U.S. DEPARTMENT OF ENERGY

Office of Science

2018 Genomic Sciences Program Annual PI Meeting

Abstract Book

February 25-28, 2018

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 26, 2018

Dear Colleague:

On behalf of the Biological Systems Science Division (BSSD) welcome to the 2018 Genomic Science Annual PI Meeting! I'd like to thank you for your efforts over the past twelve months in continuing to focus on your research and delivering the scientific results needed by this program, the Department and the Nation. Your work is vital and important and we thank you.

This past year has been marked by significant challenges but also remarkable progress. While there's no denying the considerable uncertainty with the Federal Budget process we have nonetheless continued to make necessary changes within the program to keep the research forward-looking and fresh. Over the past year we've initiated four new Bioenergy Research Centers, a new and expanded set of Biosystems Design projects, continued our efforts in Plant Feedstocks research and initiated new Microbiome projects at the DOE Labs. We are currently reconfiguring our computational biology and bioinformatics systems to a more open access format and developing new and/or improved Bioimaging capabilities for the portfolio. In short, we have continued to align and refresh the portfolio to help meet DOE's basic science goals for bioenergy and environmental research.

As in past meetings this annual event is an opportunity to view the entire program and to see where your project fits in with the larger portfolio. All funded principal investigator projects from our University portfolio and the DOE National Laboratory portfolio are present at this meeting. I would urge all researchers to take full advantage of the opportunity to meet your colleagues. You would be hard pressed to find a more focused scientific group in which to discuss your ideas for DOE research. This is also an opportunity to meet with your DOE program staff, representatives from elsewhere within the Department of Energy, and colleagues from other Federal Agencies.

Within BSSD we also strive to develop new and enabling research capabilities within the programs and at the DOE User Facilities. At this meeting there will be plenary and poster presentations of a variety of enabling capabilities supported within BER including the DOE Joint Genome Institute (JGI), the Environmental Molecular Science Laboratory (EMSL), the DOE Synchrotron Light and Neutron Sources and, the DOE Systems Biology Knowledgebase (KBase). Investigators from the Bioimaging research portfolio are also in attendance and will participate in the Bioimaging Research annual PI meeting immediately following this meeting. I would urge you to visit with representatives of these facilities/projects to learn more about new and/or upcoming capabilities available to you as a researcher.

We have planned a full agenda highlighting the exceptional research results produced within the program over the past year. We hope that these presentations will spark fruitful discussions that can be carried into the poster sessions, which are the heart of this meeting. The Bioenergy Research Center program has launched its 11th year with four new BRCs. We will hear from the center directors about their centers' research vision for the future as they begin the next five years of effort.

We are extremely pleased to welcome Dr. Katherine Yelick , the Associate Laboratory Director for Computing Sciences at Lawrence Berkeley National Laboratory. Dr. Yelick's research is in high performance computing (HPC) research where she has worked on interdisciplinary teams developing scientific applications ranging from simulations of chemistry and fusion processes to phylogenetics and genome assembly. She currently leads one of DOE's Exascale Computing Projects (ECP) entitled '*Exabiome: Exascale Solutions for Microbiome Analysis*,' for developing scalable methods for metagenomic sequence assembly and analysis on HPC systems. This exciting new capability could have profound implications for microbiome research and we are confident Dr. Yelick's presentation will be an excellent catalyst for scientific discussion throughout the meeting.

This year researchers supported by the joint USDA-DOE Plant Feedstock Genomics for Bioenergy program will also be in attendance. Recent efforts in this area of the portfolio are targeting disease resistance in bioenergy feedstocks and genetic improvements to oilseed crops. A breakout session featuring 2016 and

2017 awardees will be held to discuss some of the very important developments in feedstock genomics research that have taken place since these groups last met, and to highlight the most recent awards.

The KBase project will be hosting two afternoon sessions. The first will be a User Science session where current users will discuss science examples enabled by KBase capabilities. The second will be hands on experience for attendees to test the latest tools available in KBase in an interactive setting and discuss future capabilities with KBase staff. We also planned for a breakout session in Computational Biology that will feature new computational approaches for modelling and data analysis developed in various parts of the portfolio.

In 2017, BSSD issued a Funding Opportunity Announcement for the next iteration of the Biosystems Design program element. This year's meeting includes a breakout session featuring the twelve new Biosystems Design awards focused on the fundamental understanding needed to enable the design and engineering of plants and microbes for renewable production of biofuels and valuable chemicals.

We hope these breakout sessions are informative and help to foster new collaborations across the portfolio.

Finally, we are proud to once again host a plenary session featuring some of the most recent recipients of the Office of Science (SC) Early Career program awards. The Early Career program is one of the most competitive programs within SC and award recipients are part of an exclusive group. We are pleased to have four awardees present at this year's meeting.

Thank you again for making the program 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)

Jillian Banfield: Microbial community proteogenomic analyses indicate extensive depth-dependent CO oxidation and C1 metabolism in soil and increased capacity for N2O reduction with increased rainfall
PRESENTER: Spencer Diamond

Kristen DeAngelis: The "Who" and "How" of Microbial Control over Soil Carbon Dynamics: a Multi – omics, Stable Isotope Probing, and Modeling Approach
PRESENTER: Grace Pold

Bruce Hungate: Understanding Soil Metabolism from Carbon Release and Carbon Incorporation from Position-Specific 13C-Labeled Substrates
PRESENTER: Paul Dijkstra

Jennifer Pett-Ridge: Redox Fluctuations Control Coupled Iron-Carbon Cycling and Microbial Community Structure in Tropical Soils
PRESENTER: Amrita Bhattacharyya

Virginia Rich: Modeling the Pathways to Carbon Liberation: Microbiologically-Informed C-Cycle Modeling in a Thawing Permafrost Landscape, for Scaling from Molecules to the Earth System
PRESENTER: Eoin Brodie

Virginia Rich: Subarctic Lake Sediment Microbial Community Contributions to Methane Emission Patterns
PRESENTER: Joanne Emerson

Kelly Wrighton: Methanogenesis in Oxygenated Soils is a Substantial Fraction of Wetland Methane Emissions
PRESENTER: Kelly Wrighton

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

Trent Northen: The importance of biocrust organics revealed by metabolite sorption studies and in situ imagery
PRESENTER: Estelle Couradeau

John Dunbar: Microbial Communities As Carbon Conductors: Elucidating Universal Traits Across Litter Types
PRESENTER: Danielle Marias

Jizhong Zhou: Time-series metagenomics of experimentally warmed Alaskan tundra and Oklahoma temperate soils enables fine-resolution assessment of belowground C cycling feedbacks to climate change
PRESENTER: Kostas Konstantinidis

Louise Glass: m-CAFÉs: Microbial Community Analysis and Functional Evaluation in Soils
PRESENTER: Louise Glass

Paul Adams: Characterization of Microorganisms Resistant to Multiple Metals from the Contaminated Environment at the Oak Ridge Reservation
PRESENTER: Michael Thorgersen

Paul Adams: Coupling of Field- and Lab-based Experiments to resolve controls of Nitrate Respiration Pathway Partitioning at the Oak Ridge Shallow Aquifer
PRESENTER: Frederick von Netzer

Paul Adams: Ecological Stochasticity in Subsurface Microbial Community Assembly under Stress Gradient: Application of A General Quantitative Framework
PRESENTER: Daliang Ning

Paul Adams: ENIGMA Science Focus Area
PRESENTER: David Stahl

Paul Adams: Field-relevant Isolates from the Oak Ridge FRC and their Deep Characterization and Functional Analysis.
PRESENTER: Romy Chakraborty

Paul Adams: Isolates and their Deep Characterization
PRESENTER: Aindrila Mukhopadhyay

Paul Adams: Linking Activity to Phylogeny in Groundwater/Soil Ecosystems
PRESENTER: Heidi Smith

Paul Adams: Natural Organic Matter Dynamics and ExoMetabolomics for Microbial Cultivation from the Shallow Subsurface at the Oak Ridge FRC
PRESENTER: Matthew Fields

Paul Adams: Strain isolation, genome sequencing, and functional genomics reveals adaptive evolution of a Pseudomonas population at a human impacted field site
PRESENTER: Eric Alm

Eoin Brodie: A Trait Based Approach to Exploring the Impacts of Species Diversity and Functional Diversity on Algal Community Productivity
PRESENTER: Yiwei Cheng

Patrick Chain: Using Data Mining and High-Throughput Cultivation Methods to Better Understand Bacterial:Fungal Interactions
PRESENTER: Geoffrey House

Janet Jansson: Bridging the Soil Metagenome and Metaphenome Through Integrated Omics Analyses.
PRESENTER: Mary Lipton

Janet Jansson: Elucidation of Chemical Dark Matter in Soil Using 'Standards-free' Small Molecule Identification
PRESENTER: Ryan Renslow

Janet Jansson: Phenotypic Response of the Soil Microbiome to Environmental Perturbations
PRESENTER: Janet Jansson

Jennifer Pett-Ridge: Sequential Growth of Bacteria Following the Rewetting of a Seasonally Dried Grassland Soil
PRESENTER: Steve Blazewicz

Rhona Stuart: Multi-Component Characterizations of Algal Pond Ecosystems Reveal Algal Influence on Heterotroph Dynamics
PRESENTER: Christopher Ward

Matthew Sullivan: Host-Linked Soil Viral Ecology along a Permafrost Thaw Gradient
PRESENTER: Joanne Emerson

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

Jennifer Pett-Ridge: Understanding Microbial Stress Responses in Soil Using Metagenome and Metatranscriptome Analysis
PRESENTER: Paul Dijkstra

Maria Ghirardi: The Algal Ferredoxin Interactome
PRESENTER: Venkataramanan Subramanian

Ali Navid: In silico system-level analysis of interactions in algal-bacterial co-cultures
PRESENTER: Marc Griesemer

Janet Jansson: Modeling of Context-dependent Microbial Interactions in Biopolymer Degradation Networks
PRESENTER: Hyun-Seob Song

Philippe Noirot: CRISPRi-Mediated Analysis of Biofilm Formation in Plant Growth Promoting Pseudomonas fluorescens
PRESENTER: Philippe Noirot

Andrew Allen: A Systems Level Study of the Marine Diatom Phaeodactylum Reveals an Unexpected Mitochondrial Fatty Acid Beta Oxidation Pathway
PRESENTER: Graham Peers

Andrew Allen: Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in P. tricornutum
PRESENTER: Andrew Allen

Andrew Allen: Defining the Cellular Systems in Marine Diatoms That Provide for Their Predominant Acquisition of Dissolved Nitrate in World Oceans
PRESENTER: Andrew Allen

Andrew Allen: Illuminating Diatom Cell Biology with a Genetically Encodable Tag for Electron Microscopy and Subcellular Proteomics
PRESENTER: Andrew Allen

Andrew Allen: Nitrate sensing with a model pennate diatom, Phaeodactylum tricornutum
PRESENTER: Andrew Allen

Andrew Allen: THE METABOLIC LANDSCAPE CONNECTING CHLOROPLAST AND MITOCHONDRIA ENERGETICS IN DIATOMS
PRESENTER: Andrew Allen

Steven Allison: Fungal and Bacterial Communities Vary in their Carbon Cycling Response to Climate
PRESENTER: Sydney Glassman

Mary Lidstrom: Role of Lanthanides in the Complex Interplay of Alternative Enzyme Systems Involved in Methanotrophy
PRESENTER: Ludmila Chistoserdova

James McKinlay: Competition for a cross-fed nutrient between bacterial mutualists
PRESENTER: James McKinlay

Christof Meile: Microbial Interactions in The Anaerobic Oxidation of Methane: Model Simulations Constrained by Process Rates and Activity Patterns
PRESENTER: Xiaojia He

Sabeeha Merchant: A Day in the Life of Chlamydomonas
PRESENTER: Daniela Strenkert

Brian Pfleger: A Better Understanding of Bacterial RNA Turnover Developed Through RNA Sequencing
PRESENTER: Brian Pfleger

Christopher Rao: Quantification of Multiple Post-Translational Modifications by 'One-Pot' Affinity Enrichment - Applications in E. coli
PRESENTER: Birgit Schilling

Xavier Mayali: Nanoscale Stable Isotope Tracing and Compound-Specific Radioisotope Analysis to Investigate Metabolic Interactions between Bacteria and Biofuel-producing Algae
PRESENTER: Xavier Mayali

Christopher Rao: Regulation of acetate metabolism by pH in Escherichia coli
PRESENTER: James Orr

Paul Adams: Microbial Community Assembly and Evolution Analyzed at Single Cell Level
PRESENTER: Nitin Baliga

Jamey Young: Metabolic phenotyping of cyanobacteria with increased biofuel productivity
PRESENTER: Yi Ern Cheah

Janet Jansson: Visualizing Spatial Metabolic Interactions Within a Soil Microbiome
PRESENTER: Christopher Anderton

Paul Adams: In situ demonstration of sustained adaptation of a natural microbial community to transform substrates
PRESENTER: Terry Hazen

Bruce Hungate: Are the activities of microbial taxa consistent across ecosystems?
PRESENTER: Ember Morrissey

Bruce Hungate: Exploring Biome-scale Variation in Microbial Taxon-Specific Growth and Mortality at different temperatures through quantitative Stable Isotope Probing with H218O.
PRESENTER: Egbert Schwartz

Bruce Hungate: Field Measurements of Taxon-Specific Microbial Growth in Soil at Two Elevation Gradient Sites Using Quantitative Stable Isotope Probing (qSIP)
PRESENTER: Alicia Purcell

Louise Glass: m-CAFÉs -EcoFABs: Model Ecosystems to Advance Microbiome Science I
PRESENTER: Trent Northen

Louise Glass: m-CAFÉs -EcoFABs: Model Ecosystems to Advance Microbiome Science II
PRESENTER: Peter Andeer

Paul Adams: Integration of metagenomics and consortia data to study microbial interactions and community assembly
PRESENTER: Lauren Lui

Edward Eisenstein: Elucidating Mechanisms of Rust Pathogenesis for Engineering Resistance in Poplar
PRESENTER: Edward Eisenstein

Sarah Evans: Soil- and root-associated microbiomes across twelve switchgrass cultivars
PRESENTER: Tayler Chicoine

Mary Firestone: Deeply rooted: impacts of depth and soil type on root biomass, carbon turnover and microbial communities under switchgrass cultivation
PRESENTER: Jennifer Pett-Ridge

Mary Firestone: Exometabolomics of Switchgrass-Microbe Interactions in natural and fabricated ecosystems
PRESENTER: Kateryna Zhalnina

Mary Firestone: Identifying microbial traits driving microbial succession and C transformation in the rhizosphere of Avena sp
PRESENTER: Hee Jung Cho

Mary Firestone: Plant-microbe and microbe-microbe interactions mediate switchgrass sustainability: following rhizosphere microbial communities during switchgrass establishment
PRESENTER: Yuan Wang

Mary Firestone: Stable isotope-enabled metagenomics reveal phages active in rhizosphere soil
PRESENTER: Evan Starr

Mary Firestone: Unravelling Rhizosphere-microbial Interactions in the Rhizosphere of Alamo Switchgrass (Panicum virgatum) under Abiotic Stresses
PRESENTER: Na Ding

Kirsten Hofmockel: Molecular interactions of the plant-soil-microbe continuum of bioenergy ecosystems
PRESENTER: Kirsten Hofmockel

Thomas Juenger: Climate adaptation and sustainability in switchgrass: exploring plant-microbe-soil interactions across continental scale environmental gradients
PRESENTER: Thomas Juenger

Jared LeBoldus: Towards durable resistance to Septoria stem canker and leaf spot: a molecular understanding of resistance.
PRESENTER: Jared LeBoldus

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

James Moran: Spatially Resolved Rhizosphere Function for Elucidating Key Controls on Below-ground Nutrient Interactions
PRESENTER: James Moran

Wellington Muchero: Genomic divergence in pattern recognition receptors and its implications on endophytic microbial associations in the genus Salix
PRESENTER: Welligton Muchero

Kabir Peay: Climate Predicts the Dominant Form of Plant-Microbe Symbiosis in Forests at a Global Scale
PRESENTER: Kabir Peay

Kabir Peay: How Does Precipitation Impact the Taxonomic and Functional Diversity of the Populus trichocarpa Soil Microbiome?
PRESENTER: Michael Van Nuland

Daniel Schachtman: Using Metabolomics to Uncover Metabolic Pathways Utilized in Nitrogen Stress and Plant- Microbe Interactions in Energy Sorghum
PRESENTER: Amy Sheflin

Lawrence Smart: Dissecting Genetic Resistance to Willow Leaf Rust (Melampsora spp.) in Shrub Willow (Salix spp.)

PRESENTER: Lawrence Smart

Susannah Tringe: Diverse sorghum microbiome discovery and characterization in nitrogen- and waterlimited soils for improved biomass production

PRESENTER: Susannah Tringe

David Weston: Determining the genetic and environmental factors underlying mutualism within a plant-microbiome system: insights from genome sequencing, mass spectrometry imaging and exometabolite characterization

PRESENTER: David Weston

Jizhong Zhou: Successional Dynamics of Grassland Microbial Communities in Response to Warming, Precipitation Alternation, and Clipping

PRESENTER: Liyou Wu

Jay Chen: Plant-Microbe Interfaces: Genetic mapping and genomic resequencing identify a lectin receptor-like kinase as a regulator of Populus-Laccaria bicolor interaction

PRESENTER: Jay Chen

John Dunbar: Time-Dependent Tradeoff In Microbe-Plant Interactions Under Drought: Short-Term Stress Dampening Versus Long-Term Accelerated Mortality

PRESENTER: Danielle Marias

Robert Hettich: Plant-Microbe Interfaces: Developments in integrated omics to link microbial metabolism to community structure/function in plant/microbial systems

PRESENTER: Robert Hettich

Daniel Jacobson: Plant-Microbe Interfaces: AI-GWAPA, explainable AI-based approaches to genome wide phytobiome association

PRESENTER: Piet Jones

Daniel Jacobson: Plant-Microbe Interfaces: Phytobiome and transcriptional adaptation of Populus deltoides to acute progressive drought and cyclic drought

PRESENTER: Daniel Jacobson

Udaya Kalluri: Plant-Microbe Interfaces: Altered root metabolome composition impacts microbiome composition in Populus PdKOR1 RNAi plants

PRESENTER: Udaya Kalluri

Jessy Labbé: Plant - Microbe Interfaces: The key fungal lineage of the Russulaceae, a new resource for untangling and linking beneficial plant-fungal associations and ecosystem function

PRESENTER: Jessy Labbé

Jennifer Morrell-Falvey: Plant-Microbe Interfaces: Characterization of IAA biozynthesis pathways in Populus-associated microbes
PRESENTER: Jennifer Morrell-Falvey

Philippe Noirot: A Mycorrhizal Helper Bacteria Increases Mycorrhization in Aspen Seedlings by Modulating Expression of Defense Response Genes in Roots
PRESENTER: Peter E. Larsen

Dale Pelletier: Plant-MIcrobe Interfaces: Characterization of natural products from the Populus microbiome
PRESENTER: Dale Pelletier

Jennifer Pett-Ridge: How Plant-Microbial Interactions Shape Resource Allocation in Perennial Bioenergy Grasses
PRESENTER: Christine Hawkes

Scott Retterer: Plant-Microbe Interfaces: Emerging analytical techniques for controlling and monitoring architectural changes in developing multi-kingdom systems
PRESENTER: Scott Retterer

Christopher Schadt: Plant-Microbe Interfaces: The contribution of host chemotype as a driver of rhizosphere microbial community structure
PRESENTER: Christopher Schadt

Rytas Vilgalys: Plant-Microbe Interfaces: Dissecting the compatibility and diversity of the mycobiome of Populus trichocarpa
PRESENTER: Alejandro Rojas-Flechas

David Weston: Plant-Microbes Interfaces: Developing a synthetic community system to test preferential allocation to nitrogen-fixing bacteria
PRESENTER: David Weston

Ashley Shade: Microbiome diversity and assembly in the phyllosphere of perennial bioenergy crops
PRESENTER: Jackson Sorensen

Sarah Evans: Plant and Biogeochemical Controls on the Switchgrass Microbiome: Perspectives from a fine-scale time series
PRESENTER: Sarah Evans

Peggy Lemaux: From Leaves to Roots to Microbes: How Sorghum Responds to Drought
PRESENTER: Peggy Lemaux

Gerald Tuskan: Improving Nitrogen-Use Efficiency of Switchgrass Production: Past, Present, Future
PRESENTER: Michael Udvardi

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: Trait components of whole plant water use efficiency are defined by unique, environmentally responsive genetic signatures in the model C4 grass Setaria
PRESENTER: Ivan Baxter

Federica Brandizzi: Modulating hemicellulose to improve bioenergy crop
PRESENTER: Sang-Jin Kim

Evan DeLucia: Center for Advanced Bioenergy and Bioproducts Innovation - CABBI
PRESENTER: Evan DeLucia

Jenny Mortimer: Plant Sphingolipid Glycosylation And Its Role In Immunity And Cell Wall Organization
PRESENTER: Jenny Mortimer

Henrik Scheller: A dominant approach to reduce xylan in bioenergy crops
PRESENTER: Andrew Brandon

Gerald Tuskan: Highly Resolved Genome-Wide Association Mapping (GWAS) Reveals Novel Variants for Engineering Increased Biomass Yield in Populus trichocarpa
PRESENTER: Wellington Muchero

Gerald Tuskan: Metabolic Reconstruction, Analysis and Design of Metabolic Pathways
PRESENTER: Costas Maranas

Gerald Tuskan: Poplar GWAS and eQTL Analyses Reveal a Key Transferase Gene, HCT2, as a Missing Link in Plant Growth-Defense Tradeoffs
PRESENTER: Jay Chen

Gerald Tuskan: Switchgrass Field Experiments to Enable Rapid Domestication, Increased Sustainability, and a Genome-wide Association Study
PRESENTER: Reginald Millwood

Curtis Wilkerson: Increasing biomass in grasses by identifying genes involved in mixed-linkage glucan biosynthesis
PRESENTER: Mingzhu Fan

Peter Balint-Kurti: PHYSIOLOGICAL AND MOLECULAR-GENETIC CHARACTERIZATION OF BASAL RESISTANCE IN SORGHUM
PRESENTER: Peter Balint-Kurti

Thomas Brutnell: Leaf carbon isotope composition in Setaria: genetic contribution and potential for high throughput screening for water use efficiency in C4 plants
PRESENTER: Asaph Cousins

Thomas Brutnell: Phenomics of stomata and water use efficiency in C4 species
PRESENTER: Andrew Leakey

Thomas Brutnell: Using systems approaches to improve photosynthesis and water use efficiency in sorghum
PRESENTER: Thomas Brutnell

Thomas Brutnell: Using systems approaches to improve water use efficiency in sorghum by engineering root architecture
PRESENTER: Jose Dinneny

Luca Comai: Discovery and characterization of disease resistance loci in poplar
PRESENTER: Andrew Groover

Hugo Cuevas: Genomic dissection of anthracnose resistant response in sorghum [Sorghum bicolor (L.) Moench]
PRESENTER: Hugo Cuevas

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

Deanna Funnell-Harris: Resistance to Stalk Pathogens for Bioenergy Sorghum
PRESENTER: Deanna Funnell-Harris

Matias Kirst: Phylogenomic Discovery and Engineering of Nitrogen Fixation into the Bioenergy Woody Crop Poplar
PRESENTER: Matias Kirst

Chaofu Lu: Systems Biology to Improve Camelina Seed and Oil Quality Traits
PRESENTER: John Browse

Sabeeha Merchant: Trace Metal Storage in Metal-Specific, Lysosome-Related Vacuolar Compartments
PRESENTER: Stefan Schmollinger

Erik Sacks: Genomic-Enabled Breeding of Miscanthus
PRESENTER: Erik Sacks

Erik Sacks: Introgression of Novel Disease Resistance Genes from Miscanthus into Energycane
PRESENTER: Erik Sacks

Daniel Schachtman: Identifying productive energy sorghum germplasm for water-limited and nitrogen-limited environments using classical and advanced phenotyping methods - one piece of the puzzle for creating sustainable biomass production systems
PRESENTER: Daniel Schachtman

Daniel Schachtman: Leaf Carbon and Nitrogen Isotope Composition in Diverse Sorghum Lines Under Differential Water and Nitrogen Treatments
PRESENTER: Asaph Cousins

Danny Schnell: Engineering mitochondrial metabolic networks to increase yield and water use efficiency in 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

Eduardo Blumwald: "SyPro Poplar" Improving Poplar Biomass Production under Abiotic Stress Conditions: An Integrated Omics, Bioinformatics, Synthetic Biology and Genetic Engineering Approach
PRESENTER: Amir H. Ahkami

Jack Brown: Developing Non-food Grade Brassica Biofuel Feedstock Cultivars with High Yield, Oil Content, and Oil Quality that are Suitable for Low Input Production Dryland Systems.
PRESENTER: Jack Brown

Serge Edme: Genetics and Genomics of Pathogen Resistance in Switchgrass
PRESENTER: Serge Edme

Stephen Long: Building on success in systems design of high yielding low-input energycanes for marginal lands
PRESENTER: Stephen Long

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

George Newcombe: Characterizing the Defense Hierarchy of Populus trichocarpa and its Hybrids
PRESENTER: George Newcombe

Victoria Orphan: Comparative Genomics, Transcriptomic Analysis, and Ecophysiology of Individual Environmental Methane-oxidizing Syntrophic Consortia
PRESENTER: Connor Skennerton

Daniel Schachtman: Defining the Functional Genome Associated with Enhanced Water and Nitrogen Use Efficiency in Bioenergy Sorghum
PRESENTER: Andrea Eveland

Kenneth Sale: Co-optimization of multicomponent enzyme mixtures and saccharification conditions
PRESENTER: Carolina Araujo Barcelos

Gerald Tuskan: A Comparison of Lignocellulose Solubilization by Cellulolytic Monocultures and a Mixed Enrichment Yields Surprising Results
PRESENTER: Evert Holwerda

Gerald Tuskan: CELF/CBP for Enhanced Utilization of Polysaccharides and Lignin
PRESENTER: Charles Cai

Gerald Tuskan: Towards Understanding Plant Cell Wall Structure and Properties During Microbial Deconstruction - From Chemical Bonds to Wall Architecture
PRESENTER: Yannick Bomble

Heather Coleman: Post-harvest induction of hyperthermophilic cellulases reduces recalcitrance in poplar
PRESENTER: Heather Coleman

Jeffrey Gardner: Comprehensive Functional Characterization of the Glycoside Hydrolase Family 3 Enzymes from Cellvibrio japonicus Reveals Unique Metabolic Roles in Biomass Saccharification
PRESENTER: Estela Monge

Jonathan Schilling: Spatial Connectomics to Identify Agents Relevant to Lignocellulose Deconstruction in Fungi
PRESENTER: Jiwei Zhang

Brian Davison: Insights into Hemicellulose-Cellulose Interactions from Thermochemical Pretreatment of Model Composite Materials
PRESENTER: Brian Davison

Brian Davison: Molecular Interactions Enhanced by Co-Solvent Use for Effective Biomass Pretreatment
PRESENTER: Hugh O'Neill

James Dumesic: Fractionation of Lignocellulosic Biomass using biomass derived gamma-valerolactone as solvent
PRESENTER: Ali Hussain Motagamwala

Harry Beller: Discovery of Phenylacetate Decarboxylase, a New Glycyl Radical Enzyme Enabling First-Time Biochemical Production of Toluene
PRESENTER: Andria Rodrigues

Timothy Donohue: From Field to Fuel: How Do Feedstock Variability and Plant-Experienced Stresses Affect Microbial Conversion?
PRESENTER: Rebecca Ong

Chris Hittinger: Evolutionary Dynamics of a Secondary Metabolite Cluster in Budding Yeasts
PRESENTER: David Krause

Yong-Su Jin: Microbial metabolic engineering to produce alcohols from cellulosic hydrolysates
PRESENTER: Yong-Su Jin

Yi Lu: Expanding the Scope of Biofuels and Chemicals Produced by Microorganisms through Design of Artificial Enzymes Containing Unnatural Amino Acids and Non-native Cofactors
PRESENTER: Chang Cui

John Ralph: “Ideal lignin” facilitates full biomass utilization
PRESENTER: Yanding Li

Steven Singer: Bioconversion of Sugars and Aromatics from Engineered Plants to Methyl Ketones
PRESENTER: Steven Singer

Gerald Tuskan: Rapid Domestication of Non-Model Microbial Hosts for Biofuels Production
PRESENTER: Adam Guss

Daniel Amador-Noguez: Exposure to oxygen induces a transient bottleneck in the methylerythritol 4-phosphate pathway in Zymomonas mobilis: the role of iron-sulfur cluster assembly proteins and flavodoxin reductase in recovering pathway activity
PRESENTER: Julia Martien

Gautam Dantas: Multi-omics analysis of Rhodococcus opacus strains evolved for optimized lignocellulose bioconversion
PRESENTER: Tayte Campbell

Audrey Gasch: Rerouted PKA Signaling Coordinates Sugar And Hypoxia Responses For Anaerobic Xylose Fermentation In Yeast
PRESENTER: Kevin Myers

Costas Maranas: SteadyCom: Predicting Microbial Abundances while Ensuring Community Stability
PRESENTER: Siu Hung Joshua Chan

Christos Maravelias: A Framework for the Identification of Economically Promising Bio-Based Chemicals
PRESENTER: Tony (Wenzhao) Wu

Gabriella Papa: Development and screening of one-pot IL conversion technologies
PRESENTER: Gabriella Papa

Christopher Rao: Engineering the oleaginous yeast Rhodosporidium toruloides for the production of lipids and lipid-based chemicals
PRESENTER: Jingjing Liu

Fuzhong Zhang: Engineering Bacteria to Produce Branched-chain Fatty Acid Derived Advanced Biofuels
PRESENTER: Shaojie Wang

Gerald Tuskan: Advances in Lignin Valorization via Biological and Catalytic Transformations
PRESENTER: Gregg Beckham

Timothy Donohue: Microbial Valorization of Lignin: Using Bacteria and Their Enzymes to Develop Biological Systems for Depolymerizing Lignin
PRESENTER: Wayne Kontur

Dominique Loque: SbCOMT (Bmr12) is involved in the biosynthesis of tricin-lignin in sorghum
PRESENTER: Aymerick Eudes

TaeSeok Moon: Development of a Genetic Toolkit in Rhodococcus opacus PD630 for Lignin Valorization
PRESENTER: Drew DeLorenzo

Michelle O'Malley: Engineering Anaerobic Gut Fungi for Lignocellulose Breakdown
PRESENTER: Michelle O'Malley

Brian Davison: Impact of Hydration and Temperature History on the Structure and Dynamics of Lignin
PRESENTER: Loukas Petridis

Blake Simmons: Engineering CRISPR/Cas9 platform to industrialize lignin modifying enzymes (LMEs) using Aspergillus niger
PRESENTER: Laure Leynaud Kieffer

Huimin Zhao: Combinatorial Metabolic Engineering Using an Orthogonal Tri-functional CRISPR System
PRESENTER: Carl Schultz

Huimin Zhao: Construction of a Kinetic Model for Yeast Lipid Metabolism
PRESENTER: Shekhar Mishra

Huimin Zhao: Construction of genome-scale metabolic models for non-model yeast organisms for biofuel and bioproduct engineering
PRESENTER: Siu Hung Joshua Chan

Huimin Zhao: Genome-scale Engineering of Saccharomyces cerevisiae with Single Nucleotide Precision
PRESENTER: Pu Xue

Adam Arkin: Auxotrophy Prediction and Community Metabolic Modeling Reveals Functional Delegation in a Variety of Microbiome Systems
PRESENTER: Pamela Weisenhorn

Krishna Niyogi: Systems Analysis and Engineering of Biofuel Production in Chromochloris zofingiensis, an Emerging Model Green Alga
PRESENTER: Krishna Niyogi

Janelle Thompson: Systems biology towards a continuous platform for biofuels production: Engineering an environmentally-isolated Bacillus strain for biofuel production and recovery under supercritical CO2.
PRESENTER: Jason Boock

Gerald Tuskan: Using Multi-Omic Data to Understand the Response of Clostridium thermocellum to Deletions of Genes for Lactate and Acetate Production
PRESENTER: Daniel Olson

Costas Maranas: Development of an accelerated procedure for parameterizing kinetic metabolic models for C. thermocellum
PRESENTER: Satyakam Dash

Edward Baidoo: A Metabolomic Based Approach to Identifying Bottlenecks in Biosynthetic Pathways
PRESENTER: Edward Baidoo

Costas Maranas: novoStoic: Pathway design using de novo steps through uncharted biochemical spaces
PRESENTER: Lin Wang

George Church: Encoding new non-standard amino acid designs into proteins can address limitations of fluorescent proteins and facilitate genome editing approaches
PRESENTER: Erkin Kuru

George Church: Engineering a reduced 57-codon genetic code in Escherichia coli
PRESENTER: Alexandra Rudolph

George Church: Exploring the Species Specificity of Lambda Red Recombination
PRESENTER: Timothy Wannier

Robert Clubb: Toward Recombinant Cellulolytic Bacteria: Stabilizing Displayed Cellulases on Vegetative Bacillus subtilis Cells
PRESENTER: Jason Gosschalk

Ryan Gill: Design and engineering of native regulatory networks
PRESENTER: Emily Freed

Michael Jewett: Building a Bridge between Cell-Free Experimentation and Cellular Metabolic Engineering
PRESENTER: Ashty Karim

Himadri Pakrasi: MinGenome: An in silico top-down approach for the synthesis of minimized genomes
PRESENTER: Lin Wang

Thomas Eng: Adaptive Lab Evolution Provides a Route to Improve Biofuel Production Under Ionic Liquid Stress
PRESENTER: Thomas Eng

Adam Arkin: Assembling and Annotating Prokaryotic Genomes in KBase
PRESENTER: Benjamin Allen

Gerald Tuskan: Pleiotropic and Epistatic Network-Based Discovery: Integrated Networks for Target Gene Discovery
PRESENTER: Daniel Jacobson

Adam Arkin: Comparative Phylogenomics in KBase
PRESENTER: Priya Ranjan

Adam Arkin: Connecting JGI and KBase
PRESENTER: Shane Canon

Adam Arkin: Fungal Model Construction in KBase
PRESENTER: Janaka Edirisinghe

Adam Arkin: KBase: An Integrated Systems Biology Knowledgebase for Predictive Biological and Environmental Research
PRESENTER: Robert Cottingham

Adam Arkin: KBase's App Catalog
PRESENTER: William Riehl

Adam Arkin: KBase's Narrative Interface: A User Interface for Creating Reproducible Systems Biology Workflows
PRESENTER: Nomi Harris

Adam Arkin: Learning to Use KBase to Accelerate Your Research
PRESENTER: Meghan Drake

Adam Arkin: Microbiome Genome Extraction and Metabolic Modeling of Species Interactions in KBase
PRESENTER: Dylan Chivian

Adam Arkin: RNA-Seq Analysis in KBase
PRESENTER: Sunita Kumari

Adam Arkin: The KBase Platform
PRESENTER: Shane Canon

Adam Arkin: The KBase Software Development Kit Makes KBase an Extensible Community Resource
PRESENTER: Paramvir Dehal

Adam Arkin: Tools and Data for Metabolic Modeling in KBase
PRESENTER: José Faria

William Cannon: Multi-scale Modeling of Circadian Rhythms
PRESENTER: William Cannon

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

Mary Firestone: Using iDIRECT (Inference of Direct and Indirect Relationship with Efficient Copula-based Transitivity) to improve network modelling of rhizosphere microbial assemblages
PRESENTER: Naijia Xiao

Ronan Fleming: ARTENOLIS: Automated Reproducibility and Testing Environment for Licensed Software
PRESENTER: Ronan Fleming

Ronan Fleming: Creation and Analysis of Biochemical Constraint-based Models: the COBRA Toolbox v3.0
PRESENTER: Ronan Fleming

Ronan Fleming: Mathematical Properties Of Stoichiometric Matrices
PRESENTER: Ronan Fleming

Ronan Fleming: Numerical Properties Of Stoichiometric Matrices
PRESENTER: Ronan Fleming

Hector Garcia Martin: Quantitative metabolic modeling at the Joint BioEnergy Institute
PRESENTER: Hector Garcia Martin

Jennifer Reed: Machine learning techniques help accelarate enzyme engineering: a case study with glycoside hydrolases
PRESENTER: Sanjan Gupta

Paul Adams: Management and Analysis of ENIGMA Data using KBase
PRESENTER: Pavel Novichkov

Patrik D'haeseleer: Combining multiple functional annotation tools increases completeness of metabolic annotation
PRESENTER: Patrik D'haeseleer

Trent Northen: Synthetic Chemical Probes for Studying Lignin Deconstruction and Analysis of Biofuel Molecules Using nanostructure-initiator Mass Spectrometry (NIMS)
PRESENTER: Kai Deng

Anup Singh: An Automated Microfluidic System for Gene Editing Processes
PRESENTER: Kosuke Iwai

Jonathan Sweedler: A general, high-throughput platform for molecular prototyping of microbial cell factories via optically guided mass spectrometry
PRESENTER: Tong Si

Gerald Tuskan: Identification and Optimization of CRISPR Systems for Genome Editing in Pseudomonas putida and Clostridium thermocellum
PRESENTER: Carrie Eckert

Andrew Allen: Optimization of Genome Engineering in the Model Pennate Diatom Phaeodactylum tricornutum using CRISPR-CAS9
PRESENTER: Andrew Allen

Ana Paula Alonso: Development of Resources and Tools to Improve Oil Content and Quality in Pennycress
PRESENTER: Ana Paula Alonso

Shi-You Ding: Imaging Native Structure of Plant Cell Wall Cellulose Microfibril
PRESENTER: Wei Shen

Martin Wühr: Accurate, Sensitive, and Precise Multiplexed Proteomics using the Complement Reporter Ion Cluster

PRESENTER: Martin Wühr

Todd Yeates: Designing Self-Assembling Protein Materials for Imaging and Energy Applications

PRESENTER: Kevin Cannon

Crysten Blaby: Protein Function Discovery and Sequence-Function Understanding of Stress Tolerance in the Green Lineage

PRESENTER: Crysten Blaby

Debora Rodrigues: DEVELOPMENT OF ANALYTICAL METHODS TO INVESTIGATE ANTAGONISTIC INTERACTIONS BETWEEN BACTERIA AND FUNGI

PRESENTER: Debora Rodrigues

Louise Glass: m-CAFÉs Development of novel CRISPR-Cas technologies for precise manipulation of microbial networks

PRESENTER: Adam Deutschbauer

Nathan Hillson: DIVA Services: PCR, Full DNA Construction, and MiSeq Validation

PRESENTER: Nathan Hillson

Center for Advanced Bioenergy and Bioproducts Innovation: Sustainability Theme

Madhu Khanna[1] and Wendy Yang[1]

[1] University of Illinois Urbana-Champaign

https://cabbi.bio/

Project Goals:

CABBI aims to provide improved technologies for advanced biofuels using high yielding perennials and producing value-added products from plant-produced feedstocks or substrates and an integrated economic and environmental framework for determining feedstock supply and its sustainability. Research conducted by the Sustainability Theme of the Center seeks to improve our mechanistic understanding of ecosystem processes related to carbon, nitrogen, and water cycling to enable predictions of ecosystem service production by different feedstocks grown across many geographic regions and under variable climate. It will combine ecological and economic modeling together with big data capabilities to determine the land available for feedstock production and examine the economic viability and resilience of biomass production from the farm scale to the economywide bioproducts, and biofuels under climate- and market-induced volatility under existing and future policy scenarios.

