP in a well-mixed system. The data can be explained by channeling between hexokinase and G6PD to drive oxPPP flux. Implications and future research directions will be discussed.

This work is supported by U. S. Department of Energy Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018260.

Exploring Oleaginous Yeast *Rhodosporidium toruloides* as a Platform Organism for Production of Chemicals and Fuels

Anshu Deewan* (deewan2@illinois.edu), J Carl Schultz*, Sujit Sadashiv Jagtap, Huimin Zhao and **Christopher V Rao**

Carl R. Woese Institute for Genomic Biology, Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign

https://www.igb.illinois.edu/DOEcenter

Oleaginous yeast *Rhodosporidium toruloides* is viewed as a potential platform organism for production of biofuels and bioproducts such as biodiesel, lubricants, polyols, jet fuels, and alcohols from renewable lignocellulosic biomass[1,2]. *R. toruloides* naturally accumulates lipids from multiple sugars when some other essential nutrients such as nitrogen is limiting[3]. Recently, we have engineered *R. toruloides* for increased lipid production during growth on glucose[4,5]. In addition to lipid-based chemicals, *R. toruloides* also produces a number of sugar alcohols at high titers. For example, during growth on xylose in nitrogen-rich medium, *R. toruloides* produced D-arabitol[6]. However, present efforts to engineer the organism are hampered by the limited availability of genetic engineering tools. Targeted knock-in and knock-out are available using strains which have had the non-homologous end-joining (NHEJ) gene *KU70* removed, allowing homologous recombination (HR) to dominate, albeit at the expense of the cells' DNA repair capabilities and requiring tedious and low-throughput construction of long homology arms to target the modification.

As part of this project, we seek to both improve the fundamental understanding of *R. toruloides'* metabolism by elucidating the mechanisms of substrate utilization, metabolite identification, and identification of the key genes governing the lipogenesis process, as well as to develop more advanced genome editing tools to facilitate metabolic engineering efforts.

We performed transcriptome and metabolite analysis of *R. toruloides* IFO0880 during growth on plant-based sugars, acetic acid and lipids. The results from transcriptomics and metabolomics indicate global metabolic shifts resulting from growth on different substrates. We mapped differential gene expression and metabolite levels on the metabolic pathways for *R. toruloides*, which revealed the activation of different pathways by different substrates. We also identified and functionally characterized a few putative sugar transporters from *R. toruloides* in *Saccharomyces cerevisiae*. These results provide more clarity regarding substrate utilization in *R. toruloides* and their associated pathway.

We also report the first development of a functional CRISPR/Cas9 system in *R. toruloides* for modular, high-efficiency targeted gene knockout[7]. Different Cas9 and gRNA expression systems were evaluated, and reporter genes in the β-carotene biosynthetic pathway (phytoene synthase, phytoene desaturase) were knocked out with up to 98% deletion efficiency using a novel hybrid 5S rRNA-tRNA promoter for gRNA expression. Multiplexed deletion of phytoene synthase and β-isopropylmalate dehydrogenase (*LEU2*) was demonstrated with 78% efficiency.

References

1. Jagtap, S. S. & Rao, C. V. Microbial conversion of xylose into useful bioproducts. *Applied*

Microbiology and Biotechnology **102**, 9015–9036 (2018).

2. Isikgor, F. H. & Becer, C. R. Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polymer Chemistry* **6**, 4497–4559 (2015).

3. Ageitos, J. M., Vallejo, J. A., Veiga-Crespo, P. & Villa, T. G. Oily yeasts as oleaginous cell factories. *Applied Microbiology and Biotechnology* **90**, 1219–1227 (2011).

4. Zhang, S. *et al.* Engineering *Rhodosporidium toruloides* for increased lipid production. *Biotechnology and Bioengineering* **113**, 1056–1066 (2016).

5. Zhang, S., Ito, M., Skerker, J. M., Arkin, A. P. & Rao, C. V. Metabolic engineering of the oleaginous yeast *Rhodosporidium toruloides* IFO0880 for lipid overproduction during high-density fermentation. *Applied Microbiology and Biotechnology* **100**, 9393–9405 (2016).

6. Jagtap, S. S. & Rao, C. V. Production of D-arabitol from D-xylose by the oleaginous yeast *Rhodosporidium toruloides* IFO0880. *Applied microbiology and biotechnology* **102**, 143–151 (2018).

7. Schultz, J. C., Cao, M. & Zhao, H. Development of a CRISPR/Cas9 system for high efficiency multiplexed gene deletion in *Rhodosporidium toruloides*. *Biotechnology and Bioengineering* **116**, 2103–2109 (2019).

This work was supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award number DE-SC0018260.

Genome-scale Model Reconstruction and 13C-Metabolic Flux Analysis for Non-model Yeast Organisms *Rhodosporidium toruloides* IFO0880 and *Issatchenkia orientalis* SD108

Hoang V. Dinh[1*] (hvd5034@psu.edu), Patrick F. Suthers[1], John I. Hendry[1], Yihui Shen[2], Tianxia Xiao[2], Anshu Deewan[3], Sujit Jagtap[3], Zia Fatma[3], Joshua D. Rabinowitz[2], Christopher Rao[3], Costas D. Maranas[1], and **Huimin Zhao**[3]
[1]The Pennsylvania State University, University Park, Pennsylvania; [2]Princeton University, Princeton, New Jersey; [3]University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois

https://www.igb.illinois.edu/DOEcenter

Project Goals: Our project aims to develop new metabolic engineering, omics analysis, and computational modeling tools on a 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. To guide metabolic engineering, we aim to develop kinetic models accounting for reaction kinetics and allosteric regulations. Milestones achieved so far include reconstruction of comprehensive genome-scale metabolic models and development of large-scale carbon mapping models for ^{13}C-metabolic flux analysis used in kinetic parameterization.

Non-model yeasts are promising microbial cell factories due to their unique metabolic capabilities. *R. toruloides* is a basidiomycetes yeast that can accumulate large amount of lipids while *Issatchenkia orientalis* is a promising host for industrial production of organic acids thanks to its low-pH tolerance. To better assess these yeasts' metabolic capabilities, we reconstructed separate genome-scale metabolic models (GEMs) for each organism. Model reactions and genes were drawn from genome annotations. Biomass descriptions were derived from in-house-measured macromolecular composition and ATP maintenance requirements (calculated from chemostats data). We curated the model based on the available experimental data and ensure its quality with standardized tests (i.e., memote). Following the genome-scale models, we built carbon mapping models that are capable of explaining ^{13}C-labeling data and network cofactor balances. Flux distributions were predicted using the mapping model and labeling data (U-^{13}C-glucose and 1,2-^{13}C-glucose) capturing central carbon flux differences between the two yeasts. Energy metabolism involving reduced cofactors can also be elucidated.

This work is supported by U. S. Department of Energy Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018260.