Abstract:

The University of Illinois and 17 partner institutions are establishing the Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) for the development of transformative technologies for the economic and sustainable production of fuels and chemicals from plants. This major interdisciplinary research effort is designed to accelerate biofuel and bioproduct development while retaining the flexibility to assimilate new disruptive technologies whatever their source. CABBI is founded on the "plants as factories" paradigm, in which biofuels, bioproducts, and foundation molecules for conversion are synthesized directly in plant stems. CABBI will be built around three highly interconnected DOE priority research areas: Feedstock Development, Conversion, and Sustainability. These three themes each play an essential interconnected role into developing an overall solution to providing sustainable energy solutions for our future.

The Sustainability Theme provides an integrative framework for CABBI by assessing the environmental and economic outcomes from feedstock production and conversion to biofuels snd bioproducts. The Sustainability Theme will include five major areas of investigation and implementation – Experimentation, Ecosystem Modeling, Big Data Analysis, Refinery-scale Techno-economic Analysis, and Integrated Economic Analysis. A multi-disciplinary team of investigators will improve mechanistic understanding of the links between traits of feedstocks and the ecosystem services they provide and analyze the characteristics of refinery-scale processes that determine the economic viability and environmental outcomes associated with different combinations of biofuels and bioproducts and quantify the ecological and economic implications of large scale energy crop production. It will provide insights on where, when, and how crops are planted and managed as well as the economics and energy inputs of different

conversion processes which will ultimately influence the impact of an emerging bioeconomy on the environment. The Sustainability Theme is grounded in state-of-the-art empirical measurements of how different crops and agronomic practices affect carbon, nitrogen, and water cycles, with a focus on soil microbe-plant interactions, and extends these data in space and time through process-based modeling. A unique feature of the Sustainability Theme is the ability to integrate ecological, biophysical, and economic models to holistically evaluate sustainability. The Sustainability Theme's research will prioritize minimizing competition with food production and using deep-rooted, low input bioenergy crops to improve marginal lands. Two globally-recognized process-based models, DayCent and AgroIBIS, that have been at the forefront of bioenergy crop modeling will be used to simulate the effects of climate and environmental change on carbon and nutrient cycling in terrestrial ecosystems, employing a similar approach to simulating below ground processes and contrasting approaches to simulating C assimilation. Economic and biophysical models will be integrated to develop a multi-scale, spatially explicit capability to determine the land available to optimize feedstock productivity across a mix of feedstock types, and its potential to lead to a range of improved environmental outcomes. CABBI research will examine the economic viability and resilience of biomass production from the farm scale to the economywide scale to meet local, regional and national demands for biopower, bioproducts, and biofuels under climate- and market-induced volatility under existing and future policy scenarios.

Agile Technoeconomic and Life Cycle Assessments at the CBI

Mary J. Biddy[1*] (mary.biddy@nrel.gov), Erin Webb[2], Nicholas Grundl[1], Rebecca Hanes[1], Brian H. Davison[2] and **Gerald A. Tuskan**[2]

[1] National Renewable Energy Laboratory, Golden, Colorado; [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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

TEA quantifies the impacts that research and development (R&D) breakthroughs and discoveries have on the economic viability of an integrated process. Technoeconomic analyses (TEA) is a highly supportive research tool which can identify drivers and metrics to reduce the overall costs, as well as outline the lowest cost potential, of an integrated system. Life cycle assessments (LCA) explores the key sustainability metrics critical to ensure that an underlying integrated process offer an environmentally beneficial design, which is the ultimate desired outcome when pursuing biomass-derived fuels and chemicals. The combined application of TEA and LCA allows for understanding the tradeoffs and benefits of both economics and sustainability in these integrated systems.

The overall objective of the Economic and Sustainability Team is to quantitatively assess the impact that the CBI R&D will have on the reduction of costs and scale-up risks for the integrated production of biomass through conversion to fuels. CBI will employ agile TEA and LCA to investigate the drivers of phenotypes and process conversion metrics on the integrated system design. The analysis tools and studies developed in this project will support research prioritization to ensure that the ongoing basic science being performed by the CBI is targeted toward de-risking technology bottlenecks and overcoming critical economic and technical barriers in order to accelerate growth of the bioeconomy. Our goal is to develop analyses that not

only support the mission of the CBI, but also to release such tools as user-friendly guides to help direct and inform the broader biofuels community.

The biggest challenges to evaluating economic and sustainability drivers in an integrated fashion are that 1) the underlying models and tools used for the evaluations typically require a subject matter expert to properly utilize these tools, 2) these models and tools often require expensive licenses to operate, and 3) these models are not rigorously integrated. The analysis team will work to reduce these challenges by developing first-of-a-kind integrated, agile analytical tools to provide a prioritized order of magnitude ranking of the potential impacts of CBI research on industrial cost and sustainability barriers. Specific illustrative integrated scenarios will then be compared to the robust specialized rigorous approaches. By applying a holistic assessment of the design, these TEAs and LCAs will provide rapid feedback to experimentalists and will serve as a decision-making tool for dynamic project management. This poster will outline the approach and the models to be incorporated.

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.

Technoeconomic Analysis of Biofuel Feedstock Supplies to Inform Plant Development R&D

Erin G. Webb[1*] (webbeg@ornl.gov), Tom Richard[2], Mary J. Biddy[3], Brian H. Davison[1], Wellington Muchero[1], Katrien Devos[4] and **Gerald A. Tuskan**[1]

[1]The Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2] Penn State University, University Park, Pennsylvania; [3] National Renewable Energy Laboratory, Golden, Colorado; and [4]University of Georgia, Athens, Georgia

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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

A key focus of CBI R&D is overcoming economic barriers related to supplying high-yielding, robust feedstocks by identifying and utilizing key plant genes for growth, yield, and composition for two perennial feedstocks - poplar and switchgrass. The CBI Economics and Sustainability Analyses Team will develop a TEA model to quantify the impacts of CBI fundamental R&D in plant development and sustainability as a means of identifying barriers that impact the feedstock supply chain. This model will be integrated with a conversion TEA component and sustainability analyses to create a set of field-to-end product analytical tools that will provide a prioritized order of magnitude ranking of the potential impacts of CBI research on industrial cost and sustainability barriers. By applying a holistic assessment, the TEA model will provide rapid feedback to experimentalists and will serve as a decision-making tool for dynamic project management.

Key science questions to be addressed via TEA for biomass feedstock development will be 1) deriving assumptions for industrial-scale models from bench-scale production or single plant trials and 2) assessing tradeoffs between maximizing crop yield and uniformity. Working with the poplar and switchgrass teams, and using information from peer-reviewed literature, we will evaluate the process parameters critical to moving from bench-scale processes to a range of larger process scales (e.g., from greenhouse to field to regional production). Recognizing that, taken directly, measurements from highly controlled, small-scale experiments in the greenhouse often provide overly optimistic representations of system performance at an industrial scale we will develop scale-up correction factors and identify data gaps to fill in missing information.

Additionally, we will estimate the variability associated with key parameters and the uncertainties associated with applying such correlations. These parameter assumptions will then be applied in a TEA context to evaluate upstream supply chain impacts of plant development breakthroughs. We hypothesize there will be trade-offs between selecting plants for yield (or quality) and robustness. From a biorefinery perspective, we will determine which set of phenotypes is preferred: high yields that may vary significantly from year to year and region to region, or lower yielding, but more consistent yields, both spatially and temporally.

This work builds on previous analysis by team members to integrate economic and sustainability analysis with applied R&D to quantify the economic and life cycle benefits of engineering breakthroughs to produce biofuels. However, most previous analyses assumed "mature" technologies and typical yields that neglect the risks, delays, and cost-overruns commonly encountered in commercializing new crops or emerging technologies and ventures. As such, these analyses often underestimate the substantial challenges of scaling up complex systems with novel feedstock organisms and operations and do not consider the variability of feedstock characteristics that also translate into business risk. In this project, we will design new tools that address the economics of emerging technologies at early stages of development. In addition, stochastic variability will be implemented in the economic evaluations to better understand the current operating costs, quantify the impacts of feedstock quality and uniformity on cost and system robustness, and identify priorities to improve system resilience.

The Center for Bioenergy Innovation is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science

Simulating Near-Term Scale-Up of Cellulosic Biofuel Production Using Crop Residues and Biomass Sorghum

Xinguang Cui*[1,2], **Corinne Scown**[1,2]

[1]Joint BioEnergy Institute (JBEI), [2]Lawrence Berkeley National Laboratory

Project Goals: Model biomass feedstock availability for cellulosic biofuels at a fine geospatial resolution using publicly-available datasets and tools, and develop scenarios focused on increasing biomass availability and fuel production in the near-term (next 5-10 years).

The Renewable Fuel Standard (RFS) established in the Energy Independence and Security Act of 2007 aimed at reaching 16 billion gallons of cellulosic biofuels, out of a total of 36 billion gallons biofuel production by 2022. The project presented here focuses on near-term steps toward achieving that goal. We run scenarios in which the current biorefineries will be retrofitted to process cellulosic feedstocks including the crop residues such as the corn stover and dedicated energy crops. Then new refineries are to be built based on where additional biomass exists beyond the reach of current facilities. Thus, an important first step is to assess the biomass availability of feedstocks to provide practical suggestions for retrofitting current biorefineries and build new biorefineries. To accomplish this goal, a high-resolution feedstock analysis model is designed by making use of the county-level feedstock biomass production data set and high resolution satellite land cover data set (30mx30m). This system is created based on the open source databases in the platform of open source program (R), which aims to provide the biofuel community a powerful tool to plan and analyze availability of the cellulosic feedstocks. The model includes evaluation of locations for potential new biorefineries, aggregating clusters of crops, calculating 'biosheds' which redistribute the clusters to bio-refineries, calculation of biomass production, and post-process scripts. With this system, we have studied the biomass availability of current biorefineries and new biorefineries based on different feedstock blending constraints, and evaluated the potential contribution of biomass sorghum (an annual crop that farmers are more likely to adopt in near term). Several factors are considered such as the period, yield variation, biomass price, driving distance. In consideration of base yield, at the year of 2030, price of 50$/dt and driving distance of 50 mile, we have identified that there are 24 corn stover-exclusive biorefineries among 213 current biorefineries and 25 if the biorefineries can process other feedstocks such as switchgrass and Miscanthus by setting the minimum biorefinery size to 800,000 (dt) dry ton of biomass input per year. It is not necessary to build new biorefineries for corn stover in this case, but it is necessary to build another 181 bio-refineries that can process both the crop residuals and energy crops. In short, we have identified ~10% of current biorefineries that could feasibly be retrofitted to process corn stover exclusively, the remaining facilities will require blends of corn stover and energy crops, and there are as many as ~200 new potential biorefinery locations. Our results provide detailed maps indicating where existing facilities can be scaled up under different biomass availability scenarios, where farmers could add sorghum rotations to increase short-term biomass availability, and how many more cellulosic biorefineries could be constructed if biomass production is increased. In conclusion, our model provides the biofuel community with a powerful numerical tool to assess the supply of the biomass production for each biorefinery, at a fine geospatial scale

appropriate for analyses ranging from individual localities to national-scale. which has shown its capability to assess the availability of biomass production for both of the current and potential bio-refineries.

Economic and Environmental Modeling of Bioenergy Production – Policy Impacts on Water Quality

Authors: Kelsie Ferin, Jia Zhong, Weiwei Wang, Andy VanLoocke, Madhu Khanna

Bioenergy feedstocks can provide many potential ecosystem services but also pose many tradeoffs and challenges. Key examples include the tradeoff of provisioning energy and bioproducts, reducing carbon emissions and reducing nitrogen run-off from cropland versus competing for land with food and or natural habitat. Current production systems, in particular annual croplands, are major contributors to nitrogen pollution in surface waters and a driver for the Gulf of Mexico Hypoxic Zone. Therefore, opportunities exist to improve water quality and provide other ecosystem services by growing energy crops on cropland, but balancing these with food and feed production and costs of feedstock production remains a challenge. A predictive, policy oriented approach is required to accurately investigate these tradeoffs at scale.

This poster will describe our proposed research to conduct an integrated set of experiments that combine economic and ecosystem modelling approaches to aid in assessing tradeoffs between current production systems and advanced bioenergy systems. Our approach will combine BEPAM (Biofuel and Environmental Policy Analysis Model) economic model and Agro-IBIS (Integrated Biosphere Simulator – Agricultural Version) agroecosystem model with state-of-the-art spatial mapping of land characteristics in the Central and Eastern US. The BEPAM model is a multi-market, dynamic, open economy model that integrates the fuel sectors in the US, including agricultural and forestry sectors. Agro-IBIS is a process-based agro-ecosystem model that simulates the physical cycling of carbon, nitrogen, water and energy in managed and unmanaged systems. Both models have been used separately to quantify economic and water quality impacts of bioenergy production. The key advancement in the current project is to use the models together to constrain and increase the complexity and breadth of modeling capabilities.

The poster will present our findings using the BEPAM and Agro-IBIS models. Specifically, it will show the land types and locations where energy crops can be produced economically to meet demands posed by a biofuel mandate determined using BEPAM. It will also show the findings from the Agro-IBIS models on the implications of converting cropland to varying levels of energy crop production on nitrate run-off and hypoxia in the Gulf of Mexico. We will describe proposed integrated systems approach to link the two models and discuss the implications of our findings for the design of policy to meet both energy production and ecosystem service provision goals simultaneously.

Progress Towards N_2O Source Apportionment in Biofuel Soils: Understanding Sources of Variation in Isotopic Discrimination During Denitrification

Joshua Haslun[1,2*] (haslunjo@msu.edu), **Nathaniel Ostrom[1,2]**, Clarisse Finders[1,2], **Eric Hegg[1,2]**, Peggy Ostrom[1,2]

[1]Michigan State University, East Lansing, MI; [2]DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI

https://bmb.natsci.msu.edu/faculty/eric-l-hegg/current-research/research-project-4/

Project Goals: Maximizing the environmental performance of bioenergy crops is critical to their productivity in marginal lands. Management of nitrogen loss in the form of nitrous oxide (N_2O) emissions, an important greenhouse gas and the leading source of ozone depletion, can improve productivity and mitigate climate effects. Modelling N_2O emissions is a considerable challenge in that soil N_2O emissions are principally caused by microbes through diverse processes. Spatiotemporal variation of the microbial communities, the pathways active in individual species, and environmental factors such as soil structure and vegetation further complicate this challenge. Our goal is to employ a combination of new methodologies including the analysis of N_2O isotope values to disentangle the processes contributing to N_2O formation *in situ*. Knowledge of the active processes can be combined with that of the rhizosphere microbial communities to help identify plant traits that promote rhizosphere N_2O reduction by microbes.

Efforts to mitigate agricultural N-losses as N_2O emissions commonly focus on reducing fertilization rates; however management of the microbial processes directly responsible for N_2O production can also reduce N_2O emissions. To this end, evaluation of stable isotope values observed during production of N_2O has been applied to differentiate between the two predominant microbial processes responsible for its production, nitrification and denitrification. To better understand the factors contributing to isotopic variation during denitrification, we have characterized under a range of carbon-source conditions the $\delta^{15}N$, $\delta^{15}N^{\alpha}$, $\delta^{15}N^{\beta}$, $\delta^{18}O$, and site preference (SP; the intramolecular distribution of ^{15}N in N_2O) of N_2O produced during NO_3^- reduction by two denitrifying bacteria lacking nitrous oxide reductase. The evaluation of microbial N_2O sources is challenged because the isotopic discrimination for $\delta^{15}N$ and $\delta^{18}O$ during production, while expected to be constant, can vary markedly over the course of the reaction. Additionally, the multi-step nature of denitrification violates critical assumptions of the traditional Rayleigh approximation used to estimate changes in $\delta^{15}N$ and $\delta^{18}O$ values of a single-step unidirectional reaction. The change in isotope value of a multi-step reaction is more aptly termed a net isotope effect (NIE), because many isotopic fractionations can contribute to the observed isotopic discrimination of the final product. We demonstrate that the NIE varies by as much as 100 ‰ over the course of a single reaction, which clearly limits the use of $\delta^{15}N$ and $\delta^{18}O$ as apportionment tools. Given our observations of changes in the degree of fractionation of $\delta^{15}N$, $\delta^{15}N^{\alpha}$, $\delta^{15}N^{\beta}$, and $\delta^{18}O$ across all treatments, we developed a non-linear approach to estimate this changing isotopic discrimination associated with multi-step reactions, such as denitrification. In contrast, SP values for denitrification did not change over the course of the

reaction, although the mean SP of N_2O produced by each species differed. Therefore, SP remains a robust indicator of the origin of N_2O.

To better understand the observed non-linear change in isotope values during denitrification and the potential sources of species-specific variation in SP, we characterized isotopic fractionation during *in vivo* reduction of nitric oxide (NO) to N_2O by *P. aureofaciens* and *P. chlororaphis*. The enzyme cNOR is thought to be the primary enzyme responsible for the reduction of NO to N_2O in soils and is present in both of the aforementioned species. Therefore we also performed *in vitro* reduction of NO by isolated cytochrome *c*-dependent NO reductase (cNOR) from *Paracoccus denitrificans* to determine the intrinsic fractionation associated with this enzyme. Collectively, these data suggest that species-specific factors likely contribute to the fractionation expressed during NO reduction and provide insight into the yet unknown mechanism of N_2O production by cNOR.

Funding statement

This work was funded by the National Science Foundation's Earth Sciences Instrumentation and Facilities program (grant #1456430) and DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

Carbon Cycling in Biomass-Based Cropping Systems: Improving the Mechanistic Foundations of Plant, Microbial, and Soil Interactions

Adam C. von Haden[1]* (avh@illinois.edu), **Evan H. DeLucia**[1], and **Wendy H. Yang**[1]

[1]University of Illinois, Center for Advanced Bioenergy and Bioproducts Innovation, Urbana, IL 61801 USA

https://cabbi.bio/research/sustainability-theme/

Project Goals:

The reduction of fossil fuel-derived carbon emissions provided by the emerging bioeconomy represents a critical step toward meeting energy independence goals. Decisions regarding feedstock selection, geographic placement within the landscape, and specific management practices will impact the overall efficacy of ecosystem carbon sequestration and productivity benefits. Therefore, we seek to advance a comprehensive mechanistic understanding of plant-soil-microbe interactions within biomass-based cropping systems that can maximize the potential benefits by providing a more robust framework for predicting ecosystem carbon dynamics. The overarching goal of this project is to use field-based measurements of key ecosystem carbon processes to inform and improve next generation mechanistic ecosystem models.

Abstract text:

The use of plant biomass as feedstocks for bioenergy and bioproducts is expected to offer both economic and ecological benefits over traditional fossil-derived sources. The potential net reduction of carbon emissions is a major environmental advantage of plant biomass-derived feedstocks over fossil fuels. However, the magnitude of carbon emissions reduction within a biomass cropping system depends on upon the efficiency of ecosystem carbon sequestration. Management decisions concerning plant feedstock type, soil tillage, and residue removal rates will likely affect carbon sequestration rates. In addition, the efficacy of ecosystem carbon storage within biomass cropping systems will vary geospatially due to natural differences in edaphic properties. An extensive understanding of interactions between plants, microbes, and soil is required to adequately predict ecosystem carbon dynamics and thus maximize the potential carbon sequestration benefits across the landscape.

A key challenge of studying ecosystem carbon exchange *in situ* is isolating the soil carbon dioxide (CO_2) fluxes from plant root versus soil microbial metabolism. Although the total CO_2 flux can be measured relatively easily at the soil surface, this measurement does not provide any information about the two discrete sources and therefore is less insightful. Since the two sources of CO_2 may have different $\delta^{13}C$ isotopic ratios, they can be separated by measuring $^{13}CO_2$ from soil gas exchange. We will integrate an automated chamber-based soil gas sampling system with an advanced Cavity Ring-Down Spectrometer to provide high-frequency *in situ* measurements of

CO_2 fluxes from both plant roots and soil microbes. This information will help to better understand the underlying environmental and physiological controls on plant root CO_2 respiration, to constrain the quantity of carbon utilized by the soil microbial community, and to calibrate and validate ecosystem-level carbon models.

The soil environment represents the complex interaction of biological, chemical, and physical attributes. These properties have traditionally been simplified within the soil organic matter models that are used to make predictions of soil organic carbon change. Next generation soil organic models explicitly represent many of the complex soil processes, including microbial substrate use, carbon-nitrogen coupling, and physical protection of organic carbon by soil aggregates or mineral surfaces. However, these models have not been thoroughly tested in an agricultural setting, and therefore their performance within biomass cropping systems remains uncertain. We will provide critical empirical data on key soil factors including microbial carbon use efficiency, nitrogen mineralization, and plant root dynamics. We will work closely with model developers to refine and test the improved soil organic matter models using lab and field-based observations. These efforts will result in greater predictive capabilities for carbon sequestration in both biomass-based and traditional row crop agricultural systems.

Characterizing the Multitrophic Interactions that Mediate Carbon Flow in Soil

Javier A. Ceja-Navarro (JCNavarro@lbl.gov)[1*], Abelardo Arellano[1], Leila Ramanculova[1], Katerina Estera[2], Kateryna Zhalnina[1], Henrik Krehenwinkle[2], Nhu Nguyen[3], **Rosemary Gillespie**[2], **Jennifer Pett-Ridge**[4], **Mary Firestone**[2], **Eoin Brodie**[1,2] (elbrodie@lbnl.gov)

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of California, Berkeley, CA; [3] University of Hawaii, Honolulu, HI; [4]Lawrence Livermore National Laboratory

Project Goals: This project aims to understand the effects of changing precipitation regimes on the composition, function and interactions among biological compartments (microbial and eukaryotic) of the soil food web and the consequences for C flow in the rhizosphere. Approaches for the separation, quantification, and identification of soil fauna are being optimized and combined with high-throughput sequencing and stable isotope probing techniques, to characterize fauna populations associated with the rhizosphere of *Avena fatua* and their contribution to the dynamics of C transformation and stabilization. A primary goal is to develop robust mathematical models of soil trophic networks that extend beyond bacteria and fungi, and thus require information on faunal community composition, body size, and feeding preferences.

In all terrestrial ecosystems, numerous populations of organisms such as protists, nematodes and arthropods interact with the soil's free-living and plant-associated microbes, defining biogeochemical and nutrient cycling processes. The objective of the research proposed here is to illuminate the contribution of bacterial, archaeal and eukaryotic communities, as components of a multitrophic network, to carbon and nutrient cycling in soil, with their mechanistic basis illuminated through the application of multi-omics approaches, stable isotope tracing, and field manipulations.

To predict the responses of C and nutrient cycling to environmental change, it is important to recognize that these environmental processes are the result of the interactions of multiple groups of organisms that in concert shape an ecosystem. The first step for the construction of mathematical models to define environmental multitrophic interactions is to characterize the different trophic nodes that are part of an ecosystem. For this, we have tested and optimized different approaches to characterize populations of arthropods, nematodes, and protists from two different soil ecosystems: a switchgrass plantation in a marginal soil in southern OK, and a mediterranean annual grassland.

Characterization of soil arthropod community composition. We separate total arthropod communities from soil samples using Berlese funnels. Arthropods are surface- cleaned, individually imaged and measured, and their DNA extracted for identification by sequencing and phylogenetic analyses. These approaches have allowed us to characterize arthropod dynamics during switchgrass development. Populations of fungal-feeding collembola and mites first emerge during plant growth, followed by increased densities of root-feeding beetle larvae and plant

pathogenic thrips. Next, spider populations' bloom while collembolan populations decline, possibly impacting soil fungal population dynamics.

Old meets new: physical isolation and sequence based characterization of soil nematode and protist populations. Our initial attempts to use direct DNA extraction and "universal" molecular marker amplification to quantify and characterize nematode and protist populations yielded eukaryotic populations that were mainly dominated by plant and fungal sequences with minimal contribution of micro- and meio-fauna. Based on these results, we re-directed our efforts to develop techniques for the physical separation of nematodes and protists from the soil matrix. We developed an optimized isolation protocol based on gradient centrifugation and size selection by differential filtration. Subsamples of the isolated pools of nematodes and protist are then used for DNA extraction, fixing of specimens, and imaging. Sequenced libraries yielded 50-70% fauna sequences; remaining sequences were mainly fungal. This technique has allowed us to characterize nematode and protist densities in bulk and rhizosphere, confirming that the rhizosphere is a hotspot for nematode and protist populations. Predicted functional roles of these species suggests direct root feeding in addition to bacterial and fungal predation. These functional predictions are being evaluated using isotopic methods. We are designing new primers (targeting specific fauna groups) to enable quantification of these groups by qPCR.

Physical isolation of arthropods, nematodes, and protists has allowed us to define the body size distribution of individual fauna populations; sizes range from 20 um to 1 cm – a property directly related to metabolic rates. This information, together with predicted functional roles, will later be used to parameterize and optimize food-web models.

The approaches developed here will provide the foundation for molecular approaches to quantitative study soil trophic networks, and also have potential applications as diagnostic tools to identify and intervene for the early control of plant pathogenic arthropods and nematodes in bioenergy cropping systems.

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 Awards Numbers DE-SC0010570 and DE-SC0016247 to UC Berkeley. Work at Lawrence Berkeley National Laboratory was performed under U.S. Department of Energy Contract No. DE-AC02-05CH11231 and work at LLNL, under U.S. Department of Energy Contract No. DE-AC52-07NA27344.

Determination of the Roles of Pyrophilous Microbes in the Breakdown and Sequestration of Pyrolyzed Forms of SOM

Thea Whitman,[1,*](twhitman@wisc.edu), Matthew Traxler,[2] Igor Grigoriev,[3] and **Thomas D Bruns**[2]

[1]Dept. Soil Science, Univ. Wisconsin, Madison; [2]Dept. Plant & Microbial Biol., Univ. Cal., Berkeley, CA; [3]US DOE Joint Genome Institute, Walnut Creek, CA

http://plantandmicrobiology.berkeley.edu/profile/bruns

https://whitmanlab.soils.wisc.edu/research/

Project Goals:

The frequency of large, high severity wild fires is increasing in the western US and in regions around the world due to long-term fire suppression strategies and climate change [1]. These fires have direct, negative effects on soil carbon stocks through combustion, but they have indirect and potentially positive effects on soil carbon stocks through the production of pyrogenic organic matter (PyOM) that has a long residence time and constitutes a major pool of C in fire-prone ecosystems [2-4]. Soil microbes are likely to be involved with the degradation of all of these compounds, yet little is currently known about the organisms or metabolic processes involved. **We are dissecting the effects of microbes on post-fire soil carbon dynamics by using a systems biology approach that couples small experimental "pyrocosms", highly controlled production of ^{13}C-labeled pyrolyzed substrates, genomics, transcriptomics, stable isotope techniques, and mass spectrometry to address the following objectives:**

Objectives:

1. Develop improved genomic and other -omic resources for the dominant microbes of fire-affected soils

2. Determine the temporal response of soil microbes to fire and to PyOM additions

3. Characterize the temporal patterns of degradation of different sub-fractions of PyOM

Objective 1. An important component of this research is to develop and improve the genomic, proteomic, and metabolomic resources available for studying the microorganisms that dominate fire-affected soils. The fungal community in a model fire-affected soil has been well characterized, and fungal isolates from this community have been isolated and sequenced by the project team (see talk Monday afternoon). Similarly, early attempts at cultivating members of the bacterial community in these soils has produced >50 bacterial isolates that are able to grow with highly-aromatic PyOM as their primary carbon source. The majority of these isolates have been identified by 16S rRNA genes as members of the *Actinobacteria* phylum with the genera *Streptomyces* and *Pseudonocardia* being most prominent. Isolates are being evaluated for more complete sequencing so that they their genomes can serve as references for additional analysis, will be subjected to metabolic profiling, and possibly used as candidates for future whole-genome sequencing. From our collection of fungal isolates, we have so far submitted 13 pyrophilous fungi genomes and 11 transcriptomes as part of a JGI community sequencing project. Two genomes, *Crassisporium funariophilum* and *Pholiota highlandensis*, have been

assembled and annotated, and three more genomes are currently in sequencing. We are also in the process of growing a dominant pyrophilous fungal species, *Pyronema omphalodes*, on forest soil treated by a controlled burn to collect transcriptomic and metabolic data at individual time points, temperature profiles and soil depths.

Objectives 2 and 3. Temporal patterns in the degradation of PyOM and its sub-fractions will be assessed by extracting ^{13}C-labelled polar and non-polar compounds and adding them to unlabeled char particles prior to incubation with soils. In order to trace degradation of each fraction, we have built an automated custom gas-sampling "multiplexer" connected to a cavity ringdown spectrometer, through which we can measure the efflux and isotopic composition of respired CO_2. The ^{13}C labelled PyOM for these experiments is being produced from pine biomass grown in a labeling chamber constructed for this project. The biomass will be charred in a custom-built "charcoalator", which allows for highly-controlled production of PyOM. The PyOM can be pyrolyzed over a range of temperatures, consistent with the physicochemical gradient model that evokes the steep heat gradient produced by forest fires to predict the effects of fire on the chemical properties of soils and biomass post-fire.

References

1. Westerling, A.L., H.G. Hidalgo, D.R. Cayan, and T.W. Swetnam, Warming and earlier spring increase western US forest wildfire activity. Science, 2006. 313(5789): p. 940-943.
2. Gonzalez-Perez, J.A., F.J. Gonzalez-Vila, G. Almendros, and H. Knicker, The effect of fire on soil organic matter - a review. Environment International, 2004. 30(6): p. 855-870.
3. Santin, C., S.H. Doerr, C.M. Preston, and G. Gonzalez-Rodriguez, Pyrogenic organic matter production from wildfires: a missing sink in the global carbon cycle. Global Change Biology, 2015. 21(4): p. 1621-1633.
4. Lehmann, J., J. Skjemstad, S. Sohi, J. Carter, M. Barson, P. Falloon, K. Coleman, P. Woodbury, and E. Krull, Australian climate-carbon cycle feedback reduced by soil black carbon. Nature Geoscience, 2008. 1(12): p. 832-835.

Funding Statement

This work was funded by the Department of Energy, Systems Biology Enabled Research on the Roles of Microbial Communities in Carbon Cycle Processes program, grant DE-SC0016365 to Thomas D. Bruns, Thea Whitman, 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.

Microbial contributions to carbon cycling differ qualitatively and quantitatively in agricultural, forest and meadow soils

Samuel E. Barnett,[1]* (seb369@cornell.edu), Nicholas D. Youngblut,[1] and **Daniel H. Buckley**[1]

[1]School of Integrative Plant Science, Cornell University, Ithaca, NY

Project Goals: This research program will reveal fundamental aspects of soil C-cycling and provide ecological and metabolic insights on diverse non-cultivated soil microorganisms that play major roles in the global C-cycle. Specific goals include: 1) Map the C assimilation dynamics for thousands of non-cultivated microorganisms in soil by harnessing a full cycle microbial food web mapping approach that employs an array of ^{13}C-labeled molecules; 2) Map the C assimilation dynamics of soil microorganisms across soil systems as a function of soil characteristics; and 3) Evaluate ecological and seasonal patterns of activity and abundance for discrete microbial taxa across gradients of soil characteristics and as a function of their C-assimilation dynamics. These goals will be achieved by employing a newly developed microbial food web mapping approach, enabled by advances in ^{13}C-stable isotope probing of nucleic acids and next generation sequencing.

Soil is one of the largest terrestrial carbon stocks on Earth containing thousands of peta grams of soil organic matter (SOM). Microbes are primarily agents of SOM flux, contributing significantly to the global carbon cycle. Microbial community composition significantly varies between soil habitats, but it remains unclear to what degree these differences affect the quality and rate of C-cycling. To explore this question, we conducted a high-resolution DNA-stable isotope probing experiment in which we applied five model SOM substrates (xylose, amino acids, vanillin, cellulose, and palmitic acid) to soils from forest, meadow and agricultural lands. The sites were spatially contiguous and edaphically similar, differing primarily by historical land management practices. Over a period of 30 days we measured substrate specific respiration (as $^{13}CO_2$) and identified bacterial operational taxonomic units (OTUs) that metabolized each substrate as a function of ^{13}C assimilation into DNA.

We found that mineralization rates differed among substrates and habitats. Soluble substrates such as xylose and amino acids were mineralized rapidly while insoluble substrates such as cellulose and palmitic acid were mineralized slowly in all habitats. Vanillin showed the most variability in mineralization rate across habitat type with the highest rate observed in the agricultural soil and the lowest in the meadow. Community composition differed significantly among habitats and changed over time in response to substrate addition. Temporal variability was greatest in the agricultural soil. The identities of bacteria that assimilated ^{13}C differed by habitat and substrate type. Out of 779 responder OTUs most were habitat specific (78.6%), while a smaller set (7.5%) responded in all three habitats. Responder taxa differed at higher phylogenetic ranks among habitats. For example, at peak respiration, *Planctomycetes* were prominent assimilators of palmitic acid in forest soil, yet no *Planctomycetes* were responders in agricultural or meadow soils. Determining the ecological basis for differences among active microbial groups and habitats will improve understanding of SOM turnover and carbon fate in soil.

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.

The Role of Surface-related Microbial Ecology in Organo-Mineral Stabilization of Carbon in Soil

Roland C. Wilhelm* (rwilhelm@cornell.edu), Samuel E. Barnett and **Daniel H. Buckley**

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

Project Goals

This project will improve our understanding of fundamental ecological determinants of soil C-cycling and provide insights into the metabolism of diverse non-cultivated soil microorganisms. The specific goals include: mapping the microbial carbon food web using cultivation-independent stable isotope probing (SIP) and identifying metabolism that impacts the stabilization of carbon in soil. These efforts are aimed at placing the ecology of soil communities in the greater context of terrestrial carbon cycling to inform and improve modeling and management practices.

Abstract

Microbial transformations of soil organic matter influence its accumulation and stabilization on soil minerals. Microbial metabolism alters the sorptive and adhesive properties of organic matter (OM) both directly (via enzymatic reactions) and indirectly (by conversion into living and dead biomass). The fate of carbon within organo-mineral complexes depends on a balance of environmental and ecological factors promoting accumulation or mineralization. In previous stable isotope probing (SIP) experiments[1,2] with ^{13}C, we observed that microorganisms that degrade soluble versus insoluble forms of OM exhibit different competitive strategies. The majority of insoluble OM degraders were commonly abundant in soil, yet belonged to groups with few cultured representatives, such as *Verrucomicrobia*, *Chloroflexi,* and *Planctomycetes*. The capacity of these groups to colonize and degrade insoluble OM, and their recalcitrance to culturing, support the hypothesis that their ecology is, in part, linked to surface-mediated interactions, which are difficult to reproduce *in vitro*. We designed SIP experiments to determine whether differences in microbial metabolism of soluble and insoluble substrates in soil cause differences in C-fate. In the first experiment, we used metabolomics and metagenomics to assess the fate of ^{13}C from ^{13}C-glucose and ^{13}C-cellulose (i.e. the insoluble form of glucose) added to soil, and we assessed the fate of ^{13}C from ^{13}C-vanillin added either in soluble or mineral-sorbed form. We hypothesize that metabolite pools from surface-colonizing populations will be more heterogeneous and their genomes will exhibit signatures of metabolic dependency, reflecting greater microbe-microbe interaction in resource exploitation. We expect surface-associated processes to produce more stable OM than processes that take place in soil solution, which we will test using a suite of biological and physical tests for carbon stability. In a second experiment, we will load soil with biomass from ^{13}C-labeled *P. putida* or *B. subtilis* mutants, including both biofilm overexpression and biofilm deficient phenotypes, and use this soil in a SIP experiment to measure differences in metabolism and stability of carbon from their biomass. This on-going research aims to demonstrate how ecological tradeoffs associated with living on surfaces or in soil solution, effect soil carbon cycling and sequestration.

References

1. Pepe-Ranney C and Campbell AN, Koechli CN, Berthrong S, Buckley DH. (2016). Unearthing the ecology of soil microorganisms using a high-resolution DNA-SIP approach to explore cellulose and xylose metabolism in soil. *Front Microbiol*, **7**: 703.

2. Youngblut ND, Koechli CN and Buckley DH. (*In Preparation*). Mapping microbial food web dynamics in soil with high resolution stable isotope probing.

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.

Linking Genes to Ecosystem Processes: Microbial Nitrogen and Carbon Cycling

Michaeline B.N. Albright[1*] (malbright@lanl.gov), Renee Johansen[1], Deanna Lopez[1], Bibek Timalsina[1], Laverne Gallegos-Graves[1], Cheryl R. Kuske[1], and **John Dunbar**[1]

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

http://www.lanl.gov/org/padste/adcles/bioscience/bioenergy-biome-sciences/soil-carbon.php

Project Goals:

Project 1: Identify nitrogen (N) deposition linked molecular markers in a short-term controlled laboratory experiment to develop the capability of monitoring N- responses in more complex field settings.

Project 2: Determine the prevalence of incomplete versus complete N cycling pathways, as well as the distribution of N cycling pathways across sequenced prokaryotic genomes, to inform meta-omic studies using these genetic markers.

Project 3: Describe the spatial distribution of bacteria and fungi on plant litter during decomposition to assess the potential for nutrient exchanges among microbes.

Abstract

The LANL Genomic Science 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.

The work presented here addresses BER grand challenge 4.2: define the sphere of influence and key elements of microbial communities in space and time relevant for predicting larger-scale ecosystem phenomena for Earth system understanding. Genomic data needs to be simplified for tractable application in ecosystem models. However, current approaches of consolidation have not been effective. The abundances of functional genes are inconsistently correlated with associated process rates. Moreover, studies routinely find ecosystem-specific responses to environmental perturbations such as N deposition instead of common patterns among ecosystems. Achieving the goal of predictive biology will require new ways of characterizing microbial communities and their functions in terrestrial environments.

Here, we summarize three complementary, but independent projects. The first project focuses on identifying genetic marker response to N addition. Numerous studies have examined the long-term effect of experimental N deposition in terrestrial ecosystems, however N-specific mechanistic markers are difficult to disentangle from responses to other environmental changes. The strongest picture of N-responsive mechanistic markers is likely to arise from measurements over a short (hours to days) timescale immediately after the deposition of inorganic N. To test the hypothesis that N fertilization represses the initial expression of fungal and bacterial genes linked to N-mining from plant litter, we performed a soil meta-transcriptome study in laboratory microcosms. We examined the short-term (3-day) transcriptional response of pine forest

microbial communities in two soil strata to a high dose of N fertilization. **The observed transcriptional response to N deposition was extremely weak, suggesting that direct N repression of microbial functional gene expression is not the principle mechanism for reduced soil respiration immediately after N deposition**.

The second project explores the distribution of N cycle genetic markers across bacterial and archaeal genomes. Attempts to link microbial communities to ecosystem processes regularly use marker genes to represent relevant biochemical pathways. This approach often fails to demonstrate a positive relationship between a gene of interest and the corresponding ecosystem process rate. Inaccurate measurement of marker gene abundance may contribute to failure. In this study, nitrogen metabolism (N-cycling) pathways were mapped in over 6000 complete prokaryotic genomes in the Integrated Microbial Genomes (IMG) database. **The analysis revealed many partial and incomplete N cycling pathways in complete genomes, pointing to a substantial weakness in current marker-gene analyses.**

The third project assesses the potential for nutrient exchanges among microbes on decomposing plant litter. Scanning electron microscopy (SEM) was used to measure the extent of fungal and bacterial colonization of decomposing pine needles. **The observed microbial colonization of pine needles is sparse, with much open space, suggesting ample opportunity for nutrient loss/leaching from decomposing plant litter.** Collectively, this research points towards gaps to address for robust linkage of microbial community composition to ecosystem processes.

Funding Statement: The information presented in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2015SFAF260) to Los Alamos National Laboratory.

Microbial community proteogenomic analyses indicate extensive depth-dependent CO oxidation and C1 metabolism in soil and increased capacity for N_2O reduction with increased rainfall

Spencer Diamond[1]* (sdiamond@berkeley.edu), Zhou Li[2], Peter Andeer[3], Karthik Anantharaman[1], Cristina Butterfield[1], Brian C. Thomas[1], Susannah G. Tringe[3,4], Trent Northen[3], Chongle Pan[2], and **Jillian F. Banfield**[1]

[1]University of California, Berkeley; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [3]Lawrence Berkeley National Laboratory, Berkeley, CA; [4]DOE Joint Genome Institute, Walnut Creek, CA

Global climate change will alter the patterns of distribution of rainfall over the continents, leading to shifts in the rates of production and consumption of greenhouse gases in soil. Microbial metabolism in soil plays a large role in these processes, yet some organisms, C and N currencies and transformation pathways are incompletely accounted for. Previous studies quantified species and/or functional gene abundances, but lack of information linking the organisms present to their metabolic capacities precluded comprehensive identification of biogeochemical processes ongoing in soil. Here, we quantified organism abundance, metabolic potential and in situ function in grassland soil at three depths using genomes reconstructed from 60 deeply sampled soil metagenomes, linked to proteomic analyses. Included in the study were samples from a rainfall manipulation experiment designed to simulate predicted climate change. We detect highly abundant organisms with genomically encoded capacities for C and N turnover whose abundances did not vary with depth and rainfall input. Among these is a novel Euryarchaeote that is inferred to play important roles in methane and/or ammonia oxidation. Many organisms have the capacity to oxidize CO, and the required proteins were highly represented in the proteome. CO is produced by plant roots, and generated through breakdown of heme and other porphyrins in plant and microbial biomass. The abundances of organisms predicted to produce greenhouse gases CO_2 from CO and N_2O from nitric oxide showed a statistically significant increase in abundance between 20 cm and 40 cm soil depths. Increased rainfall caused a statistically significant increase in the abundances of two Sphingobacteria capable of conversion of N_2O to N_2. One of these, represented by a near-complete genome, also has an extensive capacity for complex carbohydrate degradation. Overall, changed water inputs are predicted to affect the capacity for denitrification and N_2O production and the importance of C1 relative to complex carbon metabolism in soil.

This research is supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy Grant DOE-SC10010566.

The "Who" and "How" of Microbial Control over Soil Carbon Dynamics: a Multi –omics, Stable Isotope Probing, and Modeling Approach

Alex Bales[1], Christine L. Bunyon[2], Erin Conlon[1], Serita Frey[2], Kevin Geyer[2], Stuart Grandy[2], Grace Pold[1*] (apold@umass.edu), Seeta Sistla[3], Shana Whitney[2], **Kristen M. DeAngelis**[1]

[1]University of Massachusetts, Amherst MA USA; [2]University of New Hampshire, Durham NH USA; [3]Hampshire College, Amherst MA USA

Project Goals

Our goal is to define the genomic basis of soil microbial carbon use efficiency (CUE) in the context of changing environments, and use this new knowledge to generate more realistic models of SOM decomposition. Our specific objectives are to: (1) Define genomic markers that indicate how terrestrial bacteria and fungi change CUE in response to environmental drivers of microbial activity; (2) Determine how well genomic markers identified in Objective 1 predict CUE in soils; and (3) Improve predictions of soil C dynamics by incorporating genomics-informed CUE estimates into ecosystem C models based on genomic features defined in Objective 1 and validated for soil in Objective 2. To scale from cells to ecosystems, we are studying CUE in pure cultures of bacteria and fungi, mock communities in artificial, organic matter-free soil, and field soil manipulated to elicit microbial community composition and CUE. With each level of complexity, we refine our understanding of the genomic and genetic basis of CUE using statistical and simulation models.

Abstract

Soils are the largest repository of organic carbon (C) in the terrestrial biosphere and represent an important source of CO_2 to the atmosphere. Microbes are key drivers of C cycling and exhibit a large range in C use efficiency (CUE), the fraction of substrate C incorporated into biomass per substrate assimilated. While small changes in CUE can have profound effects on soil CO_2 flux, our understanding of the intersection of genomic and abiotic factors driving soil CUE dynamics is still very limited. This research couples two culture collections of fungi and bacteria isolated from a long-term soil warming experiment at Harvard Forest, where soils have been heated 5^{o}C above ambient for 26 years. We have over a decade of field measurements of C cycling and microbial community analyses using metagenomics and metatranscriptomics. Because microbial feedbacks to the environment are regulated at the organismic level, the foundation of this research is the physiology and genomics of CUE among isolates that we have determined to be either dominant in our soil system, sensitive to climate change factors, or both. Experiments using artificial soil, lab incubations of field-collected soil, and a field manipulation will validate potential omics-informed, novel genomic markers to be used as proxies for CUE in models.

Towards our first goal of defining genomic markers of CUE response to environmental change, we have been defining the physiological response of bacterial and fungal isolates to changes in

substrate and temperature in the lab. The CUE of fifteen soil bacteria covering six phyla is being evaluated under three temperatures and up to four C sources, with a similar study being conducted for fungi. Both bacterial and fungal isolates show diverse temperature responses, with some bacterial isolates showing reduced efficiency at higher temperatures while others showed no change or an increase. An exploratory analysis determined that bacterial genomes enriched in energy generation and metabolism tended to show a decrease in CUE on glucose at higher temperatures, while those relatively depleted in these genes tended to show an increase. This analysis was facilitated by computational tools generated by our group for automating steps in multigene phylogenetic tree production and will be followed up in the upcoming year with the development and calculation of metrics for metabolic complexity correlated with CUE.

For objective 2, we have finalized method development for measuring microbial growth and CUE in soil using a novel, substrate-free approach with ^{18}O-H_2O. This method was found to be more sensitive and interpretable than traditional ^{13}C approaches. Cultivation of select fungal and bacterial isolates and of simplified soil-derived microbial consortia in artificial soil shows a promising route for linking the liquid-culture-based evaluation of genomic markers of CUE sensitivity in natural soils. A qPCR-based approach is being developed to estimate fungal and bacterial biomass and growth in artificial soil, and we have begun to use it to evaluate CUE of bacteria under a range of temperatures and moistures. Expansion of results to natural soil communities and the consortia will enable us to independently validate genomic markers of efficiency identified in isolates.

In pursuit of objective 3, the data derived from the field and microbial laboratory studies will be used to improve model parameterization of terrestrial system responses to environmental change. Incorporating the range of CUE temperature sensitivities observed in bacterial isolates into the litter decomposition model DEMENT confirms that predicted soil C stock response to warming is highly sensitive to this parameter. Continued work to refine the "soil" version of this model with organo-mineral interactions will enable predictions of slower-cycling soil C pools. We will also incorporate the genomics-informed data on microbial CUE combined with field measurements from a long-term warming and warming × nitrogen study into a stoichiometrically-coupled, acclimating microbe-plant-soil model (SCAMPS) that aggregates microbial dynamics at the community-scale and compare these simulations with the guild-based microbial models (DEMENT and MIMICS).

Understanding the genes, metabolic pathways, and phylogenetic signals associated with microbial CUE that drive soil community responses to key global change drivers will provide a stronger basis for modeling soil C dynamics and climate-C feedbacks. Our research will establish a path by which new and extant -omics datasets may be integrated into ecosystem models. These new proxies should create greater power to infer the sensitivity of CUE to environmental dynamics and the tendency of CUE to change and adapt over long times.

Publications related to this research

1. Grace Pold, Erin M. Conlon, Marcel Huntemann, Manoj Pillay, Natalia Mikhailova, Dimitrios Stamatis, T.B.K. Reddy, Chris Daum, Nicole Shapiro, Nikos Kyrpides, Tanja Woyke, **Kristen M. DeAngelis**. (2018) "Genome sequence of *Verrucomicrobium sp*. strain GAS474, a novel bacterium isolated from soil." *ASM Genome Announcements*, in press.

2. Grace Pold, Marcel Huntemann, Manoj Pillay, Natalia Mikhailova, Dimitrios Stamatis, T.B.K. Reddy, Chris Daum, Nicole Shapiro, Nikos Kyrpides, Tanja Woyke, **Kristen M. DeAngelis**. (2018) "Draft genome sequences of three strains of a novel Rhizobiales species isolated from forest soil." *ASM Genome Announcements*, in press.

3. Jerry M Melillo, Serita D Frey, **Kristen M DeAngelis**, William Werner, Michael Bernard, F P Bowles, Grace Pold, Melissa A. Knorr, A Stuart Grandy (2017) "Long-term Pattern and Magnitude of Soil Carbon Feedback to the Climate System in a Warming World." *Science* 358 (6359): 101-105. doi:10.1126/science.aan2874.

Funding statement. *This work is supported by the DOE Genomic Sciences Program under Contract No. DE-SC0016590. Support for this project came from the University of Massachusetts Amherst and the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

Understanding Soil Metabolism from Carbon Release And Carbon Incorporation from Position-Specific ^{13}C-Labeled Substrates

Paul Dijkstra[1,2*] (Paul.Dijkstra@nau.edu), Ayla Martinez[1,2], Xavier Mayali[3], Peter K. Weber[3], Benjamin Koch[1,2], Ember Morrissey[4], Rebecca Mau[1], Michaela Hayer[1], Kirsten Hofmockel[5], Steve Blazewicz[3], Jennifer Pett-Ridge[3], Egbert Schwartz[2], **Bruce Hungate[1,2]**

[1]Center of Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ
[2]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ
[3]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore CA
[4]Davis College of Agriculture, Natural Resources and Design, West Virginia University, Morgantown, WV
[5]Department of, Molecular Ecology Laboratory, Environmental Molecular Sciences Laboratory, Richland, WA

Project Goals

- To better understand microbial population dynamics and its relationship to microbial energy metabolism and C cycling
- To link processes of individual microbial species to whole system C and element cycling and their responses to temperature
- To discover global patterns in microbial population dynamics across ecosystems

Abstract

Our knowledge of microbial biochemistry is mostly obtained in pure culture laboratory experiments, but little is known about the regulation of cell metabolism in real, intact, and complex microbial communities. An improved understanding of metabolic processes under realistic conditions of temperature, moisture and substrate quality and availability has important implications for our ability to predict how soil carbon (C) cycling processes respond to changes in the environment.

We have developed a method to measure the microbial community C use efficiency (CUE) by measuring CO_2 production from individual C-atoms in position-specific ^{13}C-labeled compounds and, based on this, model the flux distribution over the central C metabolic network pathways and determine C use efficiency. However, more information on microbial physiology and ecology can be obtained by additionally studying the position-specific ^{13}C-incorporation into microbial products, such as lipids, DNA, amino acids, or the entire microbial biomass. Here we report on the ^{13}C-incorporation from position-specific labeled glucose and pyruvate into entire microbial cells using nano-SIMS and compare it to measurements of $^{13}CO_2$ release.

We incubated soil from a meadow in the mixed conifer forest near Flagstaff, Arizona at 25 °C for 10 days. At the end of the incubation, we added glucose (1-^{13}C and U-^{13}C) and pyruvate (1-

^{13}C and 2,3-^{13}C) isotopomers in parallel incubations. We measured $^{13}CO_2$ production 0, 20, 40 and 60 minutes after addition of the substrate. Immediately afterwards, we stored the soil at 5°C until fixing of microbial cells with formaldehyde. Cells in the soil were extracted using the Nicodenz method, and captured on 0.2 µm polycarbonate membranes. Isotope incorporation of cells on the filter was then quantified with NanoSIMS analysis. Only a small fraction of the cells exhibited significant isotope incorporation from the short exposure to low concentrations of substrate. The largest percentage of labeled cells, as well as the highest labeling on a per-cell basis, was measured in the treatment with U-^{13}C glucose. Incubations with 1-^{13}C and 2,3-^{13}C pyruvate exhibited intermediate levels of isotope incorporation, while no incorporation was detected in the 1-^{13}C glucose incubation. These results qualitatively correspond to the results from simultaneous $^{13}CO_2$ measurements and modeling that showed that much of atom 1 of glucose and atom 1 of pyruvate was released as CO_2 in the first step of the pentose phosphate pathway and by pyruvate dehydrogenase respectively. These results suggest that it is possible to quantitatively link C losses as CO_2 and C incorporation into microbial biomass to further detail microbial energy and carbon metabolism.

Redox Fluctuations Control Coupled Iron-Carbon Cycling and Microbial Community Structure in Tropical Soils

Authors: Amrita Bhattacharyya[1,2*] (abhattacharyya@lbl.gov), Ashley Campbell,[1] Yang Lin,[3] Malak M Tfaily,[4] Ljiljana Pasa-Tolic,[4] Rosalie Chu,[4] Whendee Silver,[3] Peter K. Weber,[1] Peter Nico,[2] **Jennifer Pett-Ridge[1] (pettridge2@llnl.gov)**

[1]Lawrence Livermore National Laboratory, Livermore CA, [2]Lawrence Berkeley National Laboratory, Berkeley CA, [3]University of California, Berkeley, CA [4]Pacific Northwest National Laboratory, Richland, WA

https://pls.llnl.gov/people/staff-bios/nacs/pettridge-j

Project Goals: **This Early Career research examines the genomic potential and expression 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.**

Wet tropical soils alternate frequently between fully oxygenated and anaerobic conditions, constraining both the metabolism of tropical soil microorganisms and the mineral-organic matter relationships that regulate many aspects of soil C cycling. Tropical forests are predicted to experience a 2–5°C temperature increase and substantial differences in the amount and timing of rainfall in the coming half century. Yet we have a poor understanding of how soil microbial activity and C cycling in these systems will respond to such changes in environmental variability. Using a 44-day redox manipulation and isotope (^{13}C) tracing experiment with soils from the Luquillo Experimental Forest, Puerto Rico, we examined patterns of tropical soil microorganisms, metabolites and soil chemistry in soils 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 assess how changes in redox oscillation frequency altered microbial community structure and activity, organic matter turnover and fate, and soil chemistry.

Microbial community structure was strongly affected by patterns of O_2 availability in soils incubated under different redox oscillation treatments. Communities from static anoxic soils became enriched in many Proteobacteria and Firmicutes taxa relative to the initial community, while soils that experienced either fluctuating or constant O_2 exposure retained a similar community composition to the native soil (measured at the start of the experiment). At the same time, the concentration and molecular composition (measured by FTICR-MS) of DOC shifted, corresponding to O_2 availability. DOC and Fe^{2+} concentrations were positively correlated for all four redox treatments and increased within minutes following a switch from oxic to anoxic conditions. Prolonged anoxia led to the reductive dissolution of Fe oxides, and a corresponding increase in DOC availability. Several members of the microbial community belonging to

Proteobacteria, Acidobacteria, and Firmicutes that are capable of iron oxidation, reduction (or both) were significantly enriched under anoxic conditions, suggesting the observed changes in Fe cycling were at least partly driven by microbial activity. However, in soils where redox conditions oscillated, the overall crystallinity of iron oxides increased.

We used molecular scale spectromicroscopy and secondary ion mass spectrometry (SEM/STXM/nanoSIMS) high-resolution imaging to trace the fate of added ^{13}C plant litter and compared the chemical composition of the pure ^{13}C-litter vs ^{13}C-organic matter found in the soils after incubation under oxic, anoxic, and oscillating conditions. Overall, the static oxic soils exhibited the highest gross soil respiration (CO_2 flux), however, ^{13}C-litter derived respiration was highest in static anoxic soils, suggesting decomposition of pre-existing SOM was O_2-limited in the anoxic soils. Chemical alterations of the added ^{13}C organic matter (relative to the added substrate) were dependent on redox treatment. In general, anoxic soils tended to accumulate aromatic compounds compared to the oxic soils.

Taken together, these data illustrate how microbial community activity (i.e. oxidation of carbon and reduction of iron or other acceptors) and Fe-C coupling may control redox-driven biogeochemistry in humid tropical soils. These results, along with parallel studies of biogeochemical responses (pH, P availability), suggest a highly responsive microbial and geochemical system, where the frequency of low-redox events controls exchanges of C between mineral-sorbed and aqueous pools. Our findings highlight the need for a more explicit representation of soil redox dynamics in our understanding of C cycling in dynamic tropical forest ecosystems.

This research was supported by an Early Career Research Program award to J. Pett-Ridge, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1478. Work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231.

Modeling the Pathways to Carbon Liberation: Microbiologically-Informed C-Cycle Modeling in a Thawing Permafrost Landscape, for Scaling from Molecules to the Earth System

Eoin Brodie[1], **William Riley**[1], Eric King[1], Kuangyu Chang[1,2], Jia Deng[2], **Steve Frolking**[2], Tim Morin[3], Rick Wehr[4], **Scott Saleska**[4], **Virginia Rich**[5], IsoGenie Project Team[1-6]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; [2]University of New Hampshire, Durham, NH; [3]State University of New York, Syracuse, NY; [4]University of Arizona, Tuscon, AZ; [5]Ohio State University, Columbus, OH; [6]Multiple Institutions in USA, Australia, Sweden.

https://isogenie.osu.edu/

Project goals: Understanding the fate of C in thawing permafrost is a grand challenge in both the biological and earth sciences because of its importance to biogeochemistry and climate change. Permafrost C pools are large (~1700 PgC), and C dynamics of permafrost thaw are complex: old C decomposes to carbon dioxide (CO_2) and methane (CH_4) as it is mobilized from thawing permafrost peat, while new C accumulates as thaw initiates ecological succession in both plant and associated microbial communities. The IsoGenie project is working to combine measurements and modeling to improve understanding of historical C dynamics and prediction over the 21st century. Here we present results from four models participating in the project.

BioCrunch, a genome-informed trait-based model, emerged as part of continued improvements made to the reactive transport solver, Crunchflow. Its purpose is to explicitly represent microbes and microbial functions using 'omics derived information. BioCrunch has been further extended to account for kinetic isotope fractionation in order to simulate how microbial feedbacks alter the isotopic signature of methane and other C species. Simulations allow for a faithful representation of the functional diversity of microbial populations, how microbial physiological traits impact fitness, how biogeochemical processes are impacted by emerging microbial composition, and how biogeochemistry feeds back to alter microbial fitness and community assembly. PRT (Peatland Reactive Transport) is a model of peat biochemistry and gas transport, whose purpose is to use observations of methane and CO_2 isotopes in the pore water to infer how carbon gas transport, production, and consumption rates vary with depth, allowing direct comparison to *in situ* 'omic samples of microbial composition, potential, or activity. The DNDC biogeochemistry model was modified for IsoGenie to include a more detailed representation of acetotrophic and hydrogenotrophic methanogenesis, and associated ^{13}C isotopic signatures. Its purpose is to provide new means for testing model representations of these processes against observations of isotopic fluxes and methanogen community composition (acetoclasts vs hydrogenotrophs). The *ecosys* model is a mechanistic terrestrial ecosystem simulator that has been applied in many high-latitude studies. The model represents soil biogeochemistry with explicit microbial details, plant processes, and landscape-scale thermal and hydrological processes. Its purpose is to study (a) geomorphological dynamics and their effects on plant responses, (b) hydrological controls on site-level CO_2 and CH_4 emissions, and (c) landscape-scale soil biogeochemical and plant responses to expected climate changes.

These four modeling approaches provide a diverse toolkit for investigating, testing, and predicting the role of microbial communities in permafrost-thaw associated carbon dynamics.

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-SC0004632, DE-SC0010580, DE-SC0016440. Work at Lawrence Berkeley National Laboratory was performed under the auspices of the U.S. Department of Energy Contract No. DE-AC02-05CH11231.

Subarctic Lake Sediment Microbial Community Contributions to Methane Emission Patterns

Joanne B. Emerson[1,2*] (jbemerson@ucdavis.edu), **Ruth K. Varner**[3], Donovan H. Parks[4], Martin Wik[5], Rebecca B. Neumann[6], Joel E. Johnson[3], Caitlin M. Singleton[4], Ben J. Woodcroft[4], Rodney Tollerson II[1], Akosua Owusu-Dommey[7], Morgan Binder[7], Nancy L. Freitas[7], **Patrick M. Crill**[5], **Scott R. Saleska**[7], **Gene W. Tyson**[4], and **Virginia Rich**[1]

[1]The Ohio State University, Columbus; [2]Current affiliation: University of California, Davis; [3]University of New Hampshire, Durham; [4]University of Queensland, Australia; [5]Stockholm University, Sweden; [6]University of Washington, Seattle; and [7]University of Arizona, Tucson

http://isogenie.osu.edu

Project Goals: Our objective is to discover how microbial communities mediate the fate of carbon in thawing permafrost landscapes under climate change. Our team is engaged in a systems approach integrating molecular microbial and viral ecology, organic chemistry and stable and radiocarbon isotopes, and state-of-the-art modeling along an interconnected chronosequence of thawing permafrost and post-glacial lakes in subarctic Sweden.

Abstract: Post-glacial lakes in northern landscapes have been identified as a significant source of methane to the atmosphere, largely through ebullition (bubbling) of microbially produced methane from the sediments. Ebullitive methane flux has previously been shown to correlate significantly with sediment surface temperatures in these lakes, suggesting that solar radiation is the primary driver of methane emission. However, we show here that the *slope* of this relationship (*i.e.*, the extent to which increasing temperature increases ebullitive methane emissions) differs spatially, both within and among lakes. As microbes are responsible for both methane generation and removal in lakes, we hypothesized that microbial communities—previously uncharacterized in post-glacial lake sediments—could be contributing to spatiotemporal differences in methane emission responses to temperature. We compared methane emission data with sediment microbial (metagenomic and 16S rRNA gene sequencing), isotopic, and geochemical characterizations across two post-glacial lakes in Northern Sweden. With increasing temperatures, the increase in methane emissions was higher in lake middles than lake edges, consistent with higher abundances of methanogens in sediments from lake middles than edges. Using partial least squares statistical regressions, microbial abundances (including the abundances of methane-cycling microorganisms and of reconstructed population genomes) were better predictors of porewater methane concentrations than abiotic variables. These results suggest that microbial communities contribute to the rate and magnitude of the temperature response of methane emissions in subarctic post-glacial lakes.

Funding statement. This research was funded by the U.S. Department of Energy Office of Biological and Environmental Research under the Genomic Science program, Awards DE-SC0010580 and DE-SC0016440, in addition to a U.S. National Science Foundation REU program: Northern Ecosystems Research for Undergraduates (EAR#1063037).

Methanogenesis in Oxygenated Soils is a Substantial Fraction of Wetland Methane Emissions

Jordan Angle[1], Timothy Morin[2,3], Lindsey Solden[1], Adrienne Narrowe[4], Garrett Smith[1], Mikayla Borton[1,3], Camilo Rey-Sanchez[2,3], Rebecca Daly[1], David Hoyt[5], William Riley[6], Christopher Miller[4], Gil Bohrer[2,3], **Kelly Wrighton[1,3*]** (kwrighton@gmail.com)

[1] Department of Microbiology, The Ohio State University, Columbus, OH 43210 USA; [2] Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH 43210 USA; [3] Environmental Science Graduate Program, The Ohio State University, Columbus, OH 43210 USA; [4] Department of Integrative Biology, University of Colorado Denver, Denver, CO 80217; [5] Pacific Northwest National Laboratory, Richland, WA 99352 US; [6] Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Project Goals:

Despite their relatively small land coverage, wetlands represent the largest source of atmospheric CH_4 (20-40%). However, variations in these wetland emission budgets are high, with over 25% uncertainty. Accurately predicting net CH_4 fluxes from these systems depends on multiple interacting geochemical, ecological, and metabolic constraints that are poorly understood, oversimplified, or missing in global biogeochemical models[12-14]. ***The overarching objective of my early career proposal is to identify the biogeochemical and genomic determinants impacting methane production, and the scale and physical distribution over which they operate, along freshwater wetland gradients***. Using field investigations, methanogen types, activity, and responses to geochemical conditions will be determined along seasonal and spatial gradients (**Objective 1**). Using laboratory microcosms, the formation of anoxic microsites and their capacity to facilitate methanogenic activity in wetland soils will be simulated (**Objective 2**). Lastly, these field and laboratory data will be used for multi-scale, process-level evaluation of an ecosystem biogeochemical model (the upcoming CLM-coupled version of *ecosys*) that accommodates these newly identified processes and parameterizes representation of these processes along relevant environmental gradients (**Objective 3**).

Abstract text:

The current paradigm, widely incorporated in soil biogeochemical models, is that microbial methanogenesis can only occur in anoxic habitats. Recent reports present an alternative view that in some ecosystems methanogenesis also occurs in oxic soils, lakes, and marine systems, a concept known as the methane paradox. Here we use porewater and greenhouse-gas flux measurements to provide clear evidence for methane production in oxygenated soils from a freshwater wetland. A comparison of oxic to anoxic soils revealed up to ten times greater methane production and nine times more methanogenesis activity in oxygenated soils. Metagenomic and metatranscriptomic sequencing recovered the first near complete genomes for a novel methanogen species, and showed acetoclastic production from this organism was the dominant methanogenesis pathway in oxygenated soils. This organism, *Candidatus* Methanothrix paradoxon, is prevalent across methane emitting ecosystems, suggesting a global significance. Moreover, in this wetland, we estimated that up to 80% of methane fluxes could be attributed to methanogenesis in oxygenated soils. Together our findings challenge a widely-held

assumption about methanogenesis, with significant ramifications for global methane estimates and Earth system modeling.

Funding statement:

This material is based upon work supported by Early Career Award from the U.S. Department of Energy to K.C.W., Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018022. W.J.R. was supported by the Director, Office of Science, Office of Biological and Environmental Research of the US Department of Energy under Contract No. DE-AC02-05CH11231 as part of their Regional and Global Climate Modeling program. Graduate student support was provided by the National Sciences foundation via Graduate Research Fellowship (to G.J.S.) and a Doctoral Dissertation Improvement Grant (to T.H.M.), and by the Department of Energy Office of Science Graduate Research Program, Solicitation 2 (to T.H.M.). Porewater metabolites (NMR) were performed using EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research supported by DOE contract No. DE-AC05-76RL01830. DNA and RNA sequencing was conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility that is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. Flux observations were conducted with support from the US Department of Energy's Office of Science, Ameriflux Management project.

Understanding the microbial controls on biogeochemical cycles in permafrost ecosystems

Neslihan Taş

EESA, Lawrence Berkeley National Laboratory

Project Goals: Permafrost soils contain a broad diversity of cold-adapted microbes, whose metabolic activity depends on environmental factors such as temperature changes that cause cycles of freezing and thawing in the soil. Microbial metabolism leads to decomposition of soil organic matter, substantially impacting the cycling of nutrients and significantly affecting the arctic landscape. However, the relationship between permafrost microbial properties and biogeochemical cycles is poorly understood. This project will use state of the art molecular techniques to resolve complex microbial processes governing the biogeochemical cycles in arctic soils and permafrost to better inform efforts to access uncertainties surrounding ecosystem responses.

Permafrost soils are one of the world's largest terrestrial carbon reservoirs thus an important focal point for climate change research. With increasing global temperatures, arctic landscapes are changing and becoming a potential source of greenhouse gas (GHG) emissions. 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. This project will use field experiments, 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.

The importance of biocrust organics revealed by metabolite sorption studies and *in situ* imagery

Estelle Couradeau[1]* (ecouradeau@lbl.gov), Tami L. Swenson[1], Benjamin P. Bowen[1,,2], Vincent Felde[3], Joelle Schlapfer[1], Daniel Uteau[3], Stephen Peth[3], Dula Parkinson[1] , Ferran Garcia-Pichel[4] and **Trent Northen**[1,2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA ; [2]DOE Joint Genome Institute, Walnut Creek, CA; [3]Universität Kassel, Kassel, Germany ; [4]Arizona State University, Tempe, AZ ;

http://www.northenlab.org/research/environmental-exometabolomics/

Project Goals: Understanding the role of microorganisms in soil nutrient cycling and organic matter turnover is a priority of the Department of Energy missions. Within this framework, our project aims at interrogating *in situ* microbial activity through the combination of pioneering approaches that link microbial community structure to soil organic matter dynamics and physical properties.

Abstract:

Weighing approximately 2400 Gt, soil organic carbon is by far the largest terrestrial pool of organic carbon. Small changes in soil organic matter (SOM) content could have drastic effects on the overall carbon cycle, therefore making SOM stabilization one of the most promising avenues to mitigate greenhouse gas emissions. The soil microbiome is critical in these terrestrial processes including the synthesis, turn-over and eventually the stabilization of SOM. However we still have a limited understanding of how communities encompassing thousands of members process carbon *in-situ* and how SOM might be stabilized through biological and abiological interactions.

A major challenge moving forward is being able to resolve microbial processes at the relevant spatial and temporal scales to decipher their role in SOM turnover. We use biological soil crust (BSC) as a model system to interrogate the dynamics and interactions of SOM and microbial communities. BSCs are the millimeter size epidermal layer of arid lands, which cover around 20% of Earth's continental area, a global extension that is predicted to increase due to desertification. BSCs consist of an assemblage of mineral soil particles consolidated into a crust by the addition of exopolymeric substances (EPS), forming the exopolymeric matrix (EPM), which is mainly produced by the filamentous bundle forming cyanobacterium, *Microcoleus sp*. This cyanobacterium is both the primary producer for, and architect of BSCs, sustaining the development of a diverse microbial community. Despite their ecological importance, little is known about how BSC communities endure long dry periods while remaining viable for rapid resuscitation upon wetting. We hypothesized that biocrusts have the ability to retain nutrients by the EPM and that the bundle-forming is a trait of *Microcoleus sp.* that serves as an adaptive advantage, slowing the dehydration process and allowing the cyanobacterium to prepare for desiccation.

First, we investigated the ability of biocrust, particularly the EPM, to retain nutrients which would be a critical advantage during limited hydration. A 13C-labeled bacterial lysate metabolite mixture was incubated with non-biologically active (autoclaved) biocrust, the underlying subcrust or EPS extracted from biocrust. Metabolite sorption (retention) was monitored using liquid chromatography/ mass spectrometry. Our data demonstrate that both biocrust and its EPS (or more broadly, the EPM) sorbed more metabolites than the subcrust, especially amino acids and organic acids, highlighting the potential role of biocrust in nutrient retention during a wetting event (Swenson et al, 2017).

To glean further insight into how biocrusts function during limited rain events, we used synchrotron based X-ray microtomography to track the water dynamics within a undisturbed BSC core throughout a hydration-dehydration cycle. We resolved the distribution of air, water, soil particles and *Microcoleus sp.* bundles at the microscale to quantify changes of the pore architecture during wetting and drying. We confirmed that *Microcoleus sp.* bundles hold water through desiccation, as their volume keeps shrinking even after the soil pore space dried completely.

Altogether, these results suggest that the EPM is crucial for BSC adaptation to pulsatile activity and they retain water longer, allowing more time for the community to prepare for desiccation. Biocrusts may act as a passive filter, capturing a variety of metabolites that could be further used as nutrient source upon the next wetup event. These data also clearly highlight that the use of other *in situ* techniques might bring complementary information and enhance interpretations of genomics-based microbial community analyses.

References

Swenson, T.L., Couradeau, E., Bowen, B.P. , De Philippis, R., Rossi, F., Mugnai, G. & Northen, T.R. A novel method to evaluate nutrient retention by biological soil crust exopolymeric matrix. *Plant Soil* (2017). https://doi.org/10.1007/s11104-017-3537-x.

This work conducted was supported by the Office of Science Early Career Research Program, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contracts No. DE-AC02-05CH11231.

Microbial Communities As Carbon Conductors: Elucidating Universal Traits Across Litter Types

Renee Johansen[1] (rjohansen@lanl.gov), Michaeline Albright[1], Deanna Lopez[1], La Verne Gallegos-Graves[1], Cheryl Kuske, Andreas Runde[1], Thomas Yoshida[1], John Dunbar[1]

[1]Los Alamos National Laboratory, Los Alamos, New Mexico, 87501

http://www.lanl.gov/org/padste/adcles/bioscience/bioenergy-biome-sciences/soil-carbon.php

Project goal:

Determine *common* microbial community traits driving extreme patterns of soil carbon flow across different substrates

Abstract

The LANL Genomic Science 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.

We aim to identify microbial traits that consistently predict patterns of carbon flow (soil storage versus mineralization) from decomposition of different litter types. *This research addresses BER grand challenge 2.4, by linking genotype to phenotype in communities that interact to cycle carbon in terrestrial ecosystems.* The proportion of photosynthetically fixed carbon stored in soils over long time periods, and not returned to the atmosphere via respiration, is a significant moderator of atmospheric carbon levels and global climate. The balance between soil carbon storage and mineralization is mediated by micro-organisms that transform plant litter and exudates during decomposition. A relationship between carbon fate in soils and microbial community composition has been regularly postulated and indirectly implied but empirical evidence directly linking the two is lacking. Furthermore, community traits driving major changes in carbon fate are currently unknown. Plant litter quality is a key variable influencing microbial community function. Consequently, comparison of microbial community traits linked to carbon flow from different types of plant litter allows differentiation of universal traits from litter-specific traits. Universal traits are the most useful for carbon flow modelling and management.

We found carbon flux patterns vary substantially when decomposition is undertaken by different naturally occurring microbial communities, irrespective of substrate. To demonstrate this, we placed the same natural communities extracted from soil samples collected from the drylands region of the United States on pine, oak and grass litter in sand microcosms. The microcosms were incubated in homogeneous conditions and cumulative carbon outputs were compared after 45 days. Carbon fate varied widely on all substrates despite consistent conditions within the microcosms. Although grass is generally considered more labile than pine or oak, grass

decomposition produced less CO_2 on average than pine or oak. Dissolved organic carbon outputs were similar among the litter types. Litter type impacted community performance, with the same communities generating different carbon output patterns on the three litters. **Bacteria, not fungi, were the strongest drivers of divergent functional outcomes on pine with both composition and diversity being important.** Comparison of community traits driving carbon fate across the different litter types is ongoing. Elucidating robust traits under controlled conditions is a critical first step toward steering functional variation in naturally chaotic environments.

Funding Statement: The information presented in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2015SFAF260) to Los Alamos National Laboratory.

Time-series metagenomics of experimentally warmed Alaskan tundra and Oklahoma temperate soils enables fine-resolution assessment of belowground C cycling feedbacks to climate change

Eric R. Johnston (ejohnston7@gatech.edu)[1*], C. Ryan Penton[2], Julian Yu[2], **Liyou Wu**[3], Shi Zhou[3], Kai Xue[3], **Zhili He**[3], Mengting Maggie Yuan[3], **Yiqi Luo**[3], **Edward A.G. Schuur**[4], James R. Cole[5], **James M. Tiedje**[5], **Jizhong Zhou**[3], and **Konstantinos T. Konstantinidis**[1] (kostas@ce.gatech.edu).

[1]Georgia Institute of Technology, Atlanta, GA 30332, USA; [2]Arizona State University, Phoenix, AZ 85004, USA; [3]University of Oklahoma, Norman, OK 73019, USA; [4]Northern Arizona University, Flagstaff, AZ 86001, USA; [5]Michigan State University, East Lansing, MI 48824, USA

http://www.ou.edu/ieg.html

Project goals: **The overall goal of this project is to advance systems-level predictive understanding of the feedbacks of belowground microbial communities to multiple climate change factors and their impacts on soil carbon (C) cycling processes. The specific objectives are to: (i) reveal the responses of soil microbial communities to climate warming and soil moisture alteration in both tundra and temperate grassland ecosystems; (ii) determine temperature sensitivity of recalcitrant C decomposition and characteristics of the microbial degraders; and (iii) develop integrated bioinformatics and modeling approaches to scale information across different organizational levels**.

Abstract: Soils contain more carbon (C) in the form of soil organic matter (SOM) than both aboveground plant and atmospheric pools combined. Higher land temperatures are expected to cause the release of considerable amounts of CO_2 and CH_4 to the atmosphere, primarily through the stimulation of microbial-mediated SOM turnover. However, the direction, magnitude, and underlying basis of soil feedbacks to climate warming remain poorly understood. To this end, we have investigated microbial communities from Alaskan tundra (AK) and Oklahoma temperate grassland (OK) soils, both of which have been experimentally warmed *in-situ* (~2°C above ambient temperature) and under laboratory conditions (15 and 25°C). By combining well-replicated soil metagenomes with continuous environmental monitoring, respiration data, and soil measurements, we hope to gain an improved understanding of microbial responses to climate warming, particularly those involved in the release/sequestration of greenhouse gases.

At the AK field site, communities from deeper soil layers (45-55cm depth) were more sensitive to 5 years of field warming than surface soils (15-25cm depth) - e.g., for deep-layer communities, warming induced a significant increase to α-diversity, and community composition was largely relatable to changes in annual thaw duration, which increased by ~34 days due to warming (warming only increased surface soil thaw duration by ~8 days). Warming also increased the abundance of many SOM catabolic pathways, including those for both the labile and recalcitrant fractions of SOM. These results were also consistent with GeoChip functional gene analysis of increased ecosystem respiration reported at an earlier experimental phase (Xue et al., 2016). Sequence assembly and binning techniques allowed for the recovery of several near-complete bacterial population genomes from both AK and OK ecosystems, allowing for prediction of their metabolic lifestyles, regional prevalence, and response to elevated temperatures. Several of the recovered AK populations were regionally ubiquitous, e.g., found at several locations ~100-530 kilometers apart (Johnston et al., 2016). Consistent with the community-wide shifts mentioned above, warming favored bacterial populations encoding diverse metabolisms for recalcitrant and

labile SOM degradation, including abundant members of the community (0.25-2% of total). 5 years of similar experimental warming at the OK field site altered the functional composition of microbial communities (β-diversity distances) and also increased microbial community α-diversity. To assess shifts under more pronounced temperature changes and test for similarities between AK and OK soils, soils were incubated in the lab under 15 and 25°C. Metagenomic sequencing and assembly allowed for the recovery of several hundred-population genomes, which collectively recruited up to 75% of metagenomic reads. This allowed for more resolved associations between SOM-turnover and community composition (i.e., members responsible for these activities) to be identified. For instance, a correlation coefficient of 0.83 was obtained by relating the abundance of *Acidobacteria* populations to the recalcitrant fraction of SOM respired under laboratory conditions.

References

[1] E. R. Johnston, L. M. Rodríguez-R, C. Luo, M. M. Yuan, L. Wu, Z. He, E. A. Schuur, Y. Luo, J. M. Tiedje, J. Zhou, and K. T. Konstantinidis. (2016). Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska tundra ecosystem. Frontiers in Microbiology. 7:579 doi:10.3389/fmicb.2016.00579

[2] K. Xue, M. M. Yuan, Z. J. Shi, Y. Qin, Y. Deng, L. Cheng, L. Wu, Z. He, J. D. Van Nostrand, R. Bracho, S. Natali, E. A. G. Schuur, C. Luo, K. T. Konstantinidis, Q. Wang, J. R. Cole, J. M. Tiedje, Y. Luo, and J. Zhou. (2016). Rapid microbial feedbacks reveal vulnerability of tundra soil carbon to climate warming. Nature Climate Change. 6:595-600 doi:10.1038/nclimate2940

[3] C. R. Penton, C. Yang, L. Wu, Q. Wang, J. Zhang, F. Liu, Y. Qin, Y. Deng, C. L. Hemme, T. Zheng, E. A. G. Schuur, J. M. Tiedje, and J. Zhou. (2016). NifH-Harboring Bacterial Community Composition across an Alaskan Permafrost Thaw Gradient. Frontiers in Microbiology. 7:1894. doi:10.3389/fmicb.2016.01894

Funding statement: Our work is supported by the U.S. DOE Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program, Award No. DE-SC0010715.

m-CAFÉs: Microbial Community Analysis and Functional Evaluation in Soils

N. Louise Glass*[1](nlglass@lbl.gov), Peter F. Andeer[1], Jens Heller[1], Peter Kim[2], Trenton Owens[1], Joelle Schlapfer[1], Jill Banfield[3], Eoin L. Brodie[1], Adam M. Deutschbauer[1], Jennifer A. Doudna[3], Mary K. Firestone[3], Peter S. Nico[1], Anup K. Singh[2], Rodolphe Barrangou[4], Trent R. Northen[1]

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]Sandia National Laboratory, Livermore; [3]University of California, Berkeley; [4]North Carolina State University, Raleigh

Project Goals: To derive mechanistic understandings of plant-microbe-soil interactions using reproducible, simplified ecosystems

The m-CAFÉs is a collaborative, coordinated and integrated, mission-driven program to interrogate the function of soil microbiomes with critical implications for carbon cycling and sequestration, nutrient availability and plant productivity in natural and managed ecosystems. The project targets molecular mechanisms governing carbon and nutrient transformation in soil, with a focus on microbial metabolic networks. We are developing precisely controlled ecosystem fabrications (EcoFABs) that reflect key functional attributes of plant-microbe interactions within soil, focusing on progressively increasing complexity that honor the physical, chemical and biological properties of soils. To interrogate microbial function in EcoFABs, we are pioneering CRISPR-Cas and environmental RNAi technologies to systematically determine functions of bacteria and fungi in the rhizosphere/soil. These approaches address the grand challenge determining the molecular basis of soil microbiome metabolism that governs the stabilization of carbon in soil. The m-CAFÉs team will test hypotheses about microbial functional roles within soil interaction networks in an iterative manner, by observing, recapitulating and manipulating rhizosphere/soil microbiomes to establish robust predictive models of microbial community function in soils. The results of these studies will be transformative for our understanding of soil metabolism and microbiome science, with applications to DOE missions in energy and environment.

Funding statement.

This material by m-CAFÉs 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

Characterization of Microorganisms Resistant to Multiple Metals from the Contaminated Environment at the Oak Ridge Reservation

M.P. Thorgersen[1]*(mthorger@uga.edu), X. Ge[1], F.L. Poole[1], G.M. Zane[2], E.M. Forsberg[3], A. Deutschbauer[4], R. Chakraborty[4], T.C. Hazen[5], A. Mukhopadhyay[4], G. Siuzdak[3], J.D. Wall[2], J. Chandonia[4], P. Novichkov[4], M.W.W. Adams[1], **A.P. Arkin[4]** and **P.D. Adams[4]**

[1]University of Georgia, Athens GA; [2]University of Missouri, Columbia Mo; [3]Scripps Research Institute, La Jolla CA; [4]Lawrence Berkeley National Laboratory, Berkeley CA; [5]Oak Ridge National Laboratory, Oak Ridge TN

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. The goal of this project is to isolate and characterize microbial strains that are resistant to many different metals at concentrations found in contaminated wells at the Oak Ridge Reservation environment. We will use diverse techniques such as transposon sequencing and global isotope metabalomics to investigate multi-metal resistance in these strains. In addition, the large amounts of characterization data gathered will be aggregated in KBase, where new KBase tools will be used to organize, analyze and distribute results to the public.

Abstract: The metal resistance campaign of ENIGMA is focused on investigating molecular mechanisms of microbial metal resistance. One of the defining characteristics of the Oak Ridge Reservation (ORR) environment is the presence of nitrate and mixed metal industrial waste with concentrations many metals over 1,000 times elevated in contaminated areas compared to pristine groundwater. Several metal centric high throughput anaerobic enrichments from contaminated ORR groundwater and sediment have been conducted including those using media containing a suite of metals at concentrations based on the ORR environment. These isolates are all resistant to metals such as uranium, aluminum, manganese and nickel ($\geq$ 100 μM) when grown under nitrate-reducing growth conditions. After high-throughput preliminary screening, isolates that have unusual metal resistance properties indicative of novel mechanisms will be selected for in depth characterization by an array of lab processes and technologies as part of the metal resistance campaign pipeline including; genome sequencing, screening for genetic tractability, random barcode TN-Seq, DNA affinity purification sequencing, and global isotopic metabolomics. All characterization data can then be uploaded to KBase where new tools developed to analyze generic data sets are being used to organize, analyze, and eventually distribute results to the public.

Funding statement: ENIGMA (http://enigma.lbl.gov) at LBNL supported by Office of Biological and Environmental Research US Dept of Energy Contract No: DE-AC02-05CH1123

Coupling of Field- and Lab-based Experiments to resolve controls of Nitrate Respiration Pathway Partitioning at the Oak Ridge Shallow Aquifer

Frederick von Netzer[1]* (fcrvn@uw.edu), Kristopher A. Hunt[1], Jacob Valenzula[2], Annie Otwell[2], Serdar Turkarslan[2], Nitin S. Baliga[2], Ji-Won Moon[3], Kenneth Lowe[3], Miguel Rodriguez[3], Dwayne Elias[3], Dominique C. Joyner[3,4], Charles Paradis[3,4], Susan Pfiffner[4], Dan Williams[4], Katie Fitzgerald[3,4], Sheridan Brewer[4], Benjamin Adams[3,4],Terry C. Hazen[3,4], Erica. L.-W. Majumder[5], Grant M. Zane[5], Judy D. Wall[5], Daliang Ning[6], Jizhong Zhou[6], Michael T. Thorgensen[7], Xiaoxuan Ge[7], Michael W.W. Adams[7], Lauren Lui[8], Romy Chakraborty[8], Hans Carlson[8], Adam Deutschbauer[8], David Vuono[10], Kelley Meinhardt[1], David A. Stahl[1], **Adam P. Arkin**[8,9] and **Paul D. Adams**[8]

[1]University of Washington, Seattle; [2]Institute for Systems Biology, Seattle; [3]Oak Ridge National Laboratory, Oak Ridge; [4]University of Tennessee, Knoxville; [5]University of Missouri, Columbia; [6]University of Oklahoma, Norman; [7]University of Georgia, Athens; [8]**Lawrence Berkeley National Laboratory, Berkeley**; [9]University of California, Berkeley and [10]Desert Research Institute, Reno

http://enigma.lbl.gov

Project Goals: Surveying the fate of nitrate at the Oak Ridge site through field assays and experimental lab systems for understanding the drivers of nitrate respiration pathway partitioning

The Oak Ridge site has a long history of research on biogeochemical impacts of contamination. Past studies largely focused on the fate and transport of radioisotopes, heavy metals, halogenated organic compounds, and mercury. Less studied is the consequence of the disposal of heavy metal laden nitric acid reaching nitrate concentrations of up to 2 g/l, acidifying large regions of this shallow aquifer. Thus, the combined impact of nitrate and lowered pH on microbial respiration and associated processes in this oxygen-limited environment is not well constrained in the literature. Microbially-driven nitrate respiration can operate through different pathways, resulting in either removal of nitrate through gaseous metabolites or retention through dissimilatory nitrate reduction to ammonium (DNRA). In this project, we focus on biotic and abiotic controls of nitrate transformation through field assays and experiments in lab systems.

A grand challenge is to understand the function of microbial communities in the field within the limitations of available sampling methods. Therefore, we use a two-tier approach by combining field data and model laboratory systems. In the field, we survey the potential for nitrate respiration through the acetylene block method and nitrate isotope fractionation. For experiments with model isolates and communities in the lab, we are characterizing different reactor system formats but since biology in the subsurface is a combination of sediment-associated biofilms and planktonic organisms in pore water, we here focus on fluidized bed reactor (FBR) technology. Fluidization of sediment selects for both planktonic and attached populations, which can be challeng-

ing in planktonic chemostats, thus better emulating subsurface conditions while avoiding physical heterogeneities, such as channeling, that develop in packed-bed reactors.

Abiotic characterization of FBR fluidization has shown that particles of less than 100 μm in diameter have impractically low fluidization velocities while particles larger than 300 μm in diameter require excessive flow rates to fluidize. We have established controlled fluidization under both oxic and anoxic conditions, and are evaluating the influence of medium composition, substratum size, composition as well as mass per reactor volume on biomass accumulation and activity. FBR operating conditions are initially refined with monocultures of *Desulfovibrio vulgaris* to select for both attached and planktonic populations or primarily attached populations. These studies are also evaluate the influence of physical parameters (particle size/composition, shear stress, and surface area) on colonization and factors controlling the partitioning between planktonic and attached organisms, a common issue in the field. These data inform the range of operating conditions needed to develop more complex reactor communities, now being evaluated by reactor colonization by *Desulfovibrio* spp. syntrophically coupled with different hydrogenotrophic methanogens.

A simple acetylene block test serves to measure active regions of nitrate respiration in the field, identifying active source material for the FBR experiments. Using a model mixture of carbon sources to stimulate nitrate respiration with either biomass from filtered groundwater or sediment in native sterile groundwater, we trace the formation of nitrous oxide and ammonium over relatively short incubation periods in different incubations setups. Those studies are complemented by stable isotope fractionation data to constrain abiotic and biotic sinks/sources of nitrate and its transformation products, including nitrous oxide. Initial acetylene block and nitrate isotope fractionation data are consistent with a significant abiotic source of nitrous oxide in the highly contaminated area EB106 while biotic processes dominate at lower nitrate concentrations in both groundwater and sediment fractions in a region of lower contamination (EB271). DNRA and denitrification were significant processes at EB271, with DNRA more prominent in the vadose zone and denitrification in the transition zone between capillary fringe and saturated zone.

The field data will guide FBR design and operation, as needed to identify and quantify variables governing microbial community dynamics, such as activity, resilience, and persistence as they relate to different respiratory processes at the Oak Ridge site. While the FBR reactor configuration may better emulate the subsurface environment through retention of both planktonic and attached microbial populations than standard liquid cultures, we anticipate that a more fully predictive understanding of variables controlling field site processes will derive from comparative studies of different reactor formats, including ongoing complementary studies of both field chemostats and packed-bed columns. Another connection to the field is the use isotopic fractionation signatures, as well as the metabolic and thermodynamic modeling of reactors operated under field relevant conditions.

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

Ecological Stochasticity in Subsurface Microbial Community Assembly under Stress Gradient: Application of A General Quantitative Framework

D. Ning[1]* (ningdaliang@ou.edu), Y. Deng,[1] J.D. Van Nostrand,[1] L. Wu,[1] P. Zhang,[1] Z. He,[1] Y. Fu,[1] D.J. Curtis,[1] Y. Li,[1] Y. Fan,[1] M.B. Smith,[2] A.M. Rocha,[3] C.S. Smillie,[2] S.W. Olesen,[2] C.J. Paradis,[4] J.H. Campbell,[3,12] J.L. Fortney,[4] T.L. Mehlhorn,[3] K.A. Lowe,[3] J.E. Earles,[3] J. Phillips,[3] S.M. Techtmann,[4] D.C. Joyner,[4] K.L. Bailey,[3] R.A. Hurt Jr.,[3] S.P. Preheim,[2] M.C. Sanders,[2] J. Yang,[2] M.A. Mueller,[3] W.A. Lancaster,[5] B.J. Vaccaro,[5] F.L. Poole II,[5] S. Brooks,[3] D.B. Watson,[3] A. Aaring,[6] B. Adams,[4] S. Brewer,[4] K. De Leon,[7] K. Fitzgerald,[3] G.X. Ge,[5] C. Hans,[6] S. Kosina,[6] L. Lui,[6] E. Majumder,[7] J.-W. Moon,[3] A. Ottwell,[8] S. Pfiffner,[4] H. Smith,[9] M. Thorgersen,[5] S. Turkarslan,[8] F. von Netzer,[10] D. Williams,[4] S.X. Wu,[6] G. Zane,[7] A. Zelaya,[9] E.J. Alm,[2] N.S. Baliga,[8] A.M. Deutschbauer,[6] M.W. Fields,[9] T.C. Hazen,[3,4] T.R. Northen,[6] J.D. Wall,[7] M.W.W. Adams,[5] R. Chakraborty,[6] J.-M. Chandonia,[6] D.A. Elias,[3] D.A. Stahl,[10] P.J. Walian,[6] J. Zhou[1,6], **A.P. Arkin,[6,11] and P.D. Adams,[6,11]**

[1]University of Oklahoma, Norman; [2]Massachusetts Institute of Technology, Cambridge; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4]University of Tennessee, Knoxville; [5]University of Georgia, Athens; [6]Lawrence Berkeley National Laboratory, Berkeley, CA; [7]University of Missouri, Columbia; [8]Institute for Systems Biology, Seattle, WA; [9]Montana State University, Bozeman; [10]University of Washington, Seattle; [11]University of California at Berkeley, Berkeley; [12]Northwest Missouri State University, Maryville; [13]Michigan State University, East Lansing.

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. Our goal is to understand how human activity associated with energetic processes - in particular, contamination and climate change - is affecting the ecology of critical soil, groundwater, and aquifer systems.

Despite decades of debate, many fundamental questions regarding the ecological drivers of community assembly remain unanswered. Classical stress gradient theory in ecology discussed the role of different deterministic forces (abiotic filtering, biotic interactions, etc.) under different degrees of stresses, however, we still know little about the basic question how deterministic versus stochastic forces vary along stress gradients, particularly in microbial ecology. The groundwater in the Oak Ridge Integrated Field Research Challenge site (FRC, Oak Ridge, TN) has large geochemical gradients and has been comprehensively surveyed, providing a rare opportunity to examine ecological processes and drivers shaping subsurface microbial diversity. Groundwater samples were taken from 98 wells that covered the geochemical diversity across

the site. Collected samples and wells were analyzed for 205 environmental (spatial and geochemical) variables. Aliquots (4 L) were filtered through 10-μm and 0.2-μm pore size filters, DNA was extracted from 0.2-μm filters and the 16S rRNA genes were sequenced on an Illumina MiSeq sequencer. We have applied various approaches to further disentangle the mechanisms controlling community assembly (stochasticity vs determinism) based on this dataset. Although some interesting trend of ecological stochasticity are found using previous reported approaches, the algorithms have inherent limitations and the accuracy of the estimation was doubtable. To address the challenge, we proposed a general mathematical framework to provide quantitative assessment of ecological stochasticity under different situations in which deterministic factors drive the communities more similar or dissimilar than null expectations. We developed a new normalized index, followed by testing it with simulated communities considering abiotic filtering, biotic interactions, environmental noise, and spatial scales. Comparing to previous approaches, the new index (NST) showed obviously higher accuracy and precision, of which the coefficients were over 0.9 in most simulated scenarios. However, all approaches showed limited performance at large spatial scale or under very high environmental noise. We then applied the new index to an empirical study on groundwater microbial community succession in response to emulsified vegetable oil (EVO) injection at FRC, with expected trend of ecological stochasticity has been supported by various evidences. The new index revealed that community assembly processes were shifted from deterministic to highly stochastic post-EVO input, and that, as EVO is consumed, the groundwater communities gradually returned to be more deterministic similar to pre-EVO injection. Null model algorithms and community similarity metrics showed strong effects on quantitatively estimating ecological stochasticity, among which preferred algorithm and metrics were suggested based on reasonability of the results. Then, we applied this new index to investigate how ecological stochasticity varies along stress gradients at FRC. The results suggested obvious decrease of ecological stochasticity by the increase of environmental stress. Furthermore, we explored depth profile of ecological stochasticity based on a pilot study of sediment bacterial communities in a contaminated well and a background well. We found obvious variation of ecological stochasticity from vadose to saturated layers related to both selection and dispersal limitation. Across different empirical datasets we tested, the new index generally can correct the overestimation of stochasticity by previous approaches to some extent, and revealed the obvious effects of environmental stress on the role of ecological stochasticity in governing underground microbiome.

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

ENIGMA Science Focus Area

A.P. Arkin[1,2], D.A. Stahl[13*](DAStahl@uw.edu) E.J. Alm[3], N.S. Baliga[4], A.M. Deutschbauer[1], M.W. Fields[5], T. C. Hazen[6,7], T.R. Northen[1], J.D. Wall[8],M.W.W. Adams[9], M. Auer[1], K. Bender[10], G. Butland[1], R. Chakraborty[1], J.-M. Chandonia[1], D.A. Elias[7], P.S. Novichkov[1], A. Mukhopadhyay[1], G.E. Siuzdak[12], P.J. Walian[1], J. Zhou[14], and **P.D. Adams**[1,2]

[1]**Lawrence Berkeley National Lab, Berkeley**; [2]University of California at Berkeley; [3]Massachusetts Institute of Technology, Cambridge; [4]Institute for Systems Biology, Seattle; [5]Montana State University, Bozeman; [6]University of Tennessee, Knoxville; [7]Oak Ridge National Lab, Oak Ridge; [8]University ofMissouri, Columbia; [9]University of Georgia, Athens; [10]Southern Illinois University, Carbondale; [11]Sandia National Lab, Livermore; [12]Scripps Research Institute, San Diego; [13]University of Washington, Seattle;[14]University of Oklahoma, Norman.

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.

The goal of Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) scientific focus area is to develop an unprecedented mechanistic and predictive understanding of the assembly, activity, stability and resilience of complex microbial communities in variable conditions. Established in 2009, ENIGMA is a collaborative consortium of 23 investigators at fourteen institutions located across the country and is led by principals at the Lawrence Berkeley National Laboratory. We are focused on sediment and groundwater communities and the processes that impact (and are impacted by) anthropogenic contaminants resulting from legacy nuclear programs at the Oak Ridge National Laboratory (ORNL). At the Oak Ridge Reservation (ORR) we perform sophisticated field experiments to measure the natural and anthropogenically perturbed dynamics of these geochemical processes and microbial community assembly and activity. From these we infer the chemical, physical and microbial interactions most predictive of these dynamics and estimate the ecological forces, both stochastic and deterministic, that shape community function. We then dissect the causal basis for these observations through laboratory studies of representative organisms in isolation and when they are part of synthetic or naturally derived assemblies. By using a unique array of culturing, genetic, physiological, bioreactor, and imaging technologies to map gene function and energy/material flow within and among cells, we determine how different genetic capabilities and regulatory strategies adapt microbes to ORR relevant conditions and to each other, thereby delivering an unprecedented mechanistic understanding of complex environmental bioprocesses and ecology.

Our prior high spatial resolution survey of ORR biogeochemical features identified the strongest correlations among chemical and physical environmental parameters and microbial community structure and function. We now seek to dissect the contributions of biotic, abiotic, stochastic and deterministic factors controlling community assembly, system-level processes, and system resilience. This overarching goal addresses one of the grand challenges identified in the most recent DOE Grand Challenges Report: to "define the levels of biological organization most relevant to scaling from single cells to ecosystems and global cycles; capture how that organization varies in time and space; and identify critical interactions that dictate rates of carbon, nutrient, and energy transformation in different environments."

To address this major challenge, the ENIGMA team has organized itself and its scientific/technological platforms into complementary campaigns for iterative field, laboratory, and computational analyses. We: 1) characterize stochastic and deterministic variations in community structure over time and space at the ORR in both unstimulated and stimulated conditions; 2) employ field and laboratory bioreactor systems that sustain both planktonic and attached communities to simulate simulating ORR environmental conditions, testing alternative mechanisms of community assembly and resilience; 3) design and employ methods for enriching and isolating key interacting sub-communities and active microbes representative of the condition-dependent diversity observed at the site; 4) develop and employ high-throughput experimental and computational technologies for the genetic manipulation of diverse environmental microbes to assess gene function, regulatory-network structure and adaptive genetic; 5) identify adaptive genetic elements that are laterally transferred among microbes in the field through analysis of and examination of the global plasmid pool; 6) Design and employ new metabolomics technologies and high-throughput physiological characterization platforms to gain insight into the flow of materials and energy through isolates and communities; 7) apply genetic, physiological, and bioreactor technologies to refine understanding of factors controlling the activities and persistence of sulfate and nitrate respiring populations in the field; 8) use experimental evolution of mono- or co-cultures in constant and fluctuating environments to elucidate species- and community-level mechanisms of adaptation and resilience to environmental stress.

We then integrate field observations, laboratory simulations of the environment, and genetic and metabolic experimentation to create the beginnings of a mechanistic, energetic and ecological model of how differential levels of nitrate, pH, oxygen, metals (essential, e.g. molybdenum, and toxic), and carbon coupled with physical dispersal processes affect community assembly at the ORR.

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

Field-relevant Isolates from the Oak Ridge FRC and their Deep Characterization and Functional Analysis.

R. Chakraborty[1]* (rchakraborty@lbl.gov), X. Wu[1], J. Voriskova[1], L.M Lui[1], H.K. Carlson[1], M. Garber[1], F. Song[1], A. Kothari[1], H.J. Smith, A. Mukhopadhyay[1], M.W. Fields, T.R. Northen[1], A.M. Deutschbauer[1], **A P. Arkin[1]** and **P. D. Adams[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]Montana State University, Bozeman

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. To isolate key microbes or stable consortia and develop approaches to characterize and dissect function of these key microbes from Oak Ridge Field Research Center, TN. Determine if environmental correlations similar to those found in the field are observed in the laboratory with wide diversity of single microbes or defined consortia.

In the past few years, we have developed both low and high throughput strategies to recover isolates representative of the genetic and functional potential observed in the field. Currently, we have in our collection close to 1000 diverse isolates obtained from background and contaminated ORR sites. These include isolates from field groundwater and sediment as well as laboratory enrichment and bioreactor studies conducted with field-sourced inocula. The ever increasing isolate collection includes representatives from diverse phylogenetic orders that nearly cover the diversity of our field site. Closely matching the site geochemistry to media composition has resulted in an increase of both the number and diversity of strains isolated. For example, our ability to measure the small molecule composition of the field site through development of sediment metabolomics and our ability to extract and characterize bioavailable natural organic carbon have enabled us to design media with carbon sources that closely resemble the quality and quantities of in situ nutrients. We are currently developing methods to isolate the rare biosphere, the phylogenetically diverse, uncharacterized clade members, and difficult to grow organisms. Representative isolates from processes of interest such as sulfate reduction have been obtained. Desulfoporosinus, a strictly anaerobic sulfate reducer, was isolated. Methodologies utilizing biorthogonal non-canonical amino acid tagging (BONCAT) in combination with fluorescent in situ hybridization (FISH) are being developed to cultivate translationally active phylogenies of interest. Certain strains cannot be isolated as monocultures and we are developing methods for highly rarified co-cultures for downstream analyses.

Several high-throughput tools and techniques are being developed to characterize different phenotypes and metabolisms of these diverse environmental isolates. For example, a high-throughput cultivation pipeline was used to measure the inhibitory toxicity of 80 inorganic ions against representative ENIGMA isolates and the resulting data was used to identify compounds that potentially limit the range of these organisms in the field. In addition, an automated high throughput DAP-seq analysis protocol is being developed to swiftly assess regulatory proteins and their targets in a large number of isolates that show importance in metal resistance, denitrification and sulfate reduction functions.

Funding statement: ENIGMA (http://enigma.lbl.gov) at LBNL supported by Office of Biological and Environmental Research US Dept of Energy Contract No: DE-AC02-05CH1123

Isolates and their Deep Characterization

Adam M. Deutschbauer[1], **Aindrila Mukhopadhyay[1](AMukhopadhyay@lbl.gov)**, Megan Garber[1], Lauren M. Lui[1], Alex Aaring[1], Hans Carlson[1], Alexey Kazakov[1], Markus De Raad, Benjamin Bowen[1], Suzanne Kosina[1], Lara Rajeev[1], Morgan N. Price[1], Jayashree Ray[1], Andrew Sczesnak[2], Onur Erbilgin[1], Pavel S. Novichkov[1], Judy Wall[3], Romy Chakraborty[1], Trent R. Northen[1], **Adam P. Arkin[1,2], and Paul D. Adams[1]**
[1]Lawrence Berkeley National Laboratory, Berkeley; [2]University of California at Berkeley; [3]University of Missouri, Columbia;

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. Deep characterization of isolates from the Oak Ridge Field Research Site is essential to understanding their role in microbial community assembly and key biological processes in this environment. The project goal is to develop and optimize high throughput and genome wide tools, to interrogate phenotypes and genotypes of the field isolates.

http://enigma.lbl.gov

Ground water and sediment samples from the Oak Ridge Field Research Site have yielded a large number of isolates that demonstrate desirable phenotypes for carbon utilization, metal reduction and resistance, denitrification and sulfate reduction. We aimed to characterize these isolates and obtain a genotype to phenotype understanding of molecular mechanisms for stress tolerance, cellular interactions and other biological phenomena. To assess the role of individual genes in fitness towards environmental parameters, we have developed a high-throughput genetic pipeline based on randomly barcoded transposon site sequencing (RB-TnSeq) and have applied this strategy to over 30 bacteria, including 10 ENIGMA FRC isolates. The resulting data have been used to identify phenotypes for previously hypothetical genes, improve hundreds of annotations of transport proteins and catabolic enzymes, fill in gaps in amino acid biosynthetic pathways, and identify novel catabolic pathways. To understand signaling and regulation, we have developed the DAP-seq methodology and examined bacterial response regulators in a high throughput and genome-wide manner. Our current studies, focused on denitrifying *Pseudomonas putida* strains have allowed us to elucidate multiple two-component signaling and the regulation of genes in response to metals and other environmental parameters. Coupled with other functional genomics studies (e.g. RNAseq) we map complex regulatory networks that underlie the response to key environmental factors such as metals. These data sets are also valuable in predicting and validating regulatory motifs. We have also examined the physiology of isolates in a co-culture format. Three of the isolates (*P. fluorescens* strains N1B4, N2E2, N2E3) were selected for in-depth analysis based on growth in pairwise co-cultures relative to monocultures with relevant genetic tools such as transposon mutant libraries. Our results show strain N1B4 (with truncated denitrifying pathway) grows more quickly and to greater density in co-culture than in monoculture. It was also found that some genes involved in nitrate reduction, sulfate permeability, molybdenum utilization, and anaerobic reduction are important for growth under

these conditions. In addition, a few uncharacterized genes were also shown to be positively correlated to growth. We are also using RB-TnSeq to study interactions between isolates where we determine which genes are important for susceptibility to the inhibitory compounds produced by other isolates. Barcoded transposon mutant libraries were grown in the presence of spent media from other strains. Improved fitness of a mutant indicates that the disrupted gene is related to the susceptibility of a compound. Preliminary metabolomics analysis (NIMS and RP-LC-MS) of a spent medium with known inhibitory activity has provided candidate inhibitory compounds. Analysis with MAGI provides a link between the metabolomics data and gene annotations to help predict which genes produce the compounds of interest. Together these methodologies and tools allow a deep understanding of the mechanisms and the interdependencies of key biological activities in these environments, and also a better understanding of community architecture.

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.

Linking Activity to Phylogeny in Groundwater/Soil Ecosystems

Heidi J. Smith[1],*(heidi.smith@montana.edu), Anna Zelaya[1], Isaac Miller[1], Dominique C. Joyner[2], Estelle Couradeau[3], Trent R. Northen[3], Terry C. Hazen[2], Matthew W. Fields[1], **A.P. Arkin[3], and P. Adams[3]**

[1]Montana State University, Bozeman; [2]University of Tennessee, Knoxville; [3]Lawrence Berkeley National Laboratory, Berkeley;

ENIGMA (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. A fundamental goal in the field of microbial ecology is to link the activity of specific microorganisms to processes occurring within an ecosystem. This project aims to identify the drivers of community structure and succession by quantifying activity and identifying the metabolically active fraction of microbial communities from both pristine and contaminated groundwater and sediment from the Field Research Center (FRC) at Oak Ridge National Laboratory (ORNL).

It is broadly accepted that free-living and attached communities have distinct microbial community compositions; however, due to sampling constraints there are significantly fewer studies that have simultaneously looked at the activity of microorganisms in both the sediment and groundwater fractions. Historically, it has been shown that not only are there greater densities of total cells (90-99.99% of the microbial biomass in porous aquifers) but there are also higher proportions of active cells associated with sediment compared to free-living groundwater cells. An explanation for differences in activities between attached and free-living populations is likely due to differences in the densities of cell abundances. However, it remains unresolved whether free-living cells in porous subsurface habitats are in fact metabolically slower or if the explanation for these differences in activities is merely based upon numbers respective to a given volume. Recently, it has been proposed that microbial competition selects against rapid growth in attached populations and that bulk-phase populations have faster growth rates . These findings offer a unique and contradictory perspective as to the role of free-living organisms compared to sediment biofilms which alter our current understanding of dispersal and colonization in porous environments as well as the distribution of microbial activities.

Using a combination of complementary culture-independent methods, activity rates and the active microbial assemblages were determined for groundwater and sediment associated cells from pristine and contaminated aquifers. Bioorthogonal non-canonical amino acid tagging (BONCAT) and Propidium Monoazide (PMA) were used to differentiate the contributions between different DNA pools (DNA from viable cells with intact cellular membranes versus extracellular or "relic" DNA) for two groundwater wells representing geochemical extremes. BONCAT samples were microscopically evaluated and sorted for amplicon sequencing (BONCAT-Seq). *Pseudomonadacea* and *Comamonadaceae* were the dominant

active assemblages for pristine groundwater, while *Xanthomonadaceae* and *Nocardiaceae* dominated contaminated groundwater. A greater diversity of active organisms was observed in background sites. For pristine wells, 1,268 OTUs were observed on average, with between 8.5 and 26% of identified OTUs being translationally active for 24 and 72 h of incubation, respectively. On average 346 OTUs were observed for contaminated groundwater, with between 60 and 66% of observed OTUs being active for 24 and 72 h, respectively.

With PMA analyses, the pristine groundwater showed average higher richness (4,958 OTUs) than contaminated groundwater (3,886 OTUs). For communities captured on 0.2 μM filters, OTU richness was similar with and without PMA treatment for pristine groundwater, and these results indicated that most sampled populations captured within this fraction did not have compromised cellular membranes and were viable. A similar trend was observed for contaminated groundwater, and OTU richness was similar between PMA treated and non-treated samples. Ordination analysis demonstrated that samples formed tight clusters that were primarily separated by well and secondarily by filter size. These results corroborated BONCAT analyses in that a significant portion of groundwater populations appear to be viable in terms of non-compromised membranes for both pristine and contaminated groundwater.

In pristine wells activity on a per cell basis was two to three-fold greater for planktonic cells compared to particle associated organisms, with small cells (<0.1um) contributing up to 19% of total activity. Conversely, in contaminated samples, activity was greater for sediment associated cells. We observed two to three orders of magnitude lower cell specific growth for sediment associated cells compared to planktonic groundwater cells. However, combining activity measurements, cellular abundances, porosity, and the degree of saturation, the biological activity of planktonic groundwater cells and sediment associated cells in a cubic meter of the saturated subsurface was estimated. The activity estimated to corresponding sediment associated cells accounted for up to 99% of the activity within a cubic meter of the saturated subsurface. Using a combination of methods, we show that the majority of planktonic populations in pristine aquifers are highly active and consist of intact cells. While attached populations have slower rates on a per cell basis, the sediment biofilms are responsible for the majority of the activity within a shallow aquifer.

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

Natural Organic Matter Dynamics and ExoMetabolomics for Microbial Cultivation from the Shallow Subsurface at the Oak Ridge FRC

Romy Chakraborty[1], Matthew W. Fields*[3] (matthew.fields@biofilm.montana.edu),Xiaoqin Wu[1], Terry C. Hazen[2], Heidi J. Smith[3], Trent R. Northen[1], Yina Liu[4], Nancy Hess[4], Ping Zhang[4], Jizhong Zhou[4], **Adam P. Arkin[1] and Paul D. Adams[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley CA; [2]Oak Ridge National Laboratory, Oak Ridge, TN & University of Tennessee, Knoxville, TN; [3]Montana State University, Bozeman, MT; [4]**University of Oklahoma, Norman;**

http://engima.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. Natural organic matter (NOM) is central to microbial food webs and microbially mediated NOM transformations determine much of the carbon (C) flux in subsurface environments. However, little is known about the molecular signatures from differnet C pools, the microbial activities/populations that regulate NOM turnover, nor the exometabolic markers for this activity. The goal of this project is to study the interactions between NOM (extracted from the field site) and native microbial communities present in uncontaminated and contaminated environments at Oak Ridge Field Research Center, TN.

Water-soluble NOM was extracted from sediment samples collected from two background uncontaminated sites. The amount of inorganic C in extracted NOM decreased significantly with depth. Extracted NOM was used as the sole source of carbon in controlled lab incubations. Lignin, lipid, and protein levels were similar for the different depth intervals, while the relative proportions of carbohydrate, tannin, and amino sugars declined with depth and condensed aromatics increased with depth. The nitrogen content of the extractable sediment NOM compounds varied between 13 to 28%, and the deeper sediment intervals contained more S than surface intervals. Groundwater NOM was compared to sediment-extracted NOM for an uncontaminated and contaminated well (well 106 v. 271). For groundwater, the NOM is dominated by amino acid-like fluorescence, and the fluorescence is approximately 60-fold greater in the contaminated well. The spectra also indicate the formation of more recalcitrant compounds in the sampled uncontaminated groundwater. For the sediment-extracted NOM, we used the humification index (HIX) as a measure of the complexity and the condensed (aromatic) nature of the NOM. The HIX indicates a drastic decrease in humic-like fluorescence for depths >2m for both cores compared to the shallower depths (<2m). The deeper depths had mainly amino acid-like fluorescence. NOM was extracted from the uncontaminated samples and incubated with groundwater as the inoculum. We monitored the trajectory of microbial biomass, respiration, community structure and activity over the course of the incubation. To document changes in organic matter chemistry, we applied high-resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) and soft X-ray absorption spectroscopy (sXAS). Together, these analyses provided a greater view of NOM degradation by indigenous

microbes. Transformation of NOM continued even after depletion of the labile C pool, and microbial populations shifted from presumptive copiotrophic to presumptive oligotrophic microorganisms that most likely possess greater affinity for diverse non-labile carbon. The C pool shifted during incubation, the proportion of lignin in cultures increased while overall protein levels declined. In addition, Geochip was used to identify the changes of microbial communities and expression of functional genes during transformation of the NOM, and putative gene sequences associated with chitin and lignin degradation increased after the depletion of more labile carbon. Futhermore, exometabolomic methods were developed to characterize ninety-six metabolites from sediment samples that included amino acids, carbohydrates (acid/alcohol), carboxylic acids, and nucleosides. The extraction procedure focused specifically on potentially accessible for microbial metabolites that may be present in sediment/soils. The water extractable organic carbon metabolites (WEOCs), were first qualitatively characterized using both liquid chromatography mass spectrometry (LC/MS) and gas chromatography mass spectrometry (GC/MS). Application of these complementary technologies provided orthogonal confirmation of metabolite identification while also expanding the analytical scope of the study. From these data, a subset of metabolites was selected for absolute quantification to assist in formulating a set of defined media that approximate the composition and quantity of metabolites within the sediments. One composition was demonstrated to support the growth of 25 phylogenetically diverse isolates out of 30 tested from the Oak Ridge Field Research Center (ORFRC), and detailed time series characterization of the substrate preferences are underway. Thus, the characterization of NOM and exometabolites in groundwater and sediment is described and then used to construct a defined medium for use with indigenous populations for the assessment of substrate utilization and microbial interactions.

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

Strain isolation, genome sequencing, and functional genomics reveals adaptive evolution of a *Pseudomonas* population at a human impacted field site

E.J. Alm[1]*(ejalm@mit.edu), S.J. Spencer[1], V.K. Mutalik[2,4], T. C. Hazen[5,6], J.-M. Chandonia[2], J. Zhou[3], R. Chakraborty[2], Deutchbauer A.[2], **A.P. Arkin[2,4], P.D. Adams[2,4]**

[1]Massachusetts Institute of Technology; [2]Lawrence Berkeley National Lab; [3]University of Oklahoma; [4]University of California at Berkeley; [5]University of Tennessee; [6]Oak Ridge National Lab

http://enigma.lbl.gov

Project Goals: Whole genome sequencing of environmental strains can reveal recent adaptive changes, making it a powerful tool to understand how environmental conditions impact the microbial genomes in those environments. Here we aimed to combine whole genome sequencing with functional genomics of field isolates to build a mechanistic understanding of natural selection at a human impacted groundwater site.

Microbial communities in groundwater ecosystems are inherently difficult to study due to sampling challenges paired with highly dynamic microenvironments. A number of recent efforts have characterized and mined the vast bacterial diversity within groundwater via amplicon surveys and shotgun metagenomics, but these data sources cannot reveal dynamic or historical changes without high-resolution sampling. Here we identify recent microbial adaptation within groundwater sites using isolation and whole genome sequencing, and combined these insights with functional genomic experiments directly measuring the contribution of genes to fitness under controlled laboratory conditions.

We isolated 139 *Pseudomonas* strains, that grouped into 15 lineages, and found that each lineage contained strains from multiple sampling sites across a broad geographic region. SNP analysis confirmed that recent mutational changes were impacting strains in individual sampling sites, despite a fast-flowing aquifer, and a two-component system showed evidence of positive selection. We also searched for gene gain and loss within strain isolates, and found evidence of transcriptional gene excision within otherwise clonal isolates at a single location.

We next sought to investigate the genes identified from genome sequencing with high-throughput in vitro fitness measurements. Transposon-based mutagenesis was used to generate a saturating library of knock out mutants which were grown across hundreds of conditions. Fitness measurement revealed that genes in the signaling pathway identified by genome analysis as undergoing adaptation were also among the most likely to be selected for across a variety of *in vitro* conditions. These data suggest it may be possible to not only identify the genes under selection, but the environmental factors driving that selection by combinging sequencing with high-throughput functional genomics.

Funding statement: ENIGMA (http://enigma.lbl.gov) at LBNL supported by Office of Biological and Environmental Research US Dept of Energy Contract No: DE-AC02-05CH1123

A Trait Based Approach to Exploring the Impacts of Species Diversity and Functional Diversity on Algal Community Productivity

Yiwei Cheng[1] (yiweicheng@lbl.gov), Nicholas J. Bouskill[1], Xavier Mayali[2], Todd Lane[3] and **Eoin L. Brodie[1]**

[1]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, CA, [3]Sandia 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.

From an energy perspective, eukaryotic microalgae and prokaryotic cyanobacteria (collectively known as algae) have been proposed as a promising commercially viable feedstock for biofuels. Research has been underway for decades to realize the full potential of algal biofuels at the commercial scale; however, open pond algal monocultures are frequently subject to collapse due to a range of factors including grazing, and infection by pathogens or parasites. Recently, there have been calls for the application of ecological principles to overcome such bottlenecks in algal cultivation. One approach is to replace less reliable monocultures with customized polycultures where the algal consortium composition is chosen using theoretical relationships between biodiversity, productivity and stability. For these purposes, measures of functional trait composition have been acknowledged as a better representation of biodiversity than traditional functional group assignments (e.g. based on taxonomy). Functional traits can be morphological, physiological or phenological characteristics that regulate organismal fitness in a given environment.

In order to predict the combinations of functional traits that can lead to sustained biomass yields of algal polycultures, we are taking a trait-based modeling approach, where the high biological complexity of an algal community can be reduced to being represented as members of functional guilds, each defined by distinct combinations of traits related to substrate utilization, growth rate, resource use efficiency and response to environmental factors (e.g. temperature,

light). We developed a trait-based dynamic energy budget model (TB-DEB) of microalgal polycultures to explore the relative impacts of microalgal functional diversity and species diversity on productivity across systems with varying and contrasting niches. The projected seasonal variations in algal biomass production were explored using environmental data (summer and winter temperature and photosynthetically active radiation) from operational algal ponds that are part of the ATP^3 consortium and that represent distinct climate zones (https://openei.org/wiki/ATP3_Data). In the first set of simulations, algal communities consisting of 48 algal species were selected based on combinations of physiological traits (e.g. nutrient affinity, maximum uptake rates) constrained by biochemical trade-offs. The trait values (i.e. relating to temperature, light, N and P harvesting) were based on published literature values from laboratory experiments. Trait values relating to N and P harvesting kinetics show allometric scaling as expected, while trade-offs between maximum uptake rate and substrate affinity also follow known relationships. Model simulations tracked community changes over time after starting out with equal amounts of each of the 48 algal species biomass. Simulations mimicking separate summer and winter climate regimes but with constant nutrient loading resulted in emergent community compositions with temperature differences driving the selection of algal strains from the polycultures.

In subsequent simulations, we specifically selected algal species composition based on the relative position of algal species along functional trait axes. First, algal strains were categorized based on their response to environmental temperature through the following traits: (1) optimal growth temperature, (2) minimum growth temperature, (3) maximum growth temperature. Each of these traits directly relates to resource acquisition and growth in response to environmental temperature. Next, we designed *in silico* algal polycultures with increasing functional and species diversity, and simulated growth under summer followed by winter climate regimes. The idea was to explore the manipulation of different levels of complementarity in resource acquisition. Analysis of cumulative system biomass of the designed algal polycultures at the end of winter highlighted the greater positive impact of increasing functional diversity, rather than species diversity, on system productivity, under operational environmental conditions. These simulated outcomes will be tested in ongoing laboratory experiments and further simulations are underway to explore such relationships in the presence of phycosphere-associated heterotrophic bacteria that we hypothesize can further alter algal traits such as nutrient acquisition and response to temperature changes.

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.

Using Data Mining and High-Throughput Cultivation Methods to Better Understand Bacterial:Fungal Interactions

Geoffrey L. House[1]* (ghouse@lanl.gov), Armand E. K. Dichosa[1], Debora F. Rodrigues[3], Hang N. Nguyen[3], Saskia Bindschedler[2], Jean F. Challacombe[1], Jamey D. Young[4], Pilar Junier[2], and **Patrick S. G. Chain[1]**

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

Interactions between bacteria and fungi are important determinants of ecosystem function, yet little is known about these interactions or how they operate. This is a critical knowledge gap as these interactions are important in addressing multiple DOE priorities including developing renewable energy sources, understanding the possible effects of Earth system change, and understanding how these interactions may help overcome energy and environmental challenges. Here we outline a range of research questions that we are beginning to address through a new SFA in order to better understand the diversity and function of these bacterial:fungal interactions. Using bioinformatics-based data mining of existing fungal genome sequencing data coupled with single cell isolation, microfluidics, and cultivation techniques, we are beginning to understand the diversity of bacteria that form associations with fungi, and how these associations affect both fungal and bacterial phenotypes.

There is an increasing amount of fungal genome and metagenome sequencing information available in public repositories such as JGI's Mycocosm and NCBI's sequence read archive (SRA), driven in part by the 1000 Fungal Genomes project. We have started to mine this genome sequencing data, beginning with raw sequence data obtained from fungal isolates. Our goal is to screen these sequences for any that appear to have bacterial origins. We propose to screen both assemblies and raw data since reads and/or contigs with bacterial signatures may have been discarded from the original fungal genome assembly. By using this approach across the phylogenetically broad sampling represented in the 1000 Fungal Genomes project, we will determine: 1) the range of bacterial diversity found across sequenced fungi, and 2) how this diversity is organized, (e.g. whether specific bacterial clades are general associates with many fungi or whether they associate specifically with particular fungal groups). After analyzing the genomic data of fungal isolates, we will take a similar approach to analyze metagenome samples from soil where both the fungi and the bacteria present in the sample are unknown. Because this latter step involves more complex samples and larger amounts of data, we will leverage an existing funded Exascale computing partnership between LANL, JGI, and LBNL. A critical capability for this work is the ability to better determine whether identified bacterial sequences represent bacterial contamination or true associations between bacteria and fungi. To make this determination as confidently as possible, we will use a combination of bioinformatics and experimental approaches. We will identify bacterial candidates for these associations by looking for strong phylogenetic associations between specific bacterial and fungal groups, as well as identifying any bacterial taxa that are recurrently encountered during the genome sequencing of

multiple fungi. We will then use targeted field soil collections and single cell cultivation techniques to better determine whether these identified candidate bacteria do form associations with fungi or whether they may have been contaminants in the genome sequencing.

In order to better characterize the growth responses of both partners in bacterial:fungal interactions, we will utilize a novel method, high-throughput screening of cell-to-cell interactions (Hi-SCI). Hi-SCI integrates the co-cultivation of bacteria and fungi within individual gel microdroplets (GMDs) with the efficient screening of the millions of different bacterial:fungal interactions represented within the GMDs. The GMDs allow either the random or the targeted capture of bacterial and fungal partners, and millions of GMDs can be grown simultaneously, each within its own cultivation chamber on a chip. The GMDs can then be rapidly screened and sorted using flow cytometry based on morphological or physiological signals, and then used for downstream genomics analysis. We have begun capturing fungal spores in GMDs to measure their speed of germination and other physical characteristics. We will then begin co-capturing fungal spores together with various known bacterial associates of fungi to determine how they affect fungal growth both alone and in different combinations.

The information gained about the diversity of bacteria associating with fungi and the information about the growth responses of fungi and bacteria when grown in combination within the GMDs can be tightly integrated. Better understanding the bacterial diversity associated with fungi can help guide the targeted co-capture of specific bacteria with specific fungi to access their growth responses. Conversely, better understanding of morphological and physiological interactions using GMD cultivation can guide molecular screens for specific functions or possible interactions between bacterial and fungal genomes.

We will use the results of the data mining approach to help inform the next experimental phases of this SFA aimed at better understanding bacterial:fungal interactions and their function in a range of growth conditions, and to interrogate gene, protein, and metabolite interactions. Together, these measurements and approaches will help give a better understanding of the complex interactions between bacteria and fungi in the soil, and how these interactions may exert important effects on plant growth and other important ecosystem services.

This SFA is supported under the Computational Biosciences Program of the Office of Biological and Environmental Research in the DOE Office of Science.

LA-UR-18-20257

Bridging the Soil Metagenome and Metaphenome Through Integrated Omics Analyses.

Mary S. Lipton[1] (mary.lipton@pnnl.gov) Taniya RoyChowdhury[1], Eric M. Bottos[1], Richard A. White III[1], Lisa M. Bramer[1], Joseph Brown[1], Jeremy D. Zucker[1], Young-Mo Kim[1], Colin J. Brislawn[1], Ari Jumpponen[3], Charles W. Rice[3], Sarah J. Fansler[1], Thomas O. Metz[1], Lee Ann McCue[1], Stephen J. Callister[1], **Kirsten S. Hofmockel[1], Janet K. Jansson[1]**
[1]Pacific Northwest National Laboratory
[2]Kansas State University, Department of Agronomy

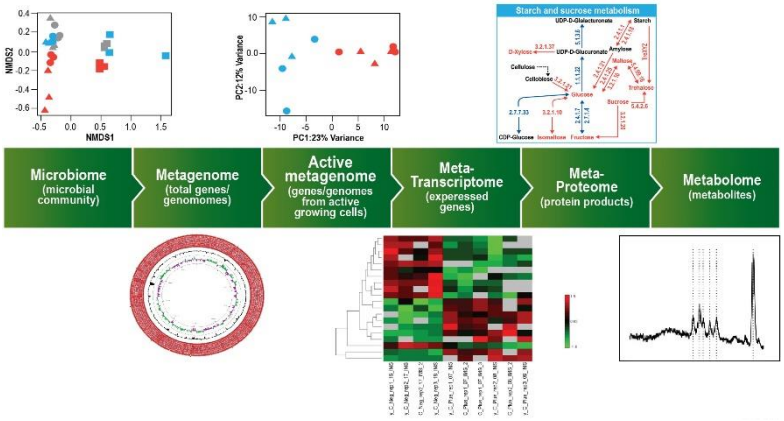
https://www.pnnl.gov/biology/programs/

Project Goals:
PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture through spatially explicit examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments will be designed to confront both the scaling challenges and inter-kingdom interactions that regulate networks of biochemical reactions. Individual- and population-based models for predicting interspecies and inter-kingdom interactions will be parameterized using experimental data, and predictions will be tested in soil to reveal spatially explicit microbial interactions. Discoveries from controlled experiments will be tested and validated in the field, using moisture gradient experiments at a new local field site. Data will be captured and shared through the establishment of a Soil Microbiome Knowledgebase (SMK). Knowledge gained will provide fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

Soil microbiome responses to changing environmental conditions are manifested as shifts in community structure and/or modifications to microbial activity. These phenomena have traditionally been characterized though bulk scale measurements of respiration, nutrient cycling and soil organic carbon decomposition. However, bulk scale measurements mask the complexity of molecular interactions that occur between specific members of soil microbial communities. Here we aim to circumvent current limitations in understanding of the soil microbiome by employing multi-omics technologies to gain a molecular understanding of inter-kingdom species interactions and biochemical processes that occur in soil and how specific processes are impacted by environmental change; specifically drought. The multi-omics data will also serve to determine how the combined phenotypes of the soil microbiomes result in the soil metaphenome.[1,2]

Currently, understanding of the soil metaphenome is hampered by the high microbial diversity and complexity of soil, experimental limitations, and incomplete genome-level understanding of biochemical pathways carried out by individual microbes and interacting community members in soil. Here, we demonstrate recent achievements in use of a multi-omics approach to link metaphenotypic observations to metagenome content. For this demonstration, we focused on a native prairie soil from Kansas where we have substantial existing metagenome data from the JGI Soil Metagenome Great Prairie pilot study.

First, we optimized a metaproteomics approach for soil. Limitations in mass spectrometric sensitivity and scan speed cycle can severely limit the number of peptides identified through this approach. However, the use of either on-line or off line 2 dimensional separations of the peptide mixture greatly increased the number of peptides identified as well as the molecular understanding of the metabolic and biochemical processes in the soil. In the future we aim to also apply specific activity based probes that are powerful tools to assess specific biochemical pathways that are active in soil microbiomes.



Potential Figure 1. A multi-omics approach enables deciphering different levels of soil microbiome information, ranging from community composition (microbiomes) to expression (metatranscriptomes) to protein production (metaproteomes) to metabolites (metabolomes). Together this multi-omics approach serves to elucidate the soil metaphenome; defined as the product of expressed genes (predicted from metagenomes) x environmental conditions.

We also used other steps in our multi-omics pipeline (Fig. 1) to assess key physiological traits expressed by a soil microbiome in response to specific nutrient and moisture pertubations[3]. Glycine was added as a mock root exudate and was found to have little effect on the community structure based on 16S measurements. However, there was a significant impact on the soil metaphenome as observed by significant changes in transcript expression and metabolite abundance. Not surprising, there was an increase in the metabolites that were products of glycine uptake and metabolism and an increase in the genes for glycine degradation and biosynthesis of serine and threonine, demonstrating the ability to discern these phenotypic responses in a completely untargeted manner in complex multi-omics datasets from highly diverse soils.

In contrast to the glycine addition, we found that moisture perturbation had an impact on the soil microbiome at all measured omics levels. In particular, soil desiccation caused a shift in community structure as well as a significant response in both the metatranscriptome and metabolome, including increases in metabolites and pathways for production of osmolytes, simple sugars, sugar alcohols and compatible solutes after drying. These results improve our understanding of the metabolic and biochemical processes occurring within soil microbial communities that ultimately lead to the soil metaphenome.

References:

1. Jansson, J.K. and Baker, E. S., (2016*) A multi-omic future for microbiome studies.* **Nat Microbiol**. *1:16049*
2. Jansson, J.K. and Hofmockel K.S., (2017) The soil microbiome – From metagenomics to metaphenomics **Current Opinion in Microbiology (in press)**
3. Roy Chowdhury, T., et al., (2017) *Soil microbiome response to nutrient and moisture perturbations.* **Microbiome (submitted).**

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.

Elucidation of Chemical Dark Matter in Soil Using 'Standards-free' Small Molecule Identification

Ryan Renslow[1]* (ryan.renslow@pnnl.gov), Dennis Thomas,[1] Sean Colby,[1] Jamie Nuñez,[1] Yasemin Yesiltepe,[2] Niranjan Govind,[1] John Cort,[1] Mark Borkum,[1] Nancy Washton,[1] Xueyun Zheng,[1] Erin Baker,[1] Justin Teeguarden,[1] Thomas O. Metz,[1] **Kirsten S. Hofmockel,[1] Janet K. Jansson[1]**

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Washington State University, Pullman, WA

https://www.pnnl.gov/biology/programs/

Project Goals: PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture through spatially explicit examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments will be designed to confront both the scaling challenges and inter-kingdom interactions that regulate networks of biochemical reactions. Individual- and population-based models for predicting interspecies and inter-kingdom interactions will be parameterized using experimental data, and predictions will be tested in soil to reveal spatially explicit microbial interactions. Discoveries from controlled experiments will be tested and validated in the field, using moisture gradient experiments at a new local field site. Data will be captured and shared through the establishment of a Soil Microbiome Knowledgebase (SMK). 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.

The ability to unambiguously and comprehensively identify metabolites and other small molecules in complex environmental samples will revolutionize our understanding of metabolic interactions occurring between members of soil microbial communities and between soil microbes and plants. In comparison to genetic information, much less is understood about the identities of small molecules comprising the metabolome, largely due to insufficiencies in molecular identification methods[1]. A significant obstacle in the field of metabolomics is the absence of methods for accurate, rapid, and comprehensive identification of small molecules in complex mixtures without relying on data obtained from analyses of authentic reference materials. This is critical for achieving our goals of identifying metabolic interactions between microbial species and for developing biochemical reaction networks of cellulose and chitin decomposition in grassland soils. We aim to compare the resulting networks to determine consistent or unique pathways across multiple soil sites. Our novel molecular identification pipeline, ISICLE (*In Silico* Chemical Library Engine), uses a large-scale computational chemistry platform that exploits PNNL's high-performance computational quantum chemistry software, NWChem, to calculate metabolite chemical properties, such as collision cross section (CCS)[2,3] and nuclear magnetic resonance (NMR) chemical shift. These properties can subsequently be used to make 'standards-free' identifications of small molecules, required to identify novel metabolic interactions for our project.

In initial tests of our platform, we investigated positional and geometric plant metabolite isomers analyzed using ion mobility spectrometry-mass spectrometry, and found that our platform was significantly more accurate at calculating CCS values compared to other methods, in part due to

Boltzmann weighting of hundreds of candidate conformers by relative energy. This level of accuracy enabled us to even distinguish *cis*/*trans* isomers[4]. Furthermore, we applied ISICLE in calculating CCS for metabolites in the Universal Natural Product Database in order to evaluate the theoretical resolving power of accurate mass and CCS. Finally, we analyzed environmental soil samples and CCSs were calculated *in silico* for possible metabolites. Several predicted degradation products, not available as authentic reference materials, were putatively identified only by accurate mass and *in silico*-derived CCS.

For novel molecule structure elucidation, ISICLE employs density functional theory (DFT) techniques to calculate NMR chemical shifts of molecule sets, with custom options for different solvents, nuclei, and user-selected chemical shift reference compounds. ISICLE calculates NMR chemical shifts of a molecule set with a variety of DFT methods while considering hundreds of conformers for each molecule. NMR chemical shift predictions were validated with experimental data from 300 molecules available in the literature. ^{1}H and ^{13}C chemical shifts were calculated with eight levels of DFT theory, with RMSD errors reaching 0.8 ppm and 5 ppm, respectively. Furthermore, we tested ISICLE on conformers obtained using DFT-based *ab initio* molecular dynamics, demonstrating the ability to reduce chemical shift errors to less than 0.1 ppm (^{1}H) and 2 ppm (^{13}C) using Boltzmann weighting of calculations for hundreds of conformers. Finally, we applied these methods to reassess the identification of wrightiadione as an isoflavonoid extracted from *Wrightia* plants. Based on our calculations, this molecule is actually the alkaloid tryptanthrin, an isobaric isostere of wrightiadione. To our knowledge, the misidentification of wrightiadione has heretofore been unrecognized, and the wrightiadione compound can no longer be said to be known to exist in nature.

Using our platform to identify active and present metabolic interaction is a central aspect of our project. This will allow us to further identify reaction modules (such as chemical and signaling pathways) that are not currently represented in soil biochemical reaction networks.

References

1. Metz, T. O. *et al.* Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: what can it add and how far can it go? *Bioanalysis* **9**, 81-98, doi:10.4155/bio-2016-0244 (2017).
2. Zhang, X. *et al.* SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. *Clinical Mass Spectrometry* **2**, 1-10, doi:https://doi.org/10.1016/j.clinms.2016.11.002 (2016).
3. Zheng, X. *et al.* Enhancing glycan isomer separations with metal ions and positive and negative polarity ion mobility spectrometry-mass spectrometry analyses. *Anal Bioanal Chem* **409**, 467-476, doi:10.1007/s00216-016-9866-4 (2017).
4. Zheng, X. *et al.* Structural elucidation of cis/trans dicaffeoylquinic acid photoisomerization using ion mobility spectrometry-mass spectrometry. *The Journal of Physical Chemistry Letters* **8**, 1381-1388, doi:10.1021/acs.jpclett.6b03015 (2017).

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.

Phenotypic Response of the Soil Microbiome to Environmental Perturbations

Janet K. Jansson[1*] (janet.jansson@pnnl.gov) and **Kirsten S. Hofmockel**[2] (kirsten.hofmockel@pnnl.gov)

[1]Biological Sciences Division, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory (PNNL), Richland, WA [2]Environmental Molecular Sciences Laboratory (EMSL), Earth and Biological Sciences Directorate, PNNL, Richland, WA

https://www.pnnl.gov/biology/programs/

Project Goals: PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture through spatially explicit examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments will be designed to confront both the scaling challenges and inter-kingdom interactions that regulate networks of biochemical reactions. Individual- and population-based models for predicting interspecies and inter-kingdom interactions will be parameterized using experimental data, and predictions will be tested in soil to reveal spatially explicit microbial interactions. Discoveries from controlled experiments will be tested and validated in the field, using moisture gradient experiments at a new local field site. Data will be captured and shared through the establishment of a Soil Microbiome Knowledgebase (SMK). 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.

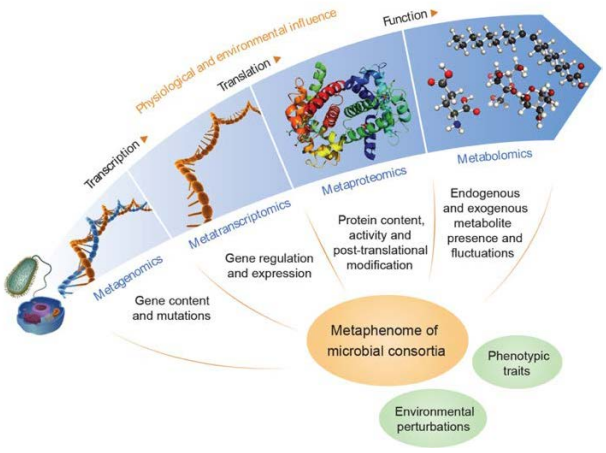


Figure 1. The metaphenome scales beyond the microbial metagenome and metatranscriptome to capture the emergent traits of the microbiome under specific environmental conditions (e.g. available resources; spatial, biotic, and abiotic constraints). Information gained from integration of omics technologies is essential for gaining mechanistic understanding of the soil microbiome metaphenome (*modified from Ref. 1*)

Abstract: Soil is a diverse ecosystem with microbial dark matter that remains to be discovered. Predicting microbial interactions in this complex system represents an exciting frontier and a grand challenge with implications for the productivity and fertility of our nation's soils. The *Pacific Northwest National Laboratory (PNNL) Soil Microbiome SFA will result in a molecular understanding of how moisture affects the interactions and phenotypes of microbial consortia, leading to our decadal vision of predicting the soil microbiome metaphenome to manage carbon*

(C) and nutrient cycling. By enhancing our understanding of the basic biology of microbial interactions, we will be able to predict ecological outcomes under changing environmental conditions. Moisture in particular drives microbial interactions and influences everything from cell function to substrate fate within soils. If we can understand how the physiology, metabolism, and interactions of soil microbes change in response to moisture, we will have a basis for modeling and predicting the soil microbiome *metaphenome* (Fig.1.).

Here we define the soil *metaphenome* as the product of expressed functions encoded in soil microbial genomes (metagenome) and the environment (resources available; spatial, biotic and abiotic constraints)[2]. The metaphenome is thus comprised of the sum of many phenotypes that are the result of biochemical interactions between members of the soil microbiome. In this SFA we aim to build an experimental and modeling framework across levels of complexity (from field to reduced complexity consortia) to gain fundamental understanding of the soil metaphenome. This knowledge gap is currently a challenge because of the high diversity of soil microbial species and the interdependencies of metabolic exchange between microbes – and across trophic levels. Our approach will thus be to define the soil metaphenome in a model grassland soil by deconstruction of the overall soil *biochemical reaction network* – that is still largely undefined – into smaller *functional modules*. Our obtained results will be transferable to more complex, non-model systems. An example of a functional module is the pathway for chitin degradation under aerobic conditions, with a defined soil moisture content, nutrient availability, temperature, pH, and other biogeochemical factors. The soil moisture content will also govern the spatial constraints of interacting members of the soil microbiome and their access to resources. We aim to focus on functional modules that result from interactions between soil microbes at the microscopic scale. These functional consortia can consist of a variety of interacting species, including bacteria, archaea, fungi and viruses that interact through metabolic exchange and chemical communication signals. The expressed functions encoded in the soil metagenome is in turn governed by the physiological status of the member populations. Only viable, active cells will contribute to the chitin degradation phenotype in the above example. *Our approach thus aims to 1) define the contributions of active cells and active metabolic pathways (expressed genes, proteins) that contribute to the soil metaphenome and 2) to define functional modules that can be plugged in to fill current gaps in the soil biogeochemical reaction network.*

References

1. Jansson, J.K. and E.S. Baker. 2016. A multi-omic future for microbiome studies. Nature Microbiology. 1:1-3.
2. Jansson, J.K. and K.S. Hofmockel. 2018. The soil microbiome: From metagenomics to metaphenomics. Current Opinion in Microbiology (in press).

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.*

Sequential Growth of Bacteria Following the Rewetting of a Seasonally Dried Grassland Soil

Steven J. Blazewicz[1,2*] (blazewicz1@llnl.gov), Erin E. Nuccio[1,2], Eoin L. Brodie[2,3], Mary K. Firestone[2,3], **Jennifer Pett-Ridge[1*] (pettridge2@llnl.gov)**

[1]Lawrence Livermore National Laboratory, Livermore, California, [2]University of California, Berkeley, California, [3]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. 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*.

Our SFA project objectives include:

1) Apply SIP-metagenomics to delineate how changing water regimes shape activity of individual microbial populations and expression of ecophysiological traits that affect the fate of microbial and plant C.
2) Identify and quantify mechanisms of mortality in the soil microbiome (focusing on phage lysis and water stress) and their contribution to C turnover and the biochemistry of microbial residues.
3) Measure how the soil microbiome and its products (cell envelope, extracellular polymeric substances, exo-enzymes) interact with contrasting mineral assemblages to control both short- and long-term soil C persistence.
4) Synthesize genome-scale ecophysiological trait data, population-specific growth and mortality, and SOM chemistry to build models of microbial functional guilds and SOM turnover, to predict the long-aspired connection between soil microbiomes and fate of soil C.

Microbial activity is stimulated by the rewetting of dry soils, resulting in a pulse of carbon mineralization and nutrient availability. This phenomenon is of particular interest because predicted changes in precipitation patterns in semi-arid life zones could profoundly change soil C dynamics and nutrient availability. While there has been much interest in the response of indigenous communities to wet-up of dry soil, to date, no work has identified the specific microorganisms in semi-arid soils that grow in response to soil wet-up. We used heavy water ($H_2{}^{18}O$) DNA stable isotope probing coupled with high throughput sequencing of bacterial 16S rRNA genes to characterize taxonomic and phylogenetic composition of bacteria following the rewetting of a seasonally dried California annual grassland soil.

Bacterial growth was detected at all time points throughout the incubation (3, 24, 72, 168 h), with patterns of sequential growth observable at the phylum and order levels. Of the 25 phyla detected in the pre-wet community, members of the Bacillales order (phylum Firmicutes) were the only detectable early responders – with an approximately 5% increase in relative abundance due to growth in the first 3 h after wet up. The second group of growing bacteria (detected at 24 h) included only Betaproteobacteria and Bacteroidetes. Members of the Burkholderiales order in the Betaproteobacteria phylum were the dominant growers during this period, with a 21% increase in relative abundance. A substantial amount of CO_2 is released within the first 24 hours after wet up, which suggests these few microbial groups may be predominantly responsible for a significant flux of CO_2 from this terrestrial ecosystem.

For actively growing bacteria, the highest richness was detected during the third time-period (between 24-72 h), with significant changes in relative abundance due to growth of 11 phyla. Nonmetric multidimensional ordination of community 16S rRNA sequences through time suggests a cyclical pattern for phylogenetic composition of growing bacteria, with the composition at 3 hours differing slightly from the pre-wet community, differing greatly at 24 h, and then becoming progressively more similar to the pre-wet community at 72 and finally 168 h. This suggests a degree of community resilience in response to this abrupt environmental change; the composition of new growth first diverges from the original community composition but eventually results in a partial return to the original composition. However, some net compositional changes were observed following wet-up. Actinobacteria were the most dominant pre-wet phylum, but Proteobacteria became the most dominant phylum by 168 h. This change in composition was likely driven by new growth, since Proteobacteria were found to increase in relative abundance across most time periods following wet-up, unlike Actinobacteria where comparatively small increases in growth were only observed during the latter two time points. Sequential growth patterns found at the phylum and order level suggest that an ecologically coherent response was observable at a high taxonomic level with members of certain groups employing different life strategies in response to rewetting of a dried 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 Number SCW1632 to the Lawrence Livermore National Laboratory. Work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52- 07NA27344 and at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231.

Multi-Component Characterizations of Algal Pond Ecosystems Reveal Algal Influence on Heterotroph Dynamics

Christopher S. Ward[1] (ward56@llnl.gov), Kunal Poorey[2], Carolyn L. Fisher[2], Laura T. Carney[3], Xavier Mayali[1], Michael P. Thelen[1], **Todd W. Lane[2]**, **Rhona K. Stuart**[1]

[1]Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, CA;
[2]Department of Systems Biology, Sandia National Laboratories, Livermore, CA;
[3]Agricultural Science and Molecular Ecology, Heliae Development LLC, Gilbert AZ

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 processes.

Algal biomass production is a promising renewable energy source and mass algal cultures in open ponds are currently grown worldwide. Considerable research effort has gone towards understanding abiotic drivers of algal productivity; nevertheless, unpredictable biomass yields continue to contribute to high production costs above the target levels needed to achieve economic feasibility. Recently, there has been a growing appreciation that the algal pond microbiome, e.g. bacterial heterotrophs and pathogens, microeukaryotic competitors, may be vital in determining outcomes of productivity and other relevant pond metrics. Thus, understanding the complex environmental roles of the algal microbiome is critical for advancing algal biofuel production. Here, we present findings from several ongoing studies in which microbiome characterizations have revealed compelling features of algal pond ecology.

Time course sampling of algal microbiomes was conducted on open ponds and raceways of *Microchloropsis salina* CCMP 1776 in AZ Sonoran Desert and TX Gulf Coast and *Chlorella sorokiniana* DOE 1412 in northern AZ. The compositions of bacterial communities associated with each microalgal culture were distinct, yet we identified several taxa held in common constituting a core algal microbiome present in all cultivation trials despite differences in, culture media, and geographical regions. The *Microchloropsis* associated microbiomes from these two sites were compared to the TARA Oceans survey to examine whether these algal-bacterial interactions also occur in natural ecosystems. Further, temporally-resolved analyses showed that progressions in bacterial community composition over culture duration were a consistent characteristic of microbiome development observed across cultivation trials.

In a separate study of *Haematococcus pluvialis* raceways frequently infected by chytrid fungi, we undertook a multi-component characterization of the algal pond ecosystem (i.e., bacterial community, chytrid infection, algal growth and exudate composition) by combining field and lab analyses. Due to the distinct growth stages of *H. pluvialis* (i.e., flagellate, palmella, aplanospore), we were able to differentiate between the influences of algal physiological state and culture age on bacterial community composition. Indeed, the pond microbiome displayed successional dynamics significantly explained by changes both in culture age and algal physiology. Similar to the *M. salina* and *C. sorokiniana* pond microbiomes, these results suggest environmental filtering of the bacterial community by phycospheric changes in carbon composition and quantity with algal culture status (i.e., age and growth stage). To further explore this, NMR-based metabolite analyses were conducted to determine differences in exudate composition between algal growth stages. Chytrid ecology was also associated with algal processes, with high pond infectivity only in cyst stages. Culture-based assays supported the dependence of chytrid infectivity on algal life stage, while also revealing a previously unknown non-infective chemoorganotrophic lifestyle regulated by dissolved organic carbon (DOC) availability. While the full significance of multiple chytrid trophic strategies in the pond setting is still unclear, it may relate to interplay with algal exudates and resource competition between heterotrophic bacteria and chytrids. The more holistic approach to studying the algal pond ecosystem described here revealed numerous ecological interactions that connect components primarily through carbon economy and flow. Together, the findings provide insight into critical ecosystem processes and dynamics that may ultimately contribute to algal health and productivity.

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.

Host-Linked Soil Viral Ecology along a Permafrost Thaw Gradient

Joanne B. Emerson[1,2*] (jbemerson@ucdavis.edu), Simon Roux[1,3], Jennifer R. Brum[1,4], Benjamin Bolduc[1], Ben J. Woodcroft[5], Ho Bin Jang[1], Caitlin M. Singleton[5], Lindsey M. Solden[1], Adrian E. Naas[6], Joel A. Boyd[5], Suzanne B. Hodgkins[7], Rachel M. Wilson[7], Gareth Trubl[1], **Changsheng Li**[8], **Steve Frolking**[9], Philip B. Pope[6], Kelly C. Wrighton[1], **Patrick M. Crill**[10], **Jeffrey P. Chanton**[7], **Scott R. Saleska**[11], **Gene W. Tyson**[5], **Virginia I. Rich**[1], and **Matthew B. Sullivan**[1]

[1]The Ohio State University, Columbus; [2]Current affiliation: University of California, Davis; [3]Current affiliation: Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek; [4]Current affiliation: Louisiana State University, Baton Rouge; [5]University of Queensland, Australia; [6]Norwegian University of Life Sciences, Norway; [7]Florida State University, Tallahassee; [8]Deceased, formerly University of New Hampshire, Durham; [9]University of New Hampshire, Durham; [10]Stockholm University, Sweden; and [11]University of Arizona, Tucson

http://isogenie.osu.edu

Project Goals: Our objective is to discover how microbial communities mediate the fate of carbon in thawing permafrost landscapes under climate change. Our team is engaged in a systems approach integrating molecular microbial and viral ecology, organic chemistry and stable and radiocarbon isotopes, and state-of-the-art modeling along an interconnected chronosequence of thawing permafrost and post-glacial lakes in subarctic Sweden.

Abstract: Climate change threatens to release large amounts of carbon sequestered at high latitudes, but constraints on the microbial metabolisms that mediate the release of greenhouse gases, such as methane and carbon dioxide, are poorly understood. While viral ecology is largely unexplored in soils, viral contributions to ecosystem processes in other systems (*e.g.*, viral impacts on microbial dynamics, metabolism, and biogeochemistry in the oceans) suggest that viruses are likely to be an important component of these microbially-mediated feedbacks to climate. Here, we investigated how viruses influence microbial ecology and carbon metabolism in peatland soils along a permafrost thaw gradient in northern Sweden. We recovered 1,907 viral populations (including genomes and genome fragments) from 197 bulk soil and size-fractionated metagenomes, approximately doubling known prokaryotic viral genera. 58% of these viruses were detected in metatranscriptomes and presumed to be active. *In silico* predictions linked 35% of the viruses to microbial host populations, highlighting likely viral predators of key carbon-cycling microorganisms, including methanogens and methanotrophs. Lineage-specific virus:host ratios varied with thaw, suggesting that viral infection dynamics may differentially impact microbial responses to a changing climate. Virus-encoded glycoside hydrolases (including an endo-mannanase with confirmed functional activity) indicated that viruses influence complex carbon degradation, while regression analyses revealed viral abundances—alone and together with host abundances—as significant predictors of methane dynamics. These findings suggest that viruses impact ecosystem function in terrestrial habitats undergoing climate change and identify multiple viral contributions to carbon cycling in thawing permafrost.

Funding statement. This research was funded by the U.S. Department of Energy Office of Biological and Environmental Research under the Genomic Science program, Awards DE-SC0004632, DE-SC0010580, and DE-SC0016440, with partial support from the Gordon and Betty Moore Foundation (GBMF#3305, 3790) and U.S. National Science Foundation (OCE#1536989).

Microbial environmental feedbacks and the evolution of soil organic matter

PI: Nicholas J. Bouskill
Lawrence Berkeley National Laboratory
Berkeley, CA, 94720

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 the 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 nutrient cycle. The chemical composition of SOM is in large measure determined by soil bacterial metabolism, which is impacted by changes in rainfall patterns. This research will conduct field and laboratory experiments and computational modeling to understand the role of microbial communities in stabilizing SOM under different water availability conditions in tropical soils. The results of this project will increase our understanding of the effects that microbes have on the global geochemical and nutrient cycles, addressing DOE's mission in energy and the environment.

Understanding Microbial Stress Responses in Soil Using Metagenome and Metatranscriptome Analysis

Paul Dijkstra[1] (Paul.Dijkstra@nau.edu), Peter Chuckran[1], Egbert Schwartz[1], Viacheslav Fofanov[2], Jennifer Pett-Ridge[3] and Bruce Hungate[1]

[1]Center of Ecosystem Science and Society and Biological Sciences, Northern Arizona University, [2]School of Informatics, Computing and Cyber Systems, Northern Arizona University, [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 will characterize this via stable isotope probing (SIP) of genome-resolved metagenomes. 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*.

Our SFA project objectives include:

1) Apply SIP-metagenomics to delineate how changing water regimes shape activity of individual microbial populations and expression of ecophysiological traits that affect the fate of microbial and plant C.
2) Identify and quantify mechanisms of mortality in the soil microbiome (focusing on phage lysis and water stress) and their contribution to C turnover and the biochemistry of microbial residues.
3) Measure how the soil microbiome and its products (cell envelope, extracellular polymeric substances, exo-enzymes) interact with contrasting mineral assemblages to control both short- and long-term soil C persistence.
4) Synthesize genome-scale ecophysiological trait data, population-specific growth and mortality, and SOM chemistry to build models of microbial functional guilds and SOM turnover, to predict the long-aspired connection between soil microbiomes and fate of soil C.

Managing soil health requires a detailed knowledge of how soil microbial metabolism and ecology affect soil organic matter formation and decomposition and respond to changes in the environment. It is often speculated that microbes in soil are C limited, either because of low C availability or low quality of the soil organic matter. This condition is thought to result in low C Use Efficiency, which reduces the microbial biomass and necromass production, with potentially negative effects on soil organic matter formation. Although we conceptually

understand how stress can affect soil organic matter formation, there are few direct ways to study stress in soil ecosystems. Here we propose to study microbial stress by analyzing the (meta)genome and (meta)transcriptome of bacterial species in soil ecosystems under high and low C availability.

Similar to "higher" organisms, bacteria have sophisticated response mechanisms to their environment. In this study, we will focus on bacterial sigma factors, transcription factors and regulators that are involved in changing gene-expression in response to stress conditions. We hypothesize that, by identifying transitions that results in the relative abundance of sigma factors, transcription factors and regulators, we can identify periods of microbial stress and measure how stress affects microbial C use efficiency, growth rates and overall community C and N cycling, and thus achieve greater insight into the processes that govern soil health.

We analyzed 10,601 bacterial genome, 4,929 metagenome and 1,753 metatranscriptome datasets from a wide range of ecosystems. Results show that of the various sigma factors, especially sigma E (regulating gene-expression in response to microbial envelope stress) is present at high relative abundance. A comparison between metagenomes and metatranscriptomes reveals that sigma factors associated with stress are less abundant in the metatranscriptomes than expected according to their abundances in metagenomes. This is interpreted to mean that microbial stress is not a dominant characteristic of microbial functioning in these environmental samples.

The Algal Ferredoxin Interactome

Venkataramanan Subramanian[1]* (venkat.subramanian@nrel.gov), Matt Wecker[2], Nitya Subrahmanian[4], Wei Xiong[1], Alexandra Dubini[3], David Gonzalez-Ballester[3], Patrice Hamel[4], **Maria Ghirardi[1]**

[1]National Renewable Energy Laboratory, Golden, CO; [2]GeneBiologics LLC, [3]Universidad de Córdoba, Córdoba, Spain; [4]The Ohio State University, Columbus; OH.

Project Goals: To unravel the specific roles of the different ferredoxin homologs in the green alga, *Chlamydomonas reinhardtii* and their contribution, either singly or in combination, in mediating electron transfer within specific metabolic pathways and under different stress conditions.

Abstract: Ferredoxins (FDX) are small, iron-sulfur cluster-containing proteins with strong negative redox potentials (-320 to -450 mV) that mediate redox transfer in various metabolic pathways. There are 13 FDX isoforms predicted in the *Chlamydomonas* genome that are differentially expressed in response to varying environmental conditions. Previous research in our lab has established a global FDX interaction network that suggests that they may have both specific and redundant functions (Peden et al. 2013).

This year, we focused our efforts on finishing our characterization of a *fdx5* mutant provided by the Grossman's group (Yang et al., 2016), and generation of a new *fdx2* strain, due to the loss of phenotype of the previously generated *fdx2* mutant. The *fdx5* strain was characterized under H_2 producing sulfur deprived conditions. When compared to its WT, CC-124, we observed (a) decreased O_2 evolution capacity under saturating light (in contrast with Yang et al., 2016) but similar respiratory capacity; (b) similar growth pattern and chlorophyll concentration under S-deprivation but a 24-h delay in starch accumulation and degradation, start of anaerobiosis, and H_2 and fermentative metabolites production; (c) lower initial level of C16:0 and C18:0 saturated fatty acids in *fdx5*, followed by decrease in C16:4 and C18:3n3 after 120 h s-deprivation in both strains (in contrast with the response of *fdx5* under dark incubation, Yang et al., 2016), resulting in similar ratio of saturated/unsaturated fatty acids for both strains at the end of the process; and (d) similar neutral lipid content, as measured by Bodipy staining, but better cell integrity in *fdx5* after 120 h S-deprivation. These results highlight the different effects of dark incubation vs. S-induced anaerobiosis on the *fdx5* strain and potentially support the observed interactions between FDX5 and the cell wall glycoprotein GP2; the Squamosa promoter binding protein, a central regulator of gene transcription; the programmed cell death protein; starch branching enzyme SB3; and the HYDEF hydrogenase maturation proteins. Clearly these interactions must be further verified by additional in vitro assays whenever possible.

An additional property of the mutant, based on potential interaction of FDX5 with peroxiredoxin 1 (PRX1), as suggested by previous yeast two-hybrid (Peden et al., 2015) and pull-down assays, was similarly examined under S-deprivation. We showed that *fdx5* was significantly more resistant to H_2O_2 following prolonged exposure to it but only at and after 72 h

of S-deprivation, suggesting changes in the cell wall protein composition under S-deprivation that are favorable to *fdx5*, which is also consistent with the well-known major membrane re-structuring following S-deprivation.

To identify a potential compensatory response to explain some of the above results by increase in the levels of other FDXs, we compared their transcript and protein levels at different points after S-deprivation. Our results show a major increase in FDX1 and FDX2 protein levels at t=0 in *fdx5* but no corresponding difference in their transcript levels. During S-deprivation, the transcript and protein levels of FDX1 are higher in *fdx5* while those of the other FDXs do not undergo significant changes with respect to the WT strain; the levels of the FDX2 protein decrease to undetectable levels soon after S-deprivation is initiated. Interestingly, the transcript levels one of the desaturases reported by Yang et al., 2016 as being dependent on FDX5 activity decreases substantially in *fdx5* during the later times of the S-deprivation process. We propose that the observed changes in FDX1 and FDX2 levels may represent a compensatory response of the mutant strain to the loss of FDX5 activity, and down-regulation of the desaturase FAD4 may be related to decreased need for its activity in *fdx5* under S-deprivation.

All the above observations are being further verified by transcriptomic analyses done with WT and *fdx5* under S-deprivation, and flux analysis under low light intensity. Moreover, since the *FDX5* mutation is present in a *nit*- background, we have generated a *fdx5* mutant in a *NIT*+ strain. This process has led to the discovery of a possible genetic suppressor of the *FDX5* mutation that can compensate for the dark no-growth phenotype of the *fdx5* mutant Genomic identification of this suppressor is in progress.

Finally, initial characterization studies of a new FDX2 down-regulated strain in a *nit*- background, is being carried out; initial results show that *fdx2*'s sensitivity to H_2O_2 is higher than WT under high light intensity but decreases to WT levels in the presence of potassium nitrate. These preliminary results strongly link FDX2 directly or indirectly to ROS detoxification or protection via the presence of nitrate. Further studies addressing this phenotype will be presented

References and publications generated by this project and cited above:

1. **Peden EA**, **Boehm M**, Mulder DW, Davis R, Old WM, King PW, **Ghirardi ML**, **Dubini A**. (2013). J. Biol. Chem. 288(49): 35192-209.
2. Yang W, Wittkopp TM, Li X, Warakanont J, **Dubini A**, Catalanotti C, Kim RG, et al. (2015). Proc Natl Acad Sci U S A. 112(48):14978-83.
3. **Subramanian, V, Dubini, A and Ghirardi, ML** (submitted review) The Role of Chlamydomonas Ferredoxins in Hydrogen Production and Other Related Metabolic Functions. In "Microbial Hydrogen Production: Achievements and Perspectives" (Eds. Seibert M. and Torzillo, G.) Springer; *submitted*.

Two additional manuscripts are in preparation, addressing (i) the physiological characterization of *fdx5* under S-deprivation, and (ii) the preliminary flux analysis of *fdx5* strain under very low light intensity.

This research is being supported by DOE's Office of Science, Biological and Environmental Research Office under a Science-Focused Area (SFA) project to NREL.

In silico system-level analysis of interactions in algal-bacterial co-cultures

Marc Griesemer[1] (griesemer1@llnl.gov), Miriam Windler[2], Jeff Kimbrel[1], Patrik D'haeseleer[3], Alfred Spormann[2], Xavier Mayali[4], **Ali Navid**[1]

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA; [2]Department of Civil and Environmental Engineering, Stanford University, Stanford, CA; [3]Global Security Computing Applications Division, 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 processes.

Microorganisms grow and adapt in a dynamic and diverse set of environments, all while interacting with other species. In fact, many organisms like algae owe their versatile set of metabolic capabilities to the community interactions they have with others [1]. In depth knowledge about causes and outcomes of these interactions are critical for assessing the robustness of natural and engineered microbial communities to various types of genetic, physical and population perturbations. However, characterizing the molecular mechanisms and regulatory processes that govern these interactions is challenging. Computational modeling that is informed by system-level experimental data is the key tool for examining the mechanisms and regulatory processes that govern these interactions and can be used to circumvent problems associated with experimental shortcomings. Here we report on our use of such modeling to characterize an algal-bacterial interaction. As part of our SFA, we are investigating the metabolic interactions of a number of microalgae with their bacterial symbionts, including the model green algae *Chlamydomonas reinhardtii* with an actinobacterium, *Arthrobacter sp.* P2b (P2b). The aim of these analyses is to understand the effect of these algal-bacterial interactions on algal biomass production. We compared *C. reinhardtii* and P2b co-culture growth to *C. reinhardtii* grown on its own (axenically). We observed that the algal biomass production in co-culture with P2b is strongly enhanced, suggesting a beneficial or commensalistic interaction between the two species. Our results show that the chlorophyll content of the algae is up to 3 times greater in the co-culture. The higher algal biomass in the co-culture with P2b is due to both higher cell densities and larger *C. reinhardtii* cells. In addition, we determined that the cell-free spent media from P2b alone

promotes biomass growth. In order to gain more knowledge about the metabotypes of the bacterium, we carried out growth experiments in phenotype microarray BIOLOG plates containing minimal media plus different carbon and nitrogen sources. We found that P2b is capable of growth on about one-third of the 190 substrates tested.
To further characterize P2b's physiology, we sequenced the genome (JGI CSP#1939), and combined the results from several different annotation tools, namely: RAST, KEGG, EFICAz, and TransportDB's transporter automatic annotation pipeline to gain a more complete functional annotation. The latter analysis resulted in identification of 557 transporter-associated genes; more than half of which were assigned to a specific substrate. Our enhanced annotation of P2b genome finds a number of interesting pathway for production of key secondary metabolites, including those for production of phytohormones that are known to improve algal growth[2].
We used DOE's Kbase (https://kbase.us) platform to generate a draft genome-scale model (GSM) of P2b's metabolism. We have begun the process of curating the draft model using our improved annotation results as well as data from our P2b BIOLOG analyses. We ultimately intend to pair the curated P2b model with a published GSM of *C. reinhardtii* [3] using Dynamic Flux Balance Analysis (dFBA) [4]. DFBA would allow us to simulate the interaction between the two organisms and to explore the influence of the growth-promoting bacterium on algal biomass production, and overall system robustness to genetic and environmental perturbations. Concurrent to curating the P2b GSM, we have been working on augmenting the DFBA method to better simulate unique biochemical environment of algal phycosphere (*i.e.*, local concentrations and relatively large flux/substrate concentration values that could necessitate varying simulation time steps). Overall, the results of our experimental and in silico analyses suggest that: a) interaction between *C. reinhardtii* and P2b is not specific to this unique pairing, i.e, P2b produces the effector compound regardless of presence or absence of algae, b) the effector molecule might be a phytohormone, c) the concentration of the effector compound is low so proximity to algae would increase the efficacy of the interactions. We are currently examining the last point using our augmented DFBA models.

References

[1] Ramanan R., Kim B-H., Oh H-M., Kim H-S. (2016). Algae-bacteria interactions: Evolution, ecology, and emerging applications. *Biotechnol Advances*, 34(1):14-29.
[2] Park W.K., Yoo G., Moon M., Kim C.W., Choi Y.E., Yang J.W. (2013). Phytohormone supplementation significantly increases growth of Chlamydomonas reinhardtii cultivated for biodiesel production. *Appl Biochem Biotechnol.*, 171(5):1128-42.
[3] Chang R.L., Ghamsari L., Manichaikul A., Hom E.F., Balaji S., Fu W., Shen Y., Hao T., Palsson B.O., Salehi-Ashtiani K., Papin J.A. (2011). Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol Syst Biol.* 7:1–13.
[4] Mahadevan R., Edwards J.S., Doyle III F.J. (2002) Dynamic Flux Balance Analysis of Diauxic Growth in *Escherichia coli*. Biophys J. 83:1331–1340.

This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NQ27344 and supported by the Genomic Sciences Program of the Office of Biological and Environmental Research under the LLNL Biofuels SFA, FWP SCW1039.

Modeling of Context-dependent Microbial Interactions in Biopolymer Degradation Networks

Hyun-Seob Song,[1]* (hyunseob.song@pnnl.gov), Nymul E. Khan,[1] Hans C. Bernstein,[1] **Kirsten S. Hofmockel**,[1] and **Janet K. Jansson**[1]

[1]Pacific Northwest National Laboratory, Richland, WA

https://www.pnnl.gov/biology/programs/

Project Goals: PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture through spatially explicit examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments will be designed to confront both the scaling challenges and inter-kingdom interactions that regulate networks of biochemical reactions. Individual- and population-based models for predicting interspecies and inter-kingdom interactions will be parameterized using experimental data, and predictions will be tested in soil to reveal spatially explicit microbial interactions. Discoveries from controlled experiments will be tested and validated in the field, using moisture gradient experiments at a new local field site. Data will be captured and shared through the establishment of a Soil Microbiome Knowledgebase (SMK). 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.

Degradation of biopolymers (such as chitin and cellulose) in soil is a key element of the terrestrial carbon cycle. Bacteria and fungi are known to play as major mediators of the degradation process in soil systems[1], but we have little understanding of how microorganisms interact with each other in degradation process under a dynamically varying environment. While *in silico* tools are often useful for improving the understanding of such a complex phenomenon in microbial communities[2], successful dynamic modeling examples for microbially-driven degradation processes are rare. A primary challenge may be ascribed to the difficulty in modeling a biological regulatory process, the mechanistic details of which are unknown in general[3, 4]. To address this challenge, we developed a modeling approach that enables simulating regulatory behavior of species by viewing microorganisms as dynamic control systems optimally modulating metabolic interactions with other species to maximize the survival chance of their own[5]. For conceptual understanding, we considered simplified degradation networks where microbes simultaneously cooperate (through the division of labor by synthesizing distinct degradation enzymes) and compete (for the consumption of digestible nutrients obtained as the final products of the degradation process) (Fig. 1). Our model not only enables predicting community dynamics that change interspecies interactions depending on the context, but also helps to characterize complex interactions by quantifying relative portions of cooperation and

competition. We also found that survival strategies of microbes could depend on a specific setting of interspecies interactions, e.g., species can choose to cooperate by providing nutrient sources to a competing partner if it can help to reduce the competition. As such, the developed conceptual model provides various interesting predictions that can be used as *ab initio* hypotheses on interspecies interactions in natural microbial communities. The developed interspecies interaction model will be parameterized based on the data collected from our field sampling and controlled experiments, ultimately to predict metabolic changes in soil microbiomes in response to moisture perturbations.

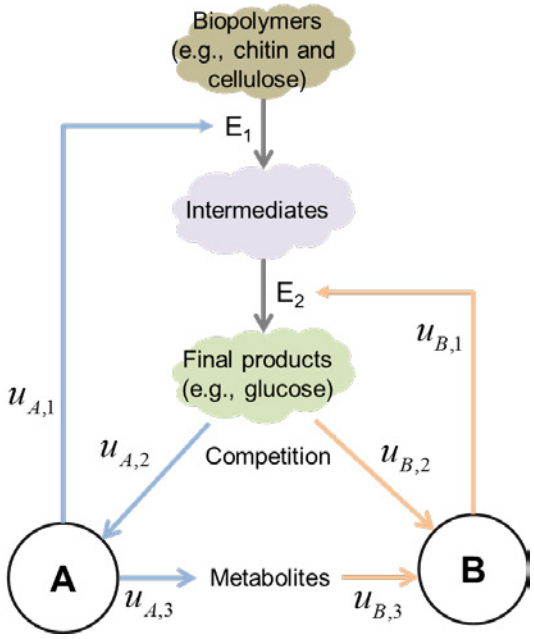


Figure 1. A conceptual biopolymer degradation network considered for understanding context-dependent interspecies interactions. Control actions of species A and B (denoted by u_A and u_B) dynamically vary depending on environment and the partner organism's response.

References

1. Beier S, Bertilsson S. Bacterial chitin degradation-mechanisms and ecophysiological strategies. *Front Microbiol* **4**, (2013).
2. Song H-S, Cannon WR, Beliaev AS, Konopka A. Mathematical modeling of microbial community dynamics: a methodological review. *Processes* **2**, 711-752 (2014).
3. Ramkrishna D, Song HS. Dynamic models of metabolism: Review of the cybernetic approach. *Aiche J* **58**, 986-997 (2012).
4. Song HS, *et al.* Regulation-Structured Dynamic Metabolic Model Provides a Potential Mechanism for Delayed Enzyme Response in Denitrification Process. *Front Microbiol* **8**, (2017).
5. Young JD, Ramkrishna D. On the matching and proportional laws of cybernetic models. *Biotechnol Progr* **23**, 83-99 (2007).

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.*

CRISPRi-Mediated Analysis of Biofilm Formation in Plant Growth Promoting *Pseudomonas fluorescens*

Marie Francoise Gros[1], Shalaka Shinde[1], Peter Larsen[1], Sara Forrester[1], Sarah Zerbs[1], Peter J. Korajczyk[1], Ken Kemner[1] and **Philippe Noirot**[1,] * (pnoirot@anl.gov).

[1]Argonne National Laboratory, Biosciences Division

Project Goals: Rhizobacterial communities provide benefits to plants in a variety of ways. To decipher specific molecular mechanisms by which rhizobacteria interact and benefit to plants, we study model rhizosphere communities in laboratory experimental systems. These simpler systems can be interrogated using a variety of approaches, including genetics.

Rhizobacteria of the *Pseudomonas fluorescens* group exhibit beneficial activities on multiple plants including *Populus* trees. *P. fluorescens* was characterized as a mycorrhizal helper bacteria (MHB) that promotes growth of the ectomycorrhizal fungus *Laccaria bicolor* at *Populus* roots. *P. fluorescens* also exhibits strain-specific promotion of aspen seedling growth in a laboratory experimental system [1]. The mechanisms underlying these beneficial interactions between roots and microbes still remain poorly characterized. To gain deeper insight into the rhizobacterial mechanisms, we investigated the spatial and temporal dynamics of the colonization by *P. fluorescens* of mycorrhizal and non-mycorrhizal roots of aspen seedlings [2]. Seedlings were grown in vertical plates in the laboratory, inoculated with a fluorescently labeled *Pseudomonas* strain, and root colonization was monitored over a period of five weeks. We observed an unsuspected diversity of bacterial assemblages at seedling roots that changed over time and were strongly affected by root mycorrhization with *Laccaria bicolor*. Biofilm assemblies on mycorrhizal and non-mycorrhizal roots were distinct, thick biofilms exhibiting internal channel-like structures were observed on non-mycorrhizal root surfaces whereas a layer of bacterial cells stacked along their long axis were found embedded within a gel-like substance at mycorrhizal roots. In the binary system, *P. fluorescens* SBW25 formed dense biofilms on aspen roots after 5 weeks and produced significant PGP phenotypes, suggesting that biofilms are associated with PGP activities [2].

In bacteria, the secondary signaling molecule cyclic diguanosine monophosphate (c-di-GMP) is a central regulator of bacterial transition from motile to biofilm life-styles [3]. We hypothesized that c-di-GMP is connecting rhizosphere signals to specific changes in cellular functions that trigger biofilm formation at roots. C-di-GMP is synthetized by enzymes called diguanylate cyclases (DGCs), degraded by phosphodiesterases (PDEs), and bound by effector proteins that regulate specific cellular functions. Among the 55 c-di-GMP-associated proteins encoded by *P. fluorescens* SBW25, six genes that were transcriptionally responsive to the presence of roots in our data sets and/or previously reported as environmentally induced genes [4], were selected for genetic interrogation. The 6 selected genes were down regulated using the CRISPR interference system (CRISPRi) for expression knockdown and phenotypes related to biofilm formation,

motility, and resistance to reactive oxygen species (ROS) were scored. Despite the high functional redundancy of the c-di-GMP regulatory network in *P. fluorescens* [3], we found that each knocked down gene exhibited measurable phenotypes relevant for interaction with plant roots (i.e., biofilm formation, ROS stress). These results support the notion that root-responsive c-di-GMP-associated proteins are important for the regulation of root-associated phenotypes. The corresponding genes will be our primary targets for future knockout experiments to study the molecular mechanisms that associate biofilm formation, patterns of root colonization, and plant growth promotion.

References

[1] S Shinde, JR Cumming, FR Collart, PH Noirot, PE Larsen (2017). Pseudomonas fluorescens Transportome Is Linked to Strain-Specific Plant Growth Promotion in Aspen Seedlings under Nutrient Stress. *Frontiers in Plant Science* 8.

[2] Marie-Francoise Noirot-Gros[1*], Shalaka Shinde[1], Peter E. Larsen[1], Sarah Zerbs[1], Peter J. Korajczyk[1], Ken Kemner[1] and Philippe H. Noirot[1]. Temporal study of Aspen roots colonization by Pseudomonads reveals strain-specific and mycorrhizal-specific patterns of biofilm formation. *Frontiers in Plant Science (under review)*

[3] Dahlstrom KM and GA O'Toole (2017). A Symphony of Cyclases: Specificity in Diguanylate Cyclase Signaling. *Annu Rev Microbiol.* 71:179-195.

[4] Silby et al., (2009). Genomic and genetic analyses of diversity and plant interactions of Pseudomonas fluorescens. Genome Biol. 2009; 10(5):R51

Funding statement. Research was supported in part by the "Environment Sensing and Response" Scientific Focus Area (SFA) program and in part from the "Small Worlds" project at Argonne National Laboratory. Development of CRISPRi approach in *P. fluorescens* was supported by LDRD funds.

A Systems Level Study of the Marine Diatom *Phaeodactylum* Reveals an Unexpected Mitochondrial Fatty Acid Beta Oxidation Pathway.

Graham Peers[1*](graham.peers@colostate.edu), Denis Jallet,[1] Denghui Xing,[1] Mark Moosburner,[2] and **Andrew Allen**[2]

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

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

Photosynthetic organisms balance heterotrophic and autotrophic metabolisms across the day to supply their requirements for growth and survival. We investigated the system level response of the diatom *Phaeodactylum tricornutum* during a shift from excess light, wherein energy is stored in compounds such as triacylglycerol (1), to light fluxes that limit the reaction rates of the photosynthetic system. Despite this shift, cells maintained maximal growth rates for 24 hours. Transcriptomic, proteomic and metabolomic data were fit to our previously established genome scale model of metabolism (2). This suggested that lipid catabolism fueled the rapid growth rates and photosynthetic remodeling required to capture a reduced light flux. *Phaeodactylum* was predicted to use both the plant-type peroxisomal fatty acid beta oxidation cycle as well as the cycle found in the mitochondria (animal-type). We found that the mitochondrial pathway was upregulated in response to low light, and is also dynamically regulated on a day/night cycle. The first step of this beta oxidation pathway is catalyzed by an Acyl-CoA Dehydrogenase (ACAD) and, in *Phaeodactylum*, this enzyme appears to be encoded by a gene resulting from a recent horizontal gene transfer from bacteria. We knocked out this gene using a CRISPR-Cas9 approach and found that mutant cells had slightly impaired growth rates compared to WT cells during growth in day/night conditions. Lipidomics studies revealed that the primary Triacylglycerol species, containing primarily 16:0 and 16:1 fatty acids, were not catabolized at night in the mutant. We also verified that this ACAD was targeted to the mitochondria and that a recombinant enzyme was able to catalyze the oxidation of palmitoyl-CoA (16:0). Mutant strains hyper-accumulated lipid following the amelioration of nitrogen starvation, suggesting that disrupting mitochondrial beta oxidation may be an effective mechanism to channel carbon to biofuel precursors in some conditions. We believe that this is the first characterization of triacylglycerol catabolism occurring by the animal-type pathway in a photosynthetic organism.

Publications

1. Jallet, D., Caballero, M.A., Gallina, A.A., Youngblood, M. and **Peers, G.** (2016) Photosynthetic physiology and biomass partitioning in the model diatom *Phaeodactylum tricornutum* grown in a sinusoidal light regime. ***Algal Research*** 18: 51-60.

2. Levering, J., Broddrick, J., Dupont, C.L., **Peers, G.**, Beeri, K., Mayers, J., Gallina, A.A., Allen, A.A., Palsson, B.O., Zengler, K. (2016) Genome scale model reveals metabolic basis of biomass partitioning in a model diatom. ***PLOS ONE*** DOI:10.1371/journal.pone.0155038

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

Title

Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in *P. tricornutum*

Authors

Niu Du[1,2], Jared T. Broddrick[3] (Co-1st author), Sarah R. Smith[2], Christopher L. DuPont[2], **Bernhard O. Palsson[3]**, B. Greg Mitchell[1], **Andrew E. Allen[1,2,*] (aallen@ucsd.edu)**

[1]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093
[2]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037
[3]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093

Project Goal

An in depth understanding of light energy utilization in photosynthesis under dynamic light environments has important implications for renewable food and fuel related industrial applications. In this study we simulated and investigated the intracellular metabolic flux in *P. tricornutum,* a unique organism that has amazing light tolerance capacity, under a wide range of acclimation light levels using state-of-the-art system biology tools. Our goal is to find out what physiological mechanisms enable this microalgal species to mediate excessive photon energy effectively.

Abstract

Photoacclimation consists of short and long-term strategies used by photosynthetic organisms to adapt to dynamic light environments. Observable photophysiology changes resulting from these strategies have been used in coarse-grained models to predict light-dependent growth and photosynthetic rates. However, the contribution of the broader metabolic network, relevant to species-specific strategies and fitness, is not accounted for in these simple models. Here we incorporate photophysiology experimental data with genome-scale modeling to characterize organism-level, light-dependent metabolic changes in the model diatom *P.tricornutum*. Oxygen evolution and photon absorption rates were combined with condition specific biomass compositions to characterize metabolic pathway usage of cells acclimated to four different light intensities. We identify photorespiration, an ornithine-glutamine shunt, and branched chain amino acid metabolism as the primary alternate electron pathways for consuming excess light energy. Additionally, simulations suggest carbon shunted through photorespiration is recycled back to the chloroplast as pyruvate, a mechanism distinct from known strategies in other phototrophs. Overall, our results suggest a flexible metabolic network in *P. tricornutum* that tunes inter-compartment metabolism to optimize energy transport between the organelles, consuming excess energy as needed. Characterization of these alternate electron flows broadens our understanding of energy partitioning strategies in this clade of ecologically important primary producers.

Funding Statement

Research supported by United States Department of Energy Genomics Science program grants DE-SC0008593 and DE-SC0018344.

Defining the Cellular Systems in Marine Diatoms That Provide for Their Predominant Acquisition of Dissolved Nitrate in World Oceans

James K. McCarthy[1], Hong Zheng[1], Anne Schulberg[1], Mark Moosburner[1,2], Andrew E. Allen[1,2]

[1]J. Craig Venter Institute and Scripps Institution of Oceanography, UCSD, La Jolla, CA; [2]Scripps Institution of Oceanography, UC San Diego, La Jolla, CA;

The ecological prominence of diatoms in the contemporary ocean environment largely results from their superior competitive ability for dissolved nitrate (NO_3^-). To investigate the cellular and genetic basis of diatom NO_3^- assimilation, we generated a knockout in the nitrate reductase gene (*NR*-KO) of the model pennate diatom *Phaeodactylum tricornutum*. In *NR*-KO cells, N assimilation was abolished although NO_3^-transport remained intact. Unassimilated NO_3^- accumulated in *NR*-KO cells resulting in swelling and associated changes in biochemical composition and physiology. Elevated expression of genes encoding putative vacuolar NO_3^- chloride channel (ClC) transporters, combined with electron micrographs (EM)indicating enlarged vacuoles, suggested vacuolar storage of NO_3^-. Triacylglycerol concentrations in the *NR*-KO cells increased immediately after the addition of NO_3^- and were concurrent with elevated gene expression of key TAG biosynthesis components. Simultaneously, notable induction of transcripts encoding proteins involved in thylakoid membrane lipid recycling suggested more abrupt repartitioning of carbon resources in *NR*-KO cells compared to WT. Conversely, ribosomal structure and photosystems genes were immediately deactivated in *NR*-KO cells after NO_3^- addition, followed within hours by those encoding for chlorophyll biosynthesis, and carbon fixation and metabolism. N-assimilation pathway genes respond uniquely, apparently induced simultaneously by both NO_3^- replete and deplete conditions. To confirm and advance our characterizations of these cellular systems, we have recently constructed gene knockouts of two putative vacuolar transporters, and two chloroplast-targeted nitrite reductases, and are screening for phenotypic changes. With plans to characterize N-stress signaling components in diatoms, we are initially using GC-MS analysis to monitor changes in metabolite concentrations between WT and *NR*-KO and *GSII*-KO mutants as cells respond N-replete and N-stressed conditions.

Funding Statement

Research supported by United States Department of Energy Genomics Science program grants DE-SC0008593 and DE-SC0018344.

Illuminating Diatom Cell Biology with a Genetically Encodable Tag for Electron Microscopy and Subcellular Proteomics

Jernej Turnšek,[1,2,3] **Andrew E. Allen**[2,3]* (aallen@jcvi.org)

[1]Harvard Medical School, Department of Systems Biology, Boston, MA; [2]Integrative Oceanography Division, Scripps Institution of Oceanography, UC San Diego, La Jolla, CA; [3]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA

Project Goals

A proteomics-based approach to identify proteins that are nearby targets of interest can provide powerful insights into biological pathways. We have implemented APEX2 in *Phaeodactylum tricornutum* to facilitate cataloging of biological pathway-specific proteomes in this model diatom species. APEX2, a soybean-derived ascorbate peroxidase, is a genetically encodable tag for electron microscopy and subcellular proteomics. When fused to a protein of interest, it permits spatially resolved proteomic mapping by oxidizing biotin-phenol to short-lived phenoxyl radicals which can covalently react with proximal proteins. We envision APEX2 will allow us to dissect a range of cell biology questions in *Phaeodactylum tricornutum* and other transformable diatoms which will collectively inform our future strategies for exploiting the biotechnological potential of these biogeochemically and evolutionarily important microeukaryotes.

Abstract

Iron is crucial for organisms across the tree of life as it plays an important role in many key enzymes linked to photosynthesis, respiration and nitrogen fixation. Primary productivity in ~30% of the modern oceans is limited by iron availability. Our laboratory has identified phytotransferrins as a new group of high affinity ferric iron-binding proteins widespread among marine microeukaryotes. Phytotransferrin ISIP2a from *Phaeodactylum tricornutum* internalizes ferric iron via endocytosis, but the molecular details behind ion liberation, chemical speciation and subsequent intracellular allocation remain elusive. We are using APEX2 in *Phaeodactylum tricornutum* to identify additional vesicle-associated proteins involved in this endocytic process. After supplementing APEX2-positive cells with biotin-phenol and hydrogen peroxide, cells are lysed, biotinylated proteins recovered with streptavidin-coated beads and analyzed using mass spectrometry. We show successful heterologous expression of APEX2-tagged ISIP2a, provide evidence for retained APEX2 activity *in* vivo, and present preliminary proteomic data containing a range of endocytosis- and trafficking-related proteins.

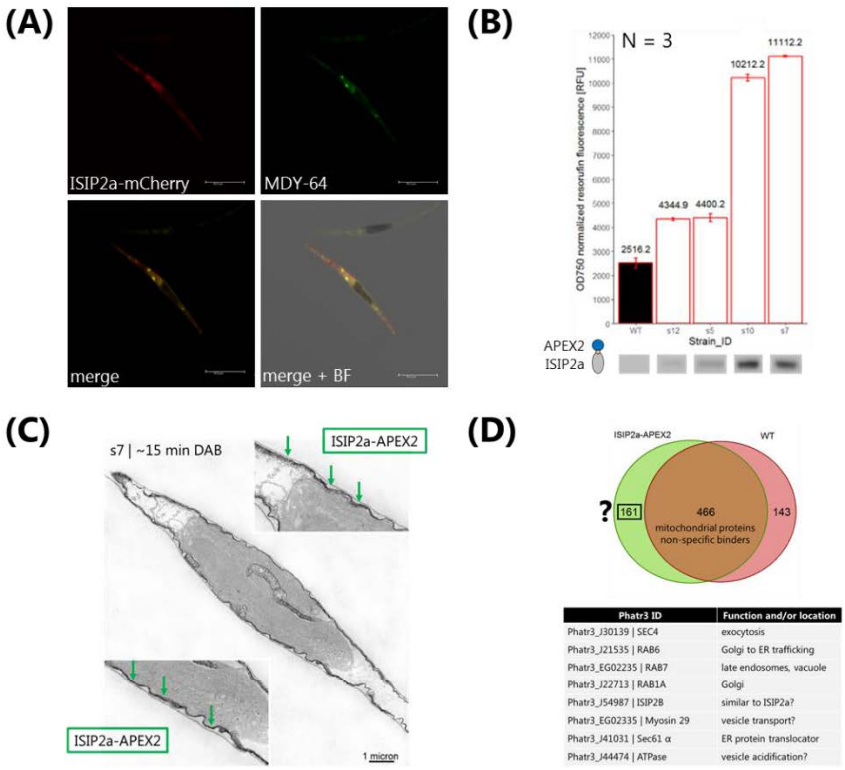


Figure 1 | ISIP2a and APEX2 in *Phaeodactylum tricornutum*. (A) ISIP2a-mCherry co-localizes with MDY-64 (membrane dye), further supporting ISIP2a internalization via endocytosis. *ΔISIP2a* genetic background. Scale bar is 10 µm. (B) A fluorescence-based assay data show retained *in vivo* APEX2 activity in 4 ISIP2a-APEX2 strains (empty bars). (C) APEX2 enables high resolution imaging of diatom proteins using electron microscopy. DAB (diaminobenzidine) is a contrast-generating substrate for APEX2. (D) Endocytosis-related proteins identified in a preliminary subcellular proteomics experiment using an ISIP2a-APEX2 strain.

References

1. McQuaid, J.B. *et al.* (2018) Carbonate sensitive phytotransferrin controls high affinity iron uptake in diatoms. Nature (*in review*).
2. Martell, J.D. *et al.* (2017) Electron microscopy using the genetically encoded APEX2 tag in cultured mammalian cells. *Nature Methods* 12, 9: 1792–1816.
3. Hung, V. *et al.* (2016) Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nature Protocols* 11: 456–475.
4. Lam, S.S. *et al.* (2015) Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods* 12, 1: 51–54.

Funding statement

This research is supported by the Department of Energy, Office of Biological and Environmental Research (BER) Grant DE-SC0018344 and by the Gordon and Betty Moore Foundation (GBMF) Grant GBMF4958.

Nitrate sensing with a model pennate diatom, *Phaeodactylum tricornutum*

Maxine H. Tan[1], Zheng Wang[2], Hong Zheng[3], Jill Meisenhelder[2], James McCarthy[3], Ahmed Moustafa[4], John McCrow[3], Tony Hunter[2], **Andrew Allen (aallen@jcvi.org)[1,3*]**

[1]Scripps Institution of Oceanography, UCSD, La Jolla, CA
[2]Salk Institute of Biological Studies, La Jolla, CA
[3]J. Craig Venter Institute, La Jolla, CA
[4]The American University of Cairo, Egypt

Nitrate sensing mechanisms have been studied for more than 30 years in organisms such as heterotrophic bacteria, cyanobacteria, green algae, yeast, fungi and plants. However, mechanisms by which diatoms sense and respond to nitrate availability are currently not known. We have identified a putative nitrate sensing protein, undescribed or characterized to date, in the genome of the model pennate diatom *Phaeodactylum tricornutum*. NitK contains a nitrate sensing Pfam (NIT) fused to a kinase domain and we hypothesize that it assists in sensing intracellular nitrate status. A search of available data indicated that this domain combination is prevalent in marine diatoms, dinoflagellates, haptophytes, and is also found in various green lineage prasinophyte species. In *P. tricornutum*, NitK transcript abundance appears to be regulated by addition of nitrate and nitrogen stress. Functional assays on NitK show autophosphorylation and that it is a serine/threonine kinase. Secondary structure predictions and analysis of membrane isolation fractions, complemented by preliminary localization of NitK, proposes that NitK is membrane bound. Mass spectrometry and yeast two-hybrid results suggest that NitK complexes with other nitrogen metabolism proteins, assisting in maintaining cellular nitrate homeostasis. Using *P. tricornutum* as a model for functional analysis, we strive to obtain an enhanced understanding of nitrogen metabolism and sensing mechanisms in major groups of marine primary producers which will improve our ability to forecast the impact of environmental change on marine ecosystems.

Funding Statement

Research supported by United States Department of Energy Genomics Science program grants DE-SC0008593 and DE-SC0018344.

THE METABOLIC LANDSCAPE CONNECTING CHLOROPLAST AND MITOCHONDRIA ENERGETICS IN DIATOMS

Sarah R. Smith1,[2], John P. McCrow[2], Christopher L. DuPont[2], **Bernhard O. Palsson[3]**, Jared T. Broddrick, **Andrew E. Allen[1,2,*] (aallen@jcvi.org)**

[1]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093
[2]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037
[3]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093

Energy obtained from light harvesting and photosynthetic electron transport is used to drive carbon fixation and other biosynthetic processes required for cellular growth in photoautotrophic organisms. In eukaryotes, these biosynthetic pathways can span multiple compartments and efficient growth requires coordinated delivery of C skeletons and energized biomolecules via primary metabolism. The primary metabolic landscape of diatoms, an ecologically important group of unicellular phototrophic eukaryotes, is highly distinct from that of plants and other algae. Diatoms possess certain metabolic pathways that are anomalous for phototrophs, such as a urea cycle, mitochondrially-targeted glycolysis pathway, and a chloroplast ornithine biosynthetic pathway, but their biological roles are largely unknown. Using phylogenetic, functional genomics, and systems biology approaches, we investigated the origin and significance of these pathways in diatoms. We found that this unique organization is a consequence of their complicated evolutionary history and functions to coordinate cellular energetics between the major energy-generating organelles, the chloroplast and mitochondria.

Funding Statement

Research supported by United States Department of Energy Genomics Science program grants DE-SC0008593 and DE-SC0018344.

Fungal and Bacterial Communities Vary in their Carbon Cycling Response to Climate

Sydney I. Glassman[1]* (sglassma@uci.edu), Claudia Weihe,[1] Michael L. Goulden,[1] Adam C. Martiny,[1] Kathleen K. Treseder,[1] Jennifer B. H. Martiny,[1] and **Steven D. Allison**[1]

[1]University of California, Irvine, California

http://allison.bio.uci.edu/projects/a-trait-based-framework-for/

http://allison.bio.uci.edu/projects/controls-over-decomposition.html

Project Goals: We aim to uncover the genomic and cellular mechanisms that underlie microbial traits important for carbon cycling under climate change. Traits related to drought tolerance, resource acquisition, and growth yield will be measured with molecular and biochemical approaches. In addition, we will quantify the extent to which trait tradeoffs occur across versus within microbial taxa and the consequences of these tradeoffs for litter carbon loss and transformation into microbial residues. Tradeoff relationships will be used to parameterize models that predict microbial community and carbon cycling responses to climate. The predictions will be validated and further refined with data from field measurements of microbial communities and carbon cycling processes.

Microbial communities are the engines of decomposition, a fundamental process in the carbon cycle. While much is understood about how changes in abiotic conditions and substrate quality affect decomposition rates, the role of microbial community composition remains elusive. This knowledge gap may be key for predicting how ecosystems will respond to climate change.

Here, we first surveyed the microbial community composition of decaying leaf litter along an elevation gradient in southern California, USA. The gradient included five sites with precipitation increasing, and temperature decreasing with elevation. To test the importance of climate for microbial composition, we conducted a transplant experiment. We used "microbial cages" to transplant leaf litter communities to different climates while preventing microbial exchange with the environment. We inoculated transplant communities onto a common, irradiated, grassland litter. In contrast with observational data, this manipulation isolates the effects of climatic conditions versus that of microbial composition on litter decomposition. To characterize fungal and bacterial community composition, we extracted DNA from the inoculum, intact litter at sites, and litter in the microbial cages, and amplified and sequenced part of the ITS and 16S rRNA regions. We also analyzed microbial biomass with fungal hyphal abundance counts and bacterial flow cytometry. We investigated the functional consequences of these transplants by measuring decomposition as mass loss and analyzing nutrient content.

We hypothesized that communities from along the climatic gradient would differ in their abundance and composition, and that communities at the extremes of the gradient would be most affected by climate. We found that microbial communities did indeed differ greatly between the

five sites (PERMANOVA; Fungi: $R^2 = 0.58$, $P < 0.01$; Bacteria: $R^2 = 0.61$, $P < 0.01$) with fungi primarily dominated by the Ascomycota. The main axis of community separation appeared to be between the colder and wetter sites versus the hotter and drier sites.

After transplantation, fungal communities retained a strong signature of the inoculum whereas bacteria were quickly influenced by local climate. Inoculum source was the strongest factor influencing fungi ($R^2 = 0.57$, $P < 0.01$), but site was also significant ($R^2 = 0.14$, $P < 0.01$). In contrast, site was the strongest factor affecting bacterial composition ($R^2 = 0.34$, $P < 0.01$), although inoculum source was also significant ($R^2 = 0.16$, $P < 0.01$). A strong site by inoculum interaction effect for both fungi ($R^2 = 0.15$, $P < 0.01$) and bacteria ($R^2 = 0.19$, $P < 0.01$) indicates that the strength of the inoculum effect varied by site, meaning that not all communities responded similarly to the climate gradient. In contrast to our expectation, communities from the extremes of the gradient were not most affected by climate. Instead, intermediate communities had the most significant impacts on decomposition and types of carbon compounds degraded when transplanted. Moreover, significant site by inoculum interactions impacting decomposition lasted a year after transplantation. These results demonstrate that microbial communities affect decomposition, but bacteria will shift more rapidly in response to climate than fungi, with fungi retaining a strong inoculum signal even after 1.5 years.

Publication

1. Glassman, SI, C Weihe, J Li, MBN Albright, C Looby, AC Martiny, KK Treseder, SD Allison, JBH Martiny. 2018. Microbial communities vary in their carbon cycling response to climate. In preparation.

This research was funded by the Genomic Science Program and the US National Science Foundation Ecosystem Studies Program. The Genomic Science Program is part of the Biological System Science Division (BSSD) of Biological and Environmental Research (BER) in the US Department of Energy Office of Science.

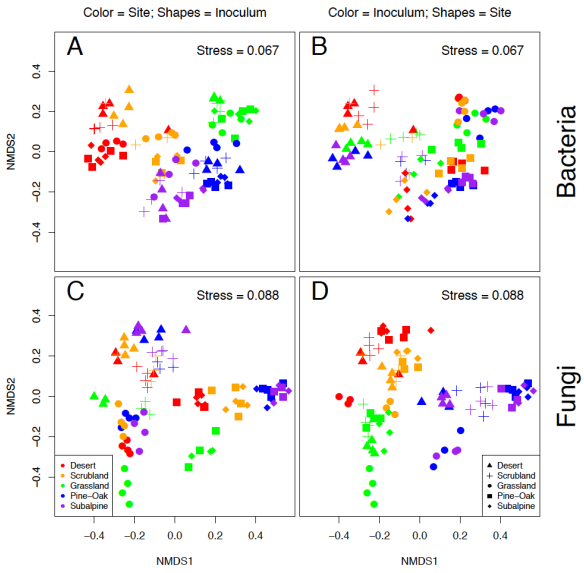


Fig. 1. NMDS of Bray-Curtis microbial community composition at 18 months for A) bacteria (16S) colored by site and shapes by inoculum and B) bacteria colored by inoculum and shapes by site. The bottom two panels are both fungal community composition with either C) colored by site or D) colored by inoculum.

Role of Lanthanides in the Complex Interplay of Alternative Enzyme Systems Involved in Methanotrophy

Yu Z.[1], Zheng Y.[1,2], Huang J.[1], **Chistoserdova L.*** (milachis@uw.edu)[1], and **Lidstrom M.E.**[1]

[1]University of Washington, Seattle; [2]Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, China

Project Goals: This project addresses the structure and function of microbial communities active in methane consumption, using lake sediment as a model. Through manipulations of synthetic communities and systems biology approaches, we are striving to understand the molecular mechanisms that form a basis for specific interspecies interactions in microbial oxidation of methane. In this phase of the project our goals were: 1) Evaluate behavior of model synthetic communities comprised of major functional guilds defined in prior research, and assess performance of these communities; 2) Evaluate the functions and the regulation of the alternative methanol dehydrogenases, previously identified as the cross-talk points among keystone species in methane-utilizing communities; 3) Obtain enzyme-level insights into lanthanide-dependent methanol oxidation systems in these keystone species.

We followed species dynamics and gene expression patterns in synthetic bacterial communities consisting of species that compete for methane, and species unable to consume methane, which could only succeed through cooperative interactions (a total of 50 species). We demonstrated that these communities mostly select for two functional guilds, methanotrophs of the family *Methylococcaceae* and non-methanotrophic methylotrophs of the family *Methylophilaceae*, these taxonomic guilds outcompeting all other species included in the synthetic mix. The metatranscriptomics analysis uncovered that in both *Methylococcaceae* and *Methylophilaceae*, some of the most highly transcribed genes were the ones encoding methanol dehydrogenases (MDH). Remarkably, expression of alternative MDH genes (*mxaFI* versus *xoxF*), previously shown to be subjects to the 'rare Earth element switch', was found to depend on environmental factors such as nitrogen source and methane and O_2 partial pressures, while also being species-specific. Along with multiple *xoxF* genes, multiple genes encoding divergent cytochromes were highly expressed in both *Methylophilaceae* and *Methylococcaceae*, suggesting their function in methanol metabolism, likely as electron acceptors from XoxF enzymes. This experiment tested a synthetic community model that is much simplified compared to natural communities consuming methane, but more complex than the previously utilized two-species model. The performance of this model identified prominent species for future synthetic ecology experiments and highlighted both advantages of this approach and the challenges that it presents.

We analyzed the physiological effects of a mutation in *xoxG*, a gene encoding a novel cytochrome XoxG(4) identified through metatranscriptomics, and compared these to the effects of mutation in XoxF, a lanthanide-dependent methanol dehydrogenase, at the enzyme activity level and also at the community function level, using *Methylomonas* sp. LW13 as a model

organism. Through comparative phenotypic characterization, we establish XoxG as the second protein directly involved in lanthanide-dependent metabolism, likely as a dedicated electron acceptor from XoxF. However, mutation in XoxG caused a phenotype that was dramatically different from the phenotype of the mutant in XoxF, suggesting a secondary function for this cytochrome, in metabolism of methane. We further purified XoxG(4) and demonstrated that this protein is a true cytochrome *c*, based on the typical absorption spectra, and we demonstrated that XoxG can be directly reduced by a purified XoxF, supporting one of its proposed physiological functions. Overall, our data continued to suggest the complex nature of the interplay between the calcium-dependent and the lanthanide-dependent alcohol oxidation systems, while they also suggested that addressing the roles of these alternative systems is important at the enzyme as well as community function level, in addition to gene transcription level.

To evaluate the properties of XoxF enzymes in major community partners, we expressed, purified and biochemically characterized three enzymes, two belonging to the XoxF4 phylogenetic clade, from *Methylotenera mobilis* JLW8, and one belonging to the XoxF5 clade, from *Methylomonas* sp. LW13, and demonstrated that, while all dependent on lanthanides as cofactors, they possess different properties in terms of kinetics and substrate specificities.

Overall, our recent data provide new insights and expand our understanding of the role of metals in methanol oxidation, an essential step in methanotrophy, and further point to the complex nature of its biochemistry. Our data strongly suggest that, likely, a balance between the two types of MDH enzymes, calcium-dependent versus lanthanide-dependent, rather than the on/off switch, provide metabolic robustness to natural populations of methanotrophs and non-methanotrophic methylotrophs.

Publications

1. Zheng Y, Huang J, Zhao F, Chistoserdova L (2018) Physiological effect of XoxG(4) on lanthanide-dependent methanotrophy. mBio Submitted.
2. Chistoserdova L (2018) Applications of methylotrophs: can single carbon be harnessed for biotechnology? Curr Opin Biotechnol Submitted.
3. Chistoserdova L, Kalyuzhnaya MG (2018) Current trends in methylotrophy. Trends Microbiol Submitted.
4. Chistoserdova L (2018) Methanotrophy: an evolving field. In: Kalyuzhnaya M, Xing X-H Eds, Methane biocatalysis: Paving the way to sustainability. Springer, Springer-Verlag, Berlin Heidelberg, Germany In Press.
5. Yu Z, Beck DAC, Chistoserdova L (2017) Natural selection in synthetic communities highlights the roles of *Methylococcaceae* and *Methylophilaceae* and suggests differential roles for alternative methanol dehydrogenases in methane consumption. Front Microbiol 8:2392.
6. Yu Z, Chistoserdova L (2017) Communal metabolism of methane and the rare Earth element switch. J Bacteriol pii:JB.00328-17.
7. Chistoserdova L (2017) Omics approaches to studying methylotrophs and methylotroph communities. In: Marco D Ed, Metagenomics: current advances and emerging concepts. Curr Isues Mol Biol 24:119-142.

This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC-0016224

Competition for a cross-fed nutrient between bacterial mutualists

Alexandra L. McCully,[1] Breah LaSarre,[1], Megan G. Behringer[2], Jennifer R. Gliessman[1], Evgeny V. Pilipenk[3] Jeffrey L. Mazny[1], Michael Lynch[2], D. Allan Drummond[3], and **James B. McKinlay *(jmckinla@indiana.edu)**[1]

[1]Indiana University, Bloomington; [2]Arizona State University, Tempe, AZ, [3]University of Chicago, Chicago IL.

Project Goals: The goals of this project are to (i) develop a stable hydrogen gas-producing coculture between *Rhodopseudomonas palustris* and fermentative microbes, such as *Escherichia coli*, (ii) use genetic, biochemical, evolutionary, and systems biology approaches to characterize and manipulate the factors governing microbial interactions and H_2 production.

Synthetic microbial communities, or cocultures, preserve natural interactions while allowing for an expanded range of experimental approaches and control. Cocultures are thus valuable for addressing ecological and evolutionary questions, such as how microbial communities transform carbon. Cocultures can also combine diverse traits to convert renewable resources into fuels and other useful chemicals. However, establishing cocultures that support stable coexistence and yield reproducible results is often challenging.

We developed an anaerobic coculture between fermentative *Escherichia coli* and an engineered strain of phototrophic *Rhodopseudomonas palustris* (Nx) that together convert carbohydrates into H_2 gas, a potential biofuel. The two bacteria form a syntrophic relationship wherein *E. coli* ferments glucose and excretes essential carbon (organic acids) for *R. palustris* while *R. palustris* fixes N_2 and excretes essential nitrogen (NH_4^+) for *E. coli*. The bidirectional exchange ensures stable coexistence as two populations converge on a common equilibrium from starting ratios spanning over six orders of magnitude in favor of either species[1]. Growth and metabolic trends are highly reproducible over serial transfers. We have also developed kinetic models that accurately describe coculture dynamics and can be used to predict the effects of various perturbations[1-3]. We are now using computational and experimental approaches to address the importance of core metabolic traits in deciding the phenotypic behavior of this community, which resembles organic acid cross-feeding interactions found in anaerobic food webs and nitrogen cross-feeding found in various environments.

Previously we determined that the level of organic acid excretion by *E. coli*, which is directly influenced by the level of NH_4^+excretion by *R. palustris,* determines the relative benefit that *R. palustris* receives from organic acids[1,2]. At high levels of excretion, organic acids become detrimental rather than beneficial as they acidify the environment and inhibit growth[1]. At low levels of excretion, organic acids can sustain the mutualism even through periods of starvation[2]. More recently we have focused on the role NH_4^+, which unlike organic acids, is not a waste

product of metabolism but rather a metabolite that both species value as a nutrient. We thus explored how interpartner competition for a communally valuable cross-fed nutrient impacts mutualism dynamics. By decreasing the affinity of each species for NH_4^+ both computationally and by genetically disrupting NH_4^+ transporters, we discovered that mutualism stability necessitates that the recipient, *E. coli*, have a competitive advantage against the producer in obtaining the cross-fed nutrient, provided that the nutrient is generated intracellularly[3]. RNAseq and proteomics analysis, along with genetic disruption of the *E. coli* master regulator of nitrogen metabolism, NtrC, revealed that the nitrogen starvation response is crucial for *E. coli* to assume a physiological state that reliably maintains coexistence with *R. palustris.* We propose that the requirement for recipient-biased competition is a general rule for mutualistic coexistence based on the transfer of intracellularly generated, communally valuable resources.

References

1. LaSarre, B, AL McCully, JT Lennon, and JB McKinlay. 2017. Microbial mutualism dynamics governed by dose-dependent toxicity of cross-fed nutrients. ISME J. 11:337–348.
2. McCully, AL, B LaSarre, and JB McKinlay. 2017. Growth-independent cross-feeding modifies boundaries for coexistence in a bacterial mutualism. Environmental Microbiology. 19: 3538-3550.
3. McCully, AL, B LaSarre, and JB McKinlay. 2017. Recipient-biased competition for an intracellulary generated cross-fed resources is required for coexistence in a bacterial mutualism. mBio. 8: e01620-17

This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under award number DE-SC0008131.

Microbial Interactions in The Anaerobic Oxidation of Methane: Model Simulations Constrained by Process Rates and Activity Patterns

He, Xiaojia[1]*(xiaojia.he25@uga.edu), Chadwick, Grayson,[2] **Kempes, Christopher**,[4] Shi, Yimeng,[1] McGlynn, Shawn[2,3], **Orphan, Victoria**,[2] and **Meile, Christof**[1]

[1] University of Georgia, Athens, GA, USA; [2] California Institute of Technology, Pasadena, CA, USA; [3]Tokyo Institute of Technology, Tokyo, Japan; and [4]Santa Fe Institute, Santa Fe, NM, USA

https://faculty.franklin.uga.edu/cmeile/content/microbial-metabolism

Project Goals: This research is focused on evaluating potential mechanisms governing anaerobic oxidation of methane (AOM) within the archaeal-bacterial consortia in marine sediments. We simulate the activity of AOM in microbial aggregates for methane oxidizing archaea, with electron transfer through diffusion of dissolved molecules to the bacteria which reduce sulfate (SRB-MIET), the disulfide mechanism in which ANME archaea coupling methane-oxidation to sulfate-reduction directly and DIET (direct interspecies electron transfer). We compare the resulting methane oxidation rates and activity distribution patterns to observational data from samples from seeps at Hydrate Ridge off the coast of Oregon, USA, and explore the effect of consortia size, the intra-consortium spatial distribution of archaeal and bacterial cells, and pH variations and the sensitivity towards poorly constrained model parameters. These model simulations are analyzed in the context of likely interactions between the archaea and bacteria in the light of this process-based quantitative analysis.

Abstract text. Anaerobic oxidation of methane (AOM) describes the conversion of methane to CO_2 in the absence of oxygen, and in marine sediments is commonly associated with the reduction of sulfate. AOM in marine sediments is estimated to consume about three-quarters of the global net methane emission to the atmosphere, thus acts as a significant sink for methane, a greenhouse gas with a warming potential ~25 times that of CO_2. However, the details of the physiological mechanism underlying AOM are still not completely understood. For microbial consortia consisting of archaea and bacteria, classical syntrophic electron transfer through the exchange of solutes (MIET) was proposed early on, where H_2, formate, acetate, have been considered as intermediates that diffuse from methane oxidizing archaea to sulfate reducing bacteria. Alternatively, Milucka et al. (2012) proposed the exchange of disulfide, produced by ANME archaea coupling methane-oxidation to sulfate-reduction directly and consumed by associated sulfur disproportionating bacteria. More recently, DIET has been proposed as a principal mechanism responsible for electron transfer in ANME-SRB consortia (McGlynn et al. 2015, Wegener et al. 2015). It has been shown that AOM by ANME-2 archaea in deep-sea sediments can be catabolically and anabolically decoupled from sulfate-reduction (Scheller et al. 2016).

We expand on preliminary modeling efforts (Orcutt and Meile 2008; McGlynn et al. 2015) by accounting for the effects of substrate availability on microbial activity, reaction energetics and chemical dynamics in a spatially explicit environmental context. A three-dimensional reaction transport model is implemented to evaluate these potential mechanisms governing AOM within the archaeal-bacterial consortia. In our model, each of these syntrophic mechanisms was implemented in a reactive transport model and the simulated activities were compared to empirical data for AOM rates and intra-consortia spatial patterns of cell-specific anabolic activity determined by FISH-nanoSIMS.

Simulated rates based on MIET were limited by the build-up of metabolites, making the reaction energetically unfavorable. The disulfide and DIET pathways yielded AOM rates that were consistent with measured values and produced intra-consortium anabolic activity distributions consistent with cell specific nanoSIMS data. Factors that significantly impacted DIET simulation results included the AOM rate constant, the concentration of redox molecules involved in cell-to-cell electron transport, and the relative contribution of conduction to the total exchange of electrons between microbial partners. Our simulations predict that electron transport and changes in pH caused by methane oxidation affected AOM only at rates exceeding those reported from methane seep environments. However, additional simulations with artificially accelerated proton diffusion allowed for higher rates, indicating the potential role of pH variations to limit AOM at higher rates and large aggregates. Consistent with these observations, our HS_2^- and DIET simulations suggest that AOM rates are largely unaffected by the spatial distribution of bacterial and archaeal cells. We conclude that the comparison of our modeling results with the intra-aggregate activity patterns are consistent with observational data presented by McGlynn et al. (2015) and Wegener et al. (2015) pointing towards DIET as a likely the pathway for electron transport within AOM consortia.

References

1. McGlynn, S.E., Chadwick, G.L., Kempes, C.P., and Orphan, V.J. (2015) Single cell activity reveals direct electron transfer in methanotrophic consortia. *Nature* **526:** 531-535.
2. Milucka, J., Ferdelman, T.G., Polerecky, L., Franzke, D., Wegener, G., Schmid, M., and Kuypers, M.M. (2012) Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* **491:** 541-546.
3. Orcutt, B., and Meile, C. (2008) Constraints on mechanisms and rates of anaerobic oxidation of methane by microbial consortia: process-based modeling of ANME-2 archaea and sulfate reducing bacteria interactions. *Biogeosciences* **5:** 1587-1599.
4. Scheller, S., Yu, H., Chadwick, G.L., McGlynn, S.E., and Orphan, V.J. (2016) Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. *Science* **351:** 703-707.
5. Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H.E., and Boetius, A. (2015) Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* **526:** 587-590.

Funding statement.

This work was supported by the Genomic Sciences Program in the DOE Office of Science, Biological and Environmental Research DE - SC0016469. The authors declare no conflict of interest.

A Day in the Life of Chlamydomonas

Daniela Strenkert[1]*(DanielaS@chem.ucla.edu), Stefan Schmollinger[1], Sean D. Gallaher[1], Samuel O. Purvine[2], Carrie Nicora[2], Mary Lipton[2] and **Sabeeha S. Merchant**[1]

[1]Institute for Genomics and Proteomics and Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA;

[2]EMSL, Richland, WA, USA

Project goals: The green alga *Chlamydomonas reinhardtii* has become a model organism for many diverse research fields, including photosynthesis research. But our understanding of how gene expression is orchestrated between different organelles is still surprisingly limited. We generated a multilayer data set of the Chlamydomonas cell cycle, using a flat panel bioreactor system to allow reproducible synchronization of *Chlamydomonas reinhardtii*. We monitored transcript abundance from all three genomes during the cell cycle by applying a modified transcriptomic approach. Parallel measurements of select metabolites and pigments, physiological parameters and a subset of proteins offered the opportunity for inferring metabolic events and for evaluating the impact of the transcriptome on the proteome.

During the cell cycle, cell division and DNA replication is initiated at the beginning of the dark period, in S phase. The high demand of the core histones during S phase is met by the coordinated expression of multiple genes encoding the core histones in the green algae Chlamydomonas. Interestingly, we see replication independent expression of two of each genes, serving as emergency histones. The helicase involved in DNA replication (MCM complex), is preceding the expression of core histones, ensuring that newly synthesized nucleosomes can be loaded on replicated DNA. After replication, cells remain in G_0 for the remainder of the night, using stored carbon sources presumably for respiration. But assessment of starch, total organic carbon and respiratory activity suggested that fermentative metabolism may dominate during the night, and co-expression pattern identified FDX9 as participant. During the light period, expression of plastid encoded subunits of photosynthetic complexes preceded their nucleus encoded counterparts in anticipation of the day. The dark to light transition is accompanied by expression of stress responsive genes. Among these, the pattern of *PSBS* and *LHCSR1* is distinguished from *LHCSR3* expression, whose pattern receives two distinct inputs from light. Although genes for tetrapyrrole biosynthesis are expressed concomitantly with those for chlorophyll-binding proteins,

those for light independent protochlorophyllide oxidoreductases are reciprocally expressed compared to the nucleus encoded, light dependent enzymes. The multi-omics approach offers an unprecedented high resolution systems level view of cellular processes as cells grown in the light period and divide in the dark from one to two cells.

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

A Better Understanding of Bacterial RNA Turnover Developed Through RNA Sequencing

Gina Gordon[1], Jeffrey C. Cameron,[2] Michael D. Engstrom[1], and **Brian F. Pfleger**[1,*]
(pfleger@engr.wisc.edu)

[1]University of Wisconsin, Madison; [2]University of Colorado, Boulder

URL: http://pflegerlab.che.wisc.edu/

Project Goals:

For this project, we use RNA-sequencing techniques to study mRNA turnover and stability in *Escherichia coli* and *Synechococcus* sp. PCC7002, a cyanobacteria with promising industrial traits that could be used as a renewable route for producing fuels and chemicals. We use RNA sequencing to study how various RNA processing enzymes affect decay of mRNA from a global perspective. We use the knowledge gained from our studies of mRNA turnover to develop design rules for (de)stabilizing transcripts and apply these rules to ongoing metabolic engineering projects in PCC 7002 for use in photosynthetic biorefineries to sequester CO_2 that would otherwise contribute to climate change.

Our specific objectives are:

(1) Quantify decay rates for each nucleotide in bacterial transcriptome with RNA sequencing
(2) Design and test strategies for (de)stabilizing transcripts in PCC 7002
(3) Apply design rules to improving biofuel production in PCC 7002

Messenger RNA (mRNA) is a labile intermediate that affects protein expression levels. For metabolic engineering purposes we would ideally be able to predict and precisely control protein expression based on a given DNA sequence. Unfortunately there is insufficient foundational knowledge about RNA stability to be able to predict the half-life of a given transcript. This arises due to the numerous and complex ribonucleases (RNases) that facilitate mRNA degradation. We are particularly interested in RNase III, a ribonuclease that cleaves long double-stranded regions of RNA, because of its potential use to alter transcript stability and process structured RNAs (e.g. CRISPR guide RNAs). RNase III recognizes mRNA secondary structure and not a conserved nucleotide sequence, but it is still more amenable to engineering than other RNases that have less selectivity. Using RNA-seq we have identified novel targets of RNase III in *Escherichia coli*, and we have shown that RNase III regulates protein expression of important metabolic enzymes, processes read-through transcripts, and is involved in the turnover of mRNA of leader peptides. With our greatly expanded list of RNase III sites and sequences, we will test if these sites can be placed next to genes of interest to alter their stability and subsequent protein expression.

The genome of the cyanobacterium *Synechococcus sp.* strain PCC 7002 encodes three homologs of RNase III, two full-length and one mini-III, that are not essential even when deleted in combination. To discern if each enzyme had distinct responsibilities, we collected and sequenced global RNA samples from the wild type strain, the single, double, and triple RNase III mutants. Approximately 20% of genes were differentially expressed in various mutants with some operons and regulons showing complex changes in expression levels between mutants. Two RNase III's had a role in 23S rRNA maturation and one of those was also involved in copy number regulation one of six native plasmids. *In vitro*, purified RNase III enzymes were capable of cleaving some of the known *E. coli* RNase III target sequences, highlighting the remarkably conserved substrate specificity between organisms yet complex regulation of gene expression.

Publications from the project

1. Cameron JC, Gordon GC, Pfleger BF. Genetic and genomic analysis of RNases in model cyanobacteria. *Photosynthesis Research.* Oct;126(1):171-83 (**2015**). DOI: 10.1007/s11120-015-0076-2
2. Markley AL, Begemann MB, Gordon GC, **Pfleger BF**. A synthetic biology toolbox for controlling gene expression in the cyanobacterium *Synechococcus* sp. PCC 7002. *ACS Synthetic Biology*. May 15;4(5):595-603 (**2015**). DOI: 10.1021/sb500260k
3. Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemann MB, Pfleger BF. CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Metabolic Engineering*. Jul 29. pii: S1096-7176(16)30062-3 (**2016**) DOI: 10.1016/j.ymben.2016.07.007
4. Pfleger BF. Microbes Paired for Biological Gas to Liquids (Bio-GTL) Process. Proceedings of the National Academy of Science USA. Apr 5;113(14):3717-9. (**2016**). DOI: 10.1073/pnas.1601926113
5. Gordon GC, Cameron JC, Pfleger BF. RNA-sequencing identifies new RNase III cleavage sites in Escherichia coli and reveals increased regulation of mRNA. *mBio*. Mar 28;8(2). pii: e00128-17. (**2017**) DOI: 10.1128/mBio.00128-17
6. Wang C, Pfleger BF, Kim SW. Reassessing *Escherichia coli* as a cell factory for biofuel production. *Current Opinion in Biotechnology*. Mar 11;45:92-103. (**2017**) DOI: 10.1016/j.copbio.2017.02.010
7. Gordon GC, Cameron JC, Pfleger BF. Distinct and redundant functions of three homologs of RNase III in the Cyanobacterium *Synechococcus* sp. PCC 7002 (*Revision submitted*)
8. Clark RL, Gordon GC, Bennett NR, Lyu H, Root TW, Pfleger BF. High-CO2 Requirement as a Mechanism for the Containment of Genetically Modified Cyanobacteria. (*Revision submitted*)
9. Gordon GC, Pfleger BF. Regulatory tools for controlling gene expression in cyanobacteria. (*invited book chapter – Submitted*)
10. Gordon GC, Cameron JC, Pfleger BF. Analysis of global RNA turnover in cyanobacteria. (*in preparation*)

Funding Statement:

This work is supported by US-Department of Energy (DE-SC0010329) and a NIH Biotechnology Training Fellowship to Ms. Gina Gordon (NIGMS- 5 T32 GM08349).

Quantification of Multiple Post-Translational Modifications by 'One-Pot' Affinity Enrichment - Applications in *E. coli*

Nathan Basisty[1], Jesse G Meyer[1], Lei Wei[1], Bradford W Gibson[2], Birgit Schilling[1]* (bschilling@buckinstitute.org)

[1]The Buck Institute for Research on Aging, 8001 Redwood Blvd, Novato, CA 94945
[2] Amgen, 1120 Veterans Blvd, South San Francisco, CA 94080

Project Goals:

The goal of this project is to gain deeper insights into posttranslational remodeling of engineered microorganisms, and specifically their protein acylomes under different growth conditions. Lysine acetylation, for example is a common post-translational modification (PTM) that eukaryotes, archaea, and bacteria employ to regulate protein activity. In fact, most metabolic enzymes are subject to lysine acetylation. Due to the dynamic nature of protein acetylation and deacetylation mechanisms in the cell, lysine acetylation can likely be considered a global mechanism to regulate metabolism in response to their energy and redox status. Here, we are presenting a novel workflow that will enrich multiple posttranslational modifications, such as acetylation and succinylation in a 'one-pot' affinity enrichment procedure, which will greatly improve throughput and enable PTM crosstalk. The significance of this work is that it will provide new tools to address a fundamental gap in our understanding of bacterial metabolism and identify new approaches for overcoming the problems associated with the production of advanced biofuels.

Abstract:

Quantitative proteomic studies of post-translational modifications (PTMs) are increasingly of widespread interest in biomedical research; however, rarely do these studies consider multiple modifications in parallel. One barrier to multi-PTM studies is the time costs for both sample preparation and instrument acquisition, which scale linearly with the number of modifications studied. Perhaps the most prohibitive requirement of PTM studies is the need for large amounts of sample material, which typically must be increased proportionally with the number of PTM enrichments. Here, we describe an innovative, streamlined, quantitative label-free proteomic workflow ("one-pot" enrichment), which allows for comprehensive identification and quantification of acetylated and succinylated peptides from a single sample containing 1 mg of mitochondrial protein. We show that simultaneous immunoaffinity enrichment of acetylated and succinylated peptides by 'one-pot' enrichment is the most efficient enrichment method, requiring the least sample preparation and instrument time without compromising data quality. We further show that coupled with a label-free, data-independent acquisition (DIA, e.g. SWATH), 'one-pot' enrichments from frozen mouse liver samples could identify and quantify 2235 acetylated and 2173 succinylated peptides from small amounts of input protein. We also demonstrate that the peak areas

of sites identified following one-pot enrichment are highly correlated with single-antibody pull downs. Finally, we show that this method makes it possible to detect both acetylation and succinylation modifications that occur on the same peptide. One-pot enrichment is a novel quantitative PTM workflow for enriching post-translational modifications as reproducibly as single-antibody enrichments, and enables the direct assessment of PTM crosstalk from biological samples with limited tissue material. This improvement in technology and workflow will be highly relevant for our acylation studies in *E. coli*, and will allow us to achieve higher throughput and better understanding of multiple PTM regulation during bacterial growth under diverse environmental and nutritional conditions.

References:

- JG Meyer, S Mukkamalla, H Steen, AI Nesvizhskii, BW Gibson, B Schilling. "PIQED: automated identification and quantification of protein modifications from DIA-MS data", Nat Methods, 2017, 14: 646-47.
- ES Nakayasu, MC Burnet, HE Walukiewicz, CS Wilkins, AK Shukla, S Brooks, MJ Plutz, BD Lee, B Schilling, AJ Wolfe, S Muller, JR Kirby, CV Rao, JR Cort, SH Payne. "Ancient Regulatory Role of Lysine Acetylation in Central Metabolism", MBio, 2017, 8.
- R Davis, A Ecija-Conesa, J Gallego-Jara, T de Diego, EV Filippova, G Kuffel, WF Anderson, BW Gibson, B Schilling, M Canovas, AJ Wolfe. "An acetylatable lysine controls CRP function in *E. coli*", Mol Microbiol. 2017, in press (DOI: 10.1128/mBio.01894-17).
- N Basisty, JG Meyer, L Wei, BW Gibson, B Schilling "Quantification of Multiple Post-Translational Modifications by 'One-Pot' Affinity Enrichment", Nature Methods, 2017, submitted.

Funding Statement:

This work was supported by the U.S Dept. of Energy, Office of Science, Office of Biological & Environmental Research (PI: Rao, DE-SC0012443), grant title 'The Systems Biology of Protein Acetylation in Fuel-Producing Microorganisms'. We also acknowledge the support of the NIH shared instrumentation program for the use of the SCIEX TripleTOF 6600 mass spectrometer (1S10 OD016281). N.B. was supported by a postdoctoral fellowship from the Glenn Foundation for Medical Research. J.G.M. was supported by a US NIH grant (T32 AG000266).

Nanoscale Stable Isotope Tracing and Compound-Specific Radioisotope Analysis to Investigate Metabolic Interactions between Bacteria and Biofuel-producing Algae

Ty Samo[1], Benjamin Stewart[1], Jeffrey Kimbrel[1], Peter K. Weber[1], and **Xavier Mayali (mayali1@llnl.gov)**[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 investigate the influence of algal-associated bacteria on photosynthetic biomass production, we are examining interactions between the model diatom *Phaeodactylum tricornutum* CCMP2561 and its associated microbiome. Using an enrichment process on outdoor pond samples of *Phaeodactylum* sp., we obtained simplified bacterial communities as well as single bacterial isolates that influence the growth of *P. tricornutum* in the laboratory under different temperature, light, and nutrient concentrations. We aim to better understand 1) the role of physical interactions (i.e., attachment) in the remineralization of excreted dissolved organic matter for improved algal productivity as well as 2) develop new approaches to measure the production and fate of metabolites that influence algal productivity (e.g. plant growth hormones and vitamins).

To address the first goal, we are using the Nano-SIP approach (Pett-Ridge and Weber, 2012) to examine cell-specific carbon fixation and carbon remineralization. This involves adding stable isotope labeled substrates followed by NanoSIMS quantification of isotope incorporation at the single cell level (Figure 1). Using ^{13}C labeled bicarbonate addition experiments, we quantified the carbon fixation activities of the algal cells when incubated on their own and when co-incubated with different bacterial strains. In the co-cultures, the high spatial resolution of the NanoSIMS enabled us to identify algal cells with and without attached bacteria, as well as quantify the amount of fixed carbon incorporated by both attached and non-attached bacteria. Our data show that a subset of the bacterial isolates can increase the carbon fixation activities of *P. tricornutum* cells, and that bacterial attachment plays a critical role in this process, likely

through remineralization of excreted fixed C back into CO_2. We further quantified the remineralized C and N incorporated by the algal cells through similar NanoSIMS analyses of one bacterial-algal co-culture (strain Pt3-2) incubated with ^{15}N and ^{13}C labeled algal exudate.

To address the second goal of metabolite tracing, we are using Parallel Accelerator Mass Spectrometry- Molecular Mass Spectrometry (PAMMS), which involves incubation of cells with ^{14}C-labeled precursors and parallel analysis by liquid chromatography time of flight mass spectrometry (LC-TOFMS) and accelerator mass spectrometry (AMS). This allows the simultaneous interrogation of metabolites, providing identification and quantitation of ^{14}C signal. One well characterized mechanism of algal growth promotion by bacteria centers on the compound indole-3-acetic acid (IAA). Using algal-produced tryptophan, bacteria produce and release this metabolite, which is incorporated by algae and stimulates growth by upshifting cell cycle progression (Amin et al., 2015). To examine this pathway, we carried out separate incubations with ^{14}C-tryptophan in two bacterial strains isolated from *Phaeodactylum* ponds to trace both the intracellular and excreted fate of this precursor molecule in the bacterial metabolomes. There was no evidence for excretion of tryptophan-derived metabolites by these two bacterial cultures. In one of the strains (*Marinobacter* sp. 19DW), the radiolabel was not traced into any metabolites, showing that it remained as tryptophan or was directly incorporated into protein. In *Algoriphagus* sp. ARW1R1, however, the radiolabel was traced into several unidentified metabolites as well as into 2-amino-3-carboxymuconate semialdehyde, an intermediate of the niacin vitamin biosynthetic pathway. These results show that *Algoriphagus* metabolizes tryptophan into compounds potentially influencing algal health and productivity.

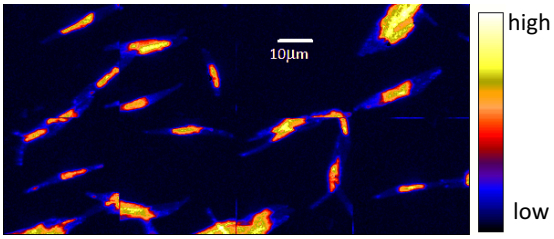


FIGURE 1: NanoSIMS $^{13}C/^{12}C$ isotope map of *Phaeodactylum* cells incubated with bacterial isolate *Muricauda* ARW7G5W

Our data suggest that bacterial taxonomic identity, including the presence of specific biosynthetic pathways, as well as physical proximity to the algal cells play critical roles in mediating the growth of algae. Future plans aim to incorporate such data into metabolic and trait-based modeling efforts to ultimately predict the impact of bacterial metabolism on biofuel production.

References

Amin, S., Hmelo, L., Van Tol, H., Durham, B., Carlson, L., Heal, K., Morales, R., Berthiaume, C., Parker, M., and Djunaedi, B. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522**,** 98-101.

Pett-Ridge, J., and Weber, P.K. (2012). "NanoSIP: NanoSIMS applications for microbial biology," in *Microbial Systems Biology: Methods and Protocols,* ed. A. Navid. Humana Press.

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

Regulation of acetate metabolism by pH in *Escherichia coli*

James S Orr[1]* (jsorr3@illinois.edu), David Christensen[2], Alan J Wolfe[2], and **Christopher V Rao**[1]

[1]University of Illinois at Urbana-Champaign; [2]Loyola University Chicago

http://www.scs.illinois.edu/rao/index.php

Project Goals: The goal of this project is to determine how protein acetylation affects metabolism in bacteria. Lysine acetylation is a common post- translational modification that eukaryotes, archaea, and bacteria employ to regulate protein activity. Multiple studies have recently shown that lysine acetylation predominantly targets metabolic enzymes – in fact, most metabolic enzymes are subject to lysine acetylation. We hypothesize that bacteria employ lysine acetylation as a global mechanism to regulate metabolism in response to their energy and redox status. Our previous work suggests that lysine acetylation may be an attractive and innovative target for metabolic engineering. We are investigating how lysine acetylation affects bacteria metabolism. The significance of this work is that it will address a fundamental gap in our understanding of bacterial metabolism and identify new approaches for overcoming the problems associated with the production of advanced biofuels.

During growth on excess sugars, the state of growth for most fermentation, *Escherichia coli* will produce acetate due to overflow metabolism. Following consumption of the sugar, *E. coli* will then reassimilate the extracellular acetate. Acetate metabolism is closely connected to protein acetylation, because acetyl phosphate, an intermediate in acetate metabolism, serves as the principal acetyl-group donor in lysine acetylation.

In the project, we investigated extracellular acetate production and reassimilation during growth at neutral and acidic pH's. We found that pH does not affect the production of extracellular acetate during growth on glucose. However, we found that *E. coli* is unable to reassimilate the extracellular acetate during growth under acidic conditions (pH < 7). These results demonstrate that acidic pH's inhibit the ability of *E. coli* to consume – but not produce – extracellular acetate.

To identify the mechanism, we investigated a number of mutants in which the different pathways for acetate metabolism were selectively deleted. These mutants also failed to reassimilate acetate under acidic growth conditions. In addition, we investigated the transcriptional regulation of the acetate metabolic genes and found that their expression was not sensitive to pH. We also investigated the impact of pH on the reversibility of the AckA/Pta pathway using purified enzymes. Much of acetate produced by *E. coli* during growth on glucose is produced via the reversible AckA/Pta pathway. We found lower pH drove acetate in the direction of consumption and inhibited its production in both proteins. This result demonstrates that differences in equilibrium acetate from AckA/Pta cannot explain the pH-dependent reassimilation of acetate. Our current hypothesis is that pH-dependent reassimilation of acetate is due to the repression of the glyoxylate bypass during growth in acidic conditions. We are currently testing this hypothesis.

This work was supported by the U.S Dept. of Energy, Office of Science, Office of Biological & Environmental Research (PI: Gibson, Rao, Wolfe DE-SC0012443), grant title 'The Systems Biology of Protein Acetylation in Fuel-Producing Microorganisms' (ER213630).

Microbial Community Assembly and Evolution Analyzed at Single Cell Level

Serdar Turkarslan[2] (sturkarslan@systemsbiology.org), *Nejc Stopnisek*[1] (nejcstopno@gmail.com), Anne W. Thompson[6], Aifen Zhou[3], Annie Otwell[2], Kristina J. Hillesland[7], Frederick von Netzer[1], Jacob Valenzuela[2], Maryann Watkins[2], Nicholas Elliott[1], Peter J. Walian[5], Manfred Auer[5], Jizhong Zhou[3], Grant M. Zane[4], Judy D. Wall[4], David A. Stahl[1], Nitin S. Baliga[2]*(nbaliga@systemsbiology.org), **Adam P. Arkin[5] and Paul D. Adams[5]**

[1]University of Washington, Seattle; [2]Institute for Systems Biology, Seattle; [3]University of Oklahoma; [4]University of Missouri, Columbia; [5]Lawrence Berkeley National Laboratory, Berkeley; [6]Portland State University, Portland; [7]University of Washington Bothell, Bothell;

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. Understanding the general principles and mechanisms for community assembly, evolution and collapse across different environments and scales.

A two-organism model community was previously established under sulfate-deplete conditions that prevented independent growth of *Desulfovibrio vulgaris* (Dv) by sulfate respiration (SR), and required its syntrophic (ST) interaction with *Methanococcus maripaludis* (Mm) to support growth of both organisms. Experimental evolution of this model community in a uniform resource environment significantly improved growth characteristics across all 12 lines at the expense of erosion of SR. On the other hand, growth in a fluctuating resource environment that required the community to switch back and forth between SR and ST led to population collapse. We probed the evolution and collapse phenomena by performing bulk and single cell genomics and transcriptomics across various stages of evolution, and compared these data with similar analyses of phenotypically characterized simplified communities, and clonal isolates. This multiscale systems analyses revealed that during evolution of obligate ST in a uniform resource environment, many independent adaptive strategies had simultaneously co-evolved within the same evolution line. This finding was confirmed upon observing that distinct pairings of Dv and Mm obtained through end-point dilutions (EPDs) from the same evolution line produced higher growth rates and yields, relative to cross-pairings of the isolates across EPDs. Multiscale analysis of the community subjected to the fluctuating environment on the other hand revealed that the phenomenon of collapse had occurred due to excessive regulation of genes, which had driven progressive dilution of essential transcripts and proteins. Together, these studies showcase how multiscale systems analysis can reveal signatures, and timelines of adaptive molecular events and the mechanisms by which they shape the assembly, evolution and collapse of microbial communities.

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

Metabolic phenotyping of cyanobacteria with increased biofuel productivity

Yi Ern Cheah[1]* (yi.ern.cheah@vanderbilt.edu), Yao Xu[2], Carl H. Johnson[2], **Jamey D. Young**[1,3]

[1] Chemical & Biomolecular Engineering, Vanderbilt University, Nashville, TN; [2] Biological Sciences, Vanderbilt University, Nashville, TN; [3] Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN

http://www.vanderbilt.edu/younglab

Project Goals: This project aims to identify and remove bottleneck reactions that limit carbon flux towards synthetic pathways in engineered autotrophic systems. The long-term goal is to develop an integrated experimental and computational workflow that can be progressively used to enhance the performance of industrially pertinent autotrophs.

To quantitatively assess *in vivo* autotrophic metabolism, our lab has previously developed isotopically nonstationary ^{13}C-MFA (INST-MFA) (*1-4*) and applied it to assess the photoautotrophic metabolism of cyanobacteria (*4*) and plant leaves (*5*). More recently, we combined INST-MFA with rational metabolic engineering to improve the productivity of an isobutyraldehyde producing mutant of the cyanobacterium *Synechococcus elongatus* PCC7942 (*6*).

This presentation describes our further efforts to examine the metabolic phenotypes of strains with improved isobutyraldehyde productivity. Our work demonstrates the promising utility of INST-MFA at guiding the metabolic engineering of autotrophs intended for industrial applications.

References

1. L. J. Jazmin, J. D. Young, in *Systems Metabolic Engineering: Methods and Protocols,* H. S. Alper, Ed. (Humana Press, Totowa, NJ, 2013), pp. 367-390.

2. A. Adebiyi, L. Jazmin, J. Young, 13C flux analysis of cyanobacterial metabolism. *Photosynthesis Research*, 1-14 (2014).

3. J. D. Young, INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* **30**, 1333-1335 (2014).

4. J. D. Young, A. A. Shastri, G. Stephanopoulos, J. A. Morgan, Mapping photoautotrophic metabolism with isotopically nonstationary C-13 flux analysis. *Metabolic Engineering* **13**, 656-665 (2011).

5. F. F. Ma, L. J. Jazmin, J. D. Young, D. K. Allen, Isotopically nonstationary C-13 flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 16967-16972 (2014).

6. L. J. Jazmin *et al.*, Isotopically nonstationary 13C flux analysis of cyanobacterial isobutyraldehyde production. *Metabolic Engineering* **42**, 9-18 (2017).

This work is funded by the U.S. Department of Energy (DOE) Award DE-SC008118.

Visualizing Spatial Metabolic Interactions Within a Soil Microbiome

Christopher R. Anderton[1] (christopher.anderton@pnnl.gov), Natalie C. Sadler[2], Sheryl L. Bell[1], Dušan Veličković[1], Rosalie K. Chu[1], Thomas W. Wietsma[1], Pubudu P. R. Handakumbura[1], **Kirsten S. Hofmockel[1], Janet K. Jansson[2]**

[1]Environmental Molecular Sciences Division; [2]Biological Sciences Division, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

https://www.pnnl.gov/biology/programs/

Project Goals: PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture through spatially explicit examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments will be designed to confront both the scaling challenges and inter-kingdom interactions that regulate networks of biochemical reactions. Individual- and population-based models for predicting interspecies and inter-kingdom interactions will be parameterized using experimental data, and predictions will be tested in soil to reveal spatially explicit microbial interactions. Discoveries from controlled experiments will be tested and validated in the field, using moisture gradient experiments at a new local field site. Data will be captured and shared through the establishment of a Soil Microbiome Knowledgebase (SMK). 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:

Understanding the basic biology that underpins soil microbiome interactions is required in order to predict the metaphenome response to environmental shifts, such as changing moisture content. A significant knowledge gap is how such changes will affect microbial community structure and its metabolic landscape. We aim to visualize the metabolome of interacting organisms within the soil habitat by attaining high resolution multidimensional maps of the compositional and functional state of soil microbial communities. This entails mapping the metabolic exchanges that occur within soil microbiomes, wherein historically it has been exceedingly difficult to measure the biochemical currency among interacting community members within a soil system.[1] For example, traditional metabolomic approaches are often limited in their ability to distinguish between molecules that remain localized within microbes and exuded molecules that are in proximity, and thus often disregard the multifaceted chemical exchange within and between interacting

species. However, visualizing metabolic interactions between interacting organisms within environmental microbiomes with unmatched sensitivity and specificity can now be accomplished using mass spectrometry imaging (MSI) methodologies we recently developed.[2,3] We are able to attain high confidence in both molecular identification and localization, offering unprecedented insights into the metabolic interactions of an inter-kingdom interaction (e.g., changes in disaccharide synthesis).[3] We will utilize these methods in a multimodal fashion with optical microscopy approaches, capable of visualizing desired taxa, in order to understand how change in soil moisture content will modify the soil microbiome community organization and its spatial metabolome.

References

1. D. Veličković, C. R. Anderton, "Mass spectrometry imaging: Towards mapping the elemental and molecular composition of the rhizosphere," *Rhizosphere* **2017**, *3*, 254
2. C.R. Anderton, R.K. Chu, N. Tolic, A.V. Creissen, L Pasa-Tolic, "Utilizing a Robotic Sprayer for High Lateral and Mass Resolution MALDI FT-ICR MSI of Microbial Cultures," *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 556
3. 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.*

In situ demonstration of sustained adaptation of a natural microbial community to transform substrates

Charles J. Paradis[1,2], Terry C. Hazen[1, 2]*(tchazen@utk.edu), **Adam P. Arkin[3], and Paul D. Adams[3]**

[1]University of Tennessee, Knoxville, TN, [2]Oak Ridge National Laboratory, Oak Ridge, TN, [3]Lawrence Berkeley National Laboratory, 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. For any microbe, it is possible to discover efficiently the genetic determinants of adaptation to life in dynamic environments in contact with other living members of its ecology and reveal hidden functional phenomena that are only possible by using a systems biology approach; For any environment, it is possible to predict the temporal changes in geochemistry with some precision, for some length of time, given geochemical and biological inputs.

Here we aim to: (1) demonstrate the exposure history dependence of microbial mediated substrate transformation rates in groundwater at the field scale and (2) elucidate the microbial mechanism(s) which control the exposure history dependence of microbial mediated substrate transformation rates

Prior exposure of a natural microbial community to a substrate can result in the increased potential of the community to transform the substrate; this phenomenon is known as adaptation. Adaptation is thought to play an important role in biogeochemical cycling at the ecosystem scale and has been demonstrated at the laboratory scale. However, in situ demonstrations of the magnitude and duration of adaptation are lacking. Ethanol was used as a substrate and was injected into a groundwater well (substrate treatment) for six consecutive weeks to establish adaptation. A second well (substrate control) was not injected with ethanol during this time. The substrate treatment demonstrated adaptation for microbial-mediated oxidation of ethanol to acetate and reduction of nitrate and sulfate as evident by sequential and significant increases in zero-order reaction rates. Both wells were then monitored for six additional weeks under natural conditions. During the final week, ethanol was injected into both wells. The substrate treatment

demonstrated sustained adaptation as evident by significantly higher reaction rates than the substrate control. Surprisingly, the selective enrichment of a microbial community within the first six weeks of the substrate treatment was not sustained after the six-week absence of ethanol, as revealed by analysis of planktonic DNA. These results demonstrated that adaptation can be induced and sustained with no apparent enrichment of a select microbial community. This suggests that the predominant mechanisms of adaptation may exist at the enzymatic- and/or genetic-levels.

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

Title: Are the activities of microbial taxa consistent across ecosystems?

Ember M. Morrissey*[1] (ember.morrissey@mail.wvu.edu), Rebecca Mau[2], Xiaojun Liu[2], Benjamin J. Koch[2], Jennifer Pett-Ridge[4], Steve Blazewicz[4], Xavier Mayali[4], Kirsten Hofmockel[5], Egbert Schwartz[2,3], Paul Dijkstra[2,3], **Bruce A. Hungate**[2,3].

[1] Division of Plant and Soil Sciences, West Virginia University, Morgantown, West Virginia
[2] Center of Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona
[3] Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona
[4] Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California
[5] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington

Project Goals: Connecting the composition of microbial communities with biogeochemical process rates has the potential to improve our understanding of, and ability to model, ecosystem function. This project aimed to determine if the growth of prokaryotic taxa is consistent across ecosystems. This work dovetails with our our larger goal to characterize *in situ* rates of biogeochemically significant microbial activity at the community scale and for specific taxa. This information will be leveraged to establish whether there is a "phylogenetic imprint" on soil carbon and nitrogen cycling processes that can facilitate better incorporation of microbial data into process-scale modeling efforts.

Soils are a huge reservoir of carbon, yet our ability to predict how this important carbon pool will respond to future climate remains poor. Understanding how soil microbial community composition influences ecosystem biogeochemistry could help resolve this knowledge gap. Our past work demonstrates clear taxonomic patterns in the activity of microorganisms; for instance, some taxonomic groups grow quickly while others grow slowly or not at all. These taxonomic patterns in activity will be most useful for modeling the influence of microbial biodiversity on ecosystem function if they appear consistently across ecosystems. Here we aimed to test the hypothesis that taxonomic identity can explain a significant amount of the variation in prokaryotic growth and carbon assimilation across ecosystems. To this end, we used qSIP with ^{18}O-water and ^{13}C-glucose to measure microbial growth and carbon assimilation in four ecosystems along a climatic gradient in Northern Arizona. These sites consisted of a mixed conifer forest, a ponderosa pine forest, a pinyon-juniper woodland, and a cool desert grassland. While data analysis is still underway, preliminary results suggest an impact of both taxonomic identity and ecosystem type on the activity of microbial populations in soil. This may indicate limited phenotypic plasticity in the activity of microbial populations wherein microorganisms respond to environmental variation but maintain relatively consistent ecological strategies. Characterizing the activity of microbial taxa could result in a paradigm shift wherein microbes are no longer treated as a 'black box' in biogeochemical models, leading to dramatic improvements in our ability to understand and predict carbon cycling dynamics in soil.

This research was supported by the Office of Biological and Environmental Research in the DOE Office of Science Project Number DE-SC0016207

Exploring Biome-scale Variation in Microbial Taxon-Specific Growth and Mortality at different temperatures through quantitative Stable Isotope Probing with $H_2{}^{18}O$.

Egbert Schwartz[1,2*] (Egbert.Schwartz@nau.edu), Rebecca Mau[1,2], Michaela Hayer[1,2], Ember Morrissey[3], Jennifer Pett-Ridge[4], Steven Blazewicz[4], Xavier Mayali[4], Kirsten Hofmockel[5], **Bruce A. Hungate[1,2]**.

[1]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, [2]Northern Arizona University, Flagstaff, AZ, [3]West Virginia University, Morgantown, WV, [4]Lawrence Livermore National Laboratory, Livermore, CA [5]Pacific Northwest National Laboratory, Richland, WA.

Project Goals:

- To better understand microbial population dynamics and its relationship to microbial energy metabolism and C cycling
- To link processes of individual microbial species to whole system C and element cycling and their responses to temperature
- To discover global patterns in microbial population dynamics across ecosystems

Abstract text: We are investigating if there is a general phylogenetic signal in the impact of temperature on growth and mortality that is consistent across biomes. We collected soils from a tundra at the Toolik LTER in Alaska, a mixed conifer forest outside Flagstaff AZ, the SPRUCE experiment in northern Minnesota, and the Luquillo Experimental Forest in Puerto Rico. These soils represented the arctic, temperate, boreal, and tropical biomes. Each soil was incubated in the laboratory with 99 atom% $H_2{}^{18}O$ at 5 different temperatures: 5, 15, 25, 35, and 45 °C, for either 5 or 10 days. Subsequently, DNA was extracted from the soils and used in quantitative stable isotope probing (qSIP). The data from the qSIP incubations will be used to calculate both taxon-specific and aggregate or average community rates of growth and mortality. Preliminary findings indicate that, while respiration rates increased with temperature, microbial growth was increasingly delayed and suppressed with higher temperatures. Particularly the tundra soil from Alaska was negatively impacted when incubated at 45 °C, and very little DNA was recovered from these samples suggesting substantial microbial death. We are keen to identify microbial populations that have a consistent response to temperature across the four different biomes. These microbial populations support the notion of microbial functional groups, analogous to plant functional groups that have proven useful for modeling plant processes at the ecosystem and global scales.

Field Measurements of Taxon-Specific Microbial Growth in Soil at Two Elevation Gradient Sites Using Quantitative Stable Isotope Probing (qSIP)

Alicia M. Purcell[1,2]* (amp753@nau.edu), Michaela Hayer[1],Benjamin J. Koch[1], Rebecca L. Mau[1], Egbert Schwartz[1,2], and **Bruce A. Hungate[1,2]**

[1]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ
[2]Department of Biology, Northern Arizona University, Flagstaff, AZ

Project Goals: The goal of this study was to determine environmentally relevant measurements of microbial growth using a quantitative stable isotope probing technique with ^{18}O-H_2O. Growth in intact soils in the field was compared with laboratory-based soil incubations, and the response of microbial growth to warming along an elevation gradient was determined.

Measuring microbial growth in environmental samples remains challenging. Determining what microbes are present, their activity, and rates of activity and growth is critical to inform their role in element and nutrient cycling, ecosystem function, and responses to environmental change. Attempts to measure *in situ* microbial cell replication include stable isotopically labeled substrate (^{13}C,^{18}O) enrichment into DNA, algorithms based on metagenome and genome sequence coverage, cell enumeration with microscopy or flow cytometry, and ^{3}H and ^{14}C substrate incorporation into cells. Many techniques, especially for substrate incorporation experiments, rely on laboratory incubations where the sample becomes highly disturbed and perhaps might not capture activity accurately. Here, we aimed to measure *in situ* microbial taxon-specific growth in disturbed and undisturbed soil as well as growth in warmed soil cores. Field and laboratory-based incubations for quantitative stable isotope probing (qSIP) with ^{18}O-H_2O were conducted to determine microbial growth. qSIP involves a combination of techniques involving CsCl density gradient separation, quantitative PCR to determine 16S rRNA gene copy number, and 16S rRNA gene sequencing, all combined to obtain microbial taxa specific growth and death. ^{18}O-H_2O was injected into undisturbed soil in a temperate mixed conifer soil and a ponderosa pine forest soil, two sites part of the C. Hart Merriam Elevation Gradient in Northern Arizona. The soil was left to incubate at field conditions for three and ten days. Parallel ten-day incubations from the mixed conifer soil were incubated in the laboratory to compare microbial taxa growth in disturbed soil incubations versus field conditions. This study, using the novel qSIP technique in a field incubation will begin to elucidate potential laboratory artifacts in studying microbial activity and give more environmentally relevant measures of microbial replication. Preliminary results of this experiment will be discussed.

Funding Statement: This work was supported by the Department of Energy's Program in Genomic Sciences.

m-CAFÉs EcoFABs: Model Ecosystems to Advance Microbiome Science I

Peter F. Andeer*[1](pfandeer@lbl.gov), Joelle Schlapfer[1], Peter Kim[2], Jens Heller[1], Trenton Owens[1], Jill Banfield[3], Eoin L. Brodie[1], Adam M. Deutschbauer[1], Jennifer A. Doudna[3], Mary K. Firestone[3], Peter S. Nico[1], Anup K. Singh[2], Rodolphe Barrangou[4], Trent R. Northen[1], and **N. Louise Glass[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]Sandia National Laboratory, Livermore; [3]University of California, Berkeley; [4]North Carolina State University, Raleigh

http://eco-fab.org

Project Goals: To derive mechanistic understandings of plant-microbe-soil interactions using reproducible, simplified ecosystems

The m-CAFÉs program is a collaborative, coordinated and integrated, mission-driven program to interrogate the function of soil microbiomes with critical implications for carbon cycling and sequestration, nutrient availability and plant productivity in natural and managed ecosystems. One of the major challenges for generalized understanding of microbiomes in these complex ecosystems is extreme variation in microbiomes and environmental conditions. In many fields, simplified model systems have been developed and adopted by many researchers to accelerate the study various aspects of biology. In contrast, there is no agreed upon model system for studying soil microbiomes, and thus, nearly every researcher in the field is studying a different soil system. This heterogeneity of study systems and an inability to replicate experiments in different laboratories limits determination of causal mechanisms and the ability for scientists to build on each other's work.

Beneficial plant-microbe interactions offer a sustainable biological solution to potentially boost low-input food and bioenergy production. However, to maximize the benefit of these plant-microbe interactions, better mechanistic understandings of these complex ecosystems are needed. Here, we present a detailed description for the production of controlled laboratory habitats (EcoFABs - http://eco-fab.org/) for creating and manipulating plant-microbe ecosystems under specific environmental conditions. EcoFABs consist of an autoclavable microfluidic chamber and modular, expandable housing units to accommodate a variety of plant species that are produced using widely accessible materials and technologies (i.e., PDMS structures from 3D-printed molds) which allows for reproducible construction of these devices without a large investment in specialized equipment. These flow-through devices allow for controlled manipulation and sampling of root microbiomes, root chemistry as well as imaging of plant root zones to examine root morphology and localize its microbial communities. This protocol includes details for the construction and sterilization of the EcoFAB system as well as methods for the manipulation of the EcoFAB chamber to vary the media, including soils, sand and liquid growth media. The versatility in the analyses methods and environmental composition available in these systems enable dynamic and detailed investigation of plant and plant-microbial consortia

including manipulation of microbiome composition (including mutants), monitoring of plant growth, root morphology, exudate composition, and microbial localization under controlled environmental conditions.

Funding statement.

This material by m-CAFÉs 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-CAFÉs -EcoFABs: Model Ecosystems to Advance Microbiome Science II

Trent Northen[1]* (TRNorthen@LBL.gov), Joelle Schlapfer Sasse[1], Peter F. Andeer[1], Peter Kim[2], Jens Heller[1], Trenton Owens[1], Andrew Klein[5], Borjana Arsova[6], Josefine Kant[6], Benjamin Cole[1], Michelle Watt[6], John Vogel[1], Eoin L. Brodie[1], Jeff Dangl[5], Adam M. Deutschbauer[1], Jennifer A. Doudna[3], Jill Banfield[3], Mary K. Firestone[3], Peter S. Nico[1], Anup K. Singh[2], Rodolphe Barrangou[4], and **N. Louise Glass[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]Sandia National Laboratory, Livermore; [3]University of California, Berkeley; [4]North Carolina State University, Raleigh; [5]University of North Carolina at Chapel Hill, NC, USA, [6]Institut für Bio-& Geowissenschaften IBG-2, Forschungszentrum Jülich, Germany

http://eco-fab.org

Project Goals: To derive mechanistic understandings of plant-microbe-soil interactions using reproducible, simplified ecosystems

The m-CAFÉs program is a collaborative, coordinated and integrated, mission-driven program to interrogate the function of soil microbiomes with critical implications for carbon cycling and sequestration, nutrient availability and plant productivity in natural and managed ecosystems. One of the major challenges for generalized understanding of microbiomes in these complex ecosystems is extreme variation in microbiomes and environmental conditions. In many fields, simplified model systems have been developed and adopted by many researchers to accelerate the study various aspects of biology. In contrast, there is no agreed upon model system for studying soil microbiomes, and thus, nearly every researcher in the field is studying a different soil system. This heterogeneity of study systems and an inability to replicate experiments in different laboratories limits determination of causal mechanisms and the ability for scientists to build on each other's work.

A major effort within the m-CAFÉs program is developing precisely controlled ecosystem fabrications (EcoFABs) that reflect key functional attributes of plant-microbe interactions within soil, focusing on progressively increasing complexity that honor the physical, chemical and biological properties of soils. Each EcoFAB system is contained within a sterile plant-sized container with independent lighting. 3D printing is used to create root chambers (1.5-5 mL) attached to microscope slides, enabling the use of hydroponics, soil, or sand as substrate, as well as high-resolution imaging. We have shown that these systems are suitable for growth of diverse plants, including *Brachypodium distachyon, Arabidopsis thaliana*, and switchgrass for >1mo. The integrated fluidics system using in the EcoFABs facilitates selective sampling and introduction of microbes, metabolites, etc. Metabolomic analysis of EcoFAB culture is used to examine metabolite exchange within soil and rhizosphere communities and sequencing is used to link activities to specific community members. We have now performed a reproducibility study in which the model grass *Brachypodium distachyon* was grown in three environmental

conditions in four different laboratories. Plant growth, root morphology, phosphate content of tissues, and root exudation were analyzed and compared between the participating laboratories. To interrogate microbial functions in EcoFABs, we are pioneering CRISPR-Cas and environmental RNAi technologies to systematically determine functions of bacteria and fungi in the rhizosphere/soil. The results of these studies will be transformative for our understanding of soil metabolism and microbiome science, with applications to DOE missions in energy and environment.

Funding statement.

This material by m-CAFÉs 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

Integration of metagenomics and consortia data to study microbial interactions and community assembly

L.M. Lui[1*] (lmlui@lbl.gov), H.K. Carlson[1], A.W. Sczesnak[1,2], O. Erbilgin[1], M. De Raad[1], V. Mutalik[1], A. Kazakov[1], X. Wu[1], A. Aaring[1], J. Voriskova[1], J. Kuehl[1], M. Price[1], D. Chivian[1], N.B. Justice[1,2], T. Simmons[1,2], K. De Leon[3], Dominique Joyner[4,5], H. Smith[6], P.S. Novichov[1], A.M. Deutschbauer[1], R. Chakraborty[1], T.R. Northen[1], J.D. Wall[3], T.C. Hazen[4,5], M.W. Fields[6], **A.P. Arkin**[1,2] and **P.D. Adams**[1,2]

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

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.

ENIGMA uses a systems biology approach to build a predictive understanding how phenomena at the genetic, community, ecological, and environmental level influence microbial community assembly and activity. Sampling of sediment and groundwater microbial communities from the Oakridge National Laboratory Field Research Center (ORNL FRC) are coordinated with measurements of biogeochemical parameters to provide a rich set of samples, isolates, and data for integrated analysis across ENIGMA labs. Three categories characterize our efforts to mechanistically understand microbial community assembly: (1) direct analysis of isolate interactions and use of genetic tools to study them, (2) enrichment of communities under controlled laboratory conditions, and (3) analysis of amplicon and whole genome shotgun sequencing data. Each of these types of analyses is informed by the biogeochemical measurements from the field site and functional and genetic fitness characterization of isolates (see other ENIGMA posters).

We study microbial interactions directly with synthetic communities and are developing novel microfluidics and isolation methods to directly observe interacting cells. Work with a consortium of *Pseudomonas* spp., the most common type of isolate from the FRC, has revealed that the type of carbon source available can influence the production of inhibitory secondary metabolites. Analysis of metabolomics data (NIMS and RP-LC-MS) derived from spent media provides candidate inhibitory compounds and the program MAGI is used to predict genes that produce them. Growth of transposon mutant libraries in spent media can provide information about which genes lead to susceptibility to these compounds. Preliminary data from experiments with combined carbon sources suggest that there is a “tipping point” in the proportion of a particular

carbon source that is needed for an isolate to produce an inhibitory effect on other microbes. Currently, we are developing methods to contain sediment particles or transposon mutant libraries in microfluidics droplets to observe microbial interactions and irreducible communities. We are also developing a method that studies flocs as we hypothesize that in ORNL FRC groundwater direct interactions of microbes are likely to occur within flocs of two or more cells. We are collecting flocs for identification of its members as well as for isolations and enrichments that will be used to reconstruct these floc communities in the laboratory. To provide a means to edit microbes in a group and dissect ecosystem function, we also have a discovery project to find bacteriophages at ORNL FRC.

Cultivation of enrichments provides a means to study sample communities under field relevant conditions and to study how ecological and environmental selection pressures in the same experiment affect community assembly. Building from experiments of isolates obtained with specific carbon sources, an experiment with 293 microcosms inoculated with ORNL FRC groundwater with varied type and number of sugars as a carbon source suggests that (1) a higher diversity of carbon sources leads to higher species diversity and (2) some carbon sources have a higher selection effect on the community structure than others. To study sulfate and nitrate reduction, two important metabolic activities at ORNL FRC, we are developing techniques for cultivation of enriched microbial communities under sulfate and nitrate reducing conditions. Enrichments are initiated under anoxic conditions using a variety of field relevant carbon sources. Additionally, we are using field geochemical information to guide media design.

Finally, we are using amplicon and whole genome metagenomics shotgun sequencing to study community structure across ORNL FRC and to study predicted microbial functional profiles in conjunction with field data. Comparison of the 16S rDNA sequences that were most abundant at the field site (either in groundwater or sediment) to the sequences of >1,000 isolates from the site informs our isolation efforts. We predict that that many of the uncultivated yet abundant microorganisms are aerobic chemolithoautotrophs or slow-growing facultative heterotrophs that utilize more recalcitrant carbon. We have also developed a computational pipeline to do a fast, focused analysis of nitrogen cycle related genes in metagenomics data and used it to analyzed six samples that represented sections of a 20 foot sediment core. This pipeline can be adapted to other functional roles. This pipeline yielded a view of how nitrification changes with depth and predicts a consortium of an ammonia-oxidizing thaumarchaeon and a nitrite-oxidizing bacterial species. In addition to this computational pipeline, we are working with KBase to integrate other metagenomics tools for fast analysis of other ENIGMA samples. Currently we are working on metagenomics of two more cores and accompanying groundwater from ORNL FRC that are coordinated with biogeochemical data and activity data.

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.

Elucidating Mechanisms of Rust Pathogenesis for Engineering Resistance in Poplar

Edward Eisenstein,[1,2,*] (eisenstein@umd.edu), Gary Coleman,[3] and Shunyuan Xiao[1.3]

[1]Institute for Bioscience and Biotechnology Research (IBBR), Rockville, MD and the [2]Fischell Department of Bioengineering, University of Maryland, College Park, MD; [3]Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD

Project Goals: The formidable challenge of engineering durable resistance in poplar against leaf rust is being addressed by investigation of the molecular basis for the virulence of *Melampsora larici-populina* towards *Populus* spp. Comparative genomics of small secreted proteins from other rust or fungal pathogens has provided an initial set of candidate secreted effector proteins (CSEPs) that is being screened for immune suppression in tobacco and poplar with the aim of revealing common mechanisms and host factors targeted by pathogens. Molecular interactions of poplar host factors and rust effectors probed by proteomics will provide targets for manipulation in transgenic poplar for use by the research community.

Abstract: An increasingly important goal in bioenergy research is to enhance the quality and availability of lignocellulosic feedstocks to match improvements in conversion processes to extend our diminishing supply of fossil fuels. *Populus* species present an attractive and sustainable target for enhancement via modern genetic engineering. Poplar trees grow rapidly and can produce significant biomass in short times when cultivated as short rotation woody crops. Additionally, because poplar has favorable cell wall characteristics, and also relatively high cellulose and low lignin levels, it can yield substantial energy. However, rust disease arising from *Melampsora* species is one of the most significant threats to poplar.

Our research aims to address the challenge of deciphering the interactions between *M. larici-populina* and poplar in an effort to engineer resistance in poplar through molecular intervention of pathogenesis. Our approach applies complementary molecular, genomic, biochemical and plant transformation approaches in order to understand the physiological disruption of the poplar defense and nutrient acquisition systems by rust fungus effectors, and to develop a set of transgenic poplar biotypes for broad use to more deeply probe the mechanisms of rust-poplar interactions and thereby develop disease resistance. The objectives are to use high-throughput screens in tobacco and poplar in order to identify key effectors of *Melampsora larici-populina*, as well as their poplar targets, and to construct an experimental system to

identify the most important genes involved in host defense against rust infection. This approach and new tools will enable us to test our overarching hypothesis that key rust effectors will either bypass the poplar immune system and/or suppress effector-triggered immunity through conserved mechanisms, and that an analysis of these interaction networks will provide for new approaches to develop rust-resistant poplar.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0017886

Soil- and Root-associated Microbiomes Across Twelve Switchgrass Cultivars

Tayler Chicoine (chicoin1@msu.edu),[1,2*] Sarah S. Roley[3,4], G. Philip Robertson[1,4,5], Sarah Evans[1,2]

[1]W.K. Kellogg Biological Station, Michigan State University (MSU), Hickory Corners, MI
[2]Department of Integrative Biology, MSU, East Lansing, MI
[3]School of the Environment, Washington State University, Richland, WA
[4]Great Lakes Bioenergy Research Center MSU, East Lansing, MI
[5]Department of Plant, Soil and Microbial Sciences, MSU, East Lansing, MI
*corresponding author

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

Switchgrass (*Panicum virgatum*), a C4 perennial grass, is known to associate with beneficial microbial communities that may enhance its potential as a low-input bioenergy crop. Genetically distinct cultivars adapted to southern and northern regions of the United States differ in their tolerance to marginal, low-input production systems, but the extent to which microbial communities influence this variation is unknown. We hypothesized that different microbial communities, and specifically more abundant free-living Nitrogen(N)-fixers bolster the ability of some cultivars to tolerate N-limited soils.[1,2] Further, we hypothesized that cultivars' specific root length (total root length/total dry weight), previously shown to correlate with switchgrass root-derived carbon, may contribute to differences in the cultivars' microbial communities.[3,4] Here, we measured N-fixing potential, root traits, and fungal and bacterial communities (16S, ITS, *nifH* abundance) in soils and roots of 12 switchgrass cultivars (including upland and lowland ecotypes) at the Great Lakes Bioenergy Research Center at Kellogg Biological Station in southwest Michigan. Preliminary findings suggest that bacterial and fungal community compositions do not differ among the cultivars or by ecotype. We will also present results on the relationships between N-fixation potential, N-fixer abundance, and root morphology among the cultivars. This study will inform our understanding of how plant-microbial interactions can support sustainable switchgrass bioenergy production.

References:

1. Bahulikar, R. A., Torres-Jerez, I., Worley, E., Craven, K., & Udvardi, M. K. (2014). Diversity of nitrogen-fixing bacteria associated with switchgrass in the native tallgrass prairie of Northern Oklahoma. Applied and Environmental Microbiology, 80(18), 5636–5643. https://doi.org/10.1128/AEM.02091-14

2. Rodrigues, R. R., Moon, J., Zhao, B., Williams, M. A., Biology, C., Tech, V., & Tech, V. (2016). Microbial Communities and Diazotrophic Activity Differ in the Root-Zone of Alamo and Dacotah Switchgrass Feedstocks. https://doi.org/10.1111/gcbb.12396

3. Graaff, M. A. de, Six, J., Jastrow, J. D., Schadt, C. W., & Wullschleger, S. D. (2013). Variation in root architecture among switchgrass cultivars impacts root decomposition rates. *Soil Biology and Biochemistry*, *58*(MARCH), 198–206. https://doi.org/10.1016/j.soilbio.2012.11.015

4. Adkins, J., Jastrow, J. D., Morris, G. P., Six, J., & de Graaff, M. A. (2016). Effects of switchgrass cultivars and intraspecific differences in root structure on soil carbon inputs and accumulation. *Geoderma*, *262*, 147–154. https://doi.org/10.1016/j.geoderma.2015.08.019

Support for this research is provided by the Great Lakes Bioenergy Research Center (US DOE Office of Science: DE-FCO2-07ER64494 and Office of Energy Efficiency and Renewable Energy: E-ACO5-76RL01830) and Department of Energy BER Genomics Program grant (DE-SC0014108).

Deeply rooted: impacts of depth and soil type on root biomass, carbon turnover and microbial communities under switchgrass cultivation

Jennifer Pett-Ridge[1], Karis McFarlane[1], Yuan Wang[2], Erin Nuccio[1], Colin Bates[3], Lauren Hale[3], Amrita Bhattacharyya[1,4], Yonatan Sher[5], **Malay Saha[2], Jizhong Zhou[3], Kelly Craven[2]** and **Mary Firestone[5]** (mkfstone@berkeley.edu)

[1]Lawrence Livermore National Laboratory, Livermore, California; [2]Noble Research Institute, Ardmore, Oklahoma; [3]University of Oklahoma, Norman, Oklahoma; [4]Lawrence Berkeley National Laboratory, Berkeley, California; and [5]University of California, Berkeley, California

Project Goals: Switchgrass (SG; *Panicum virgatum* L.) is a perennial C_4 grass native to the tallgrass prairies and a promising feedstock for bioenergy production in the U.S.A. Producing abundant biomass yield with minimal fertilizer or water, SG can survive on marginal soils, and thrives once established. Thus, SG is a model for low-input agriculture. Due to its deep rooting system, it may also serve as a means to augment soil carbon stocks. We are investigating the impact of switchgrass cultivation on different in nutrient-limited soil types in southern Oklahoma. Our goal is to determine how low-input switchgrass production in marginal soils may enhance ecosystem sustainability metrics such as C storage, nutrient availability, and soil food webs. 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 microbiomes, soil nutrient capital, plants and ecosystems, and ultimately guide efforts for translation into agronomic row crops.

Two-thirds of terrestrial C is stored as soil organic matter (SOM), yet soil C stocks (particularly in agricultural systems) are not at capacity, and could accommodate an increased sink of up to 50t ha^{-1}. But the fundamental mechanisms that regulate this vast pool (1500-1600 PgC in the top 1 m) remain elusive. What is clear, is that stabilized soil C is root carbon. Roots provide 30-40% of total soil organic C inputs, form a nexus for microbial transformations, and the primary source of C that becomes long-term stabilized. Deep soil C is stored for centuries or millennia, in contrast to C stored in vegetation, topsoils, or via other C sequestration strategies. We hypothesize that the soil C sink could be enhanced (e.g. in marginal and agricultural soils) via a greater emphasis on crops with deep rooting phenotypes such as switchgrass, *Panicum virgatum*.

Analyses were conducted in field plots planted with 'high' and 'low' performing switchgrass (Alamo cultivar) genotypes. Field sites are at the Nobel Research Institute's Red River Farm near the OK-TX border (sandy loam soil, low in NO_3-N and organic matter), and the 3rd Street Farm in Ardmore TX (silt loam soil, low in available P). Using a 'whole root ball' harvest approach, we measured root biomass and architecture for both high and low performing plants, as well as root elemental chemistry, sugar and protein content and overall digestibility. We compared this massively labor-intensive approach to a geospatial coring approach, and also a Tomographic Electrical Rhizosphere Imaging (TERI) approach designed to provide high throughput, dynamic, in-situ root phenotyping. Using accelerator mass spectrometry ^{14}C analyses of 2 m deep soil profile samples from our switchgrass fields, we characterized soil C stocks and turnover time prior to planting, and in several years' time will compare this to analogous in-plot measurements to assess C accrual. We also established a capability to make deep *in situ* $^{14}CO_2$ measurements to assess the radiocarbon age of actively respired carbon and have measured ^{14}C on soil density fractions to quantify soil C distribution and turnover of particulate, aggregate occluded, and mineral-associated C pools. Samples for analysis of microbial community

composition (16S rRNA and ITS) were also collected with depth and are currently being processed.

We found significant variation between high- and low-performing genotypes for root, shoot, and crown weight across locations with better shoot and crown growth in 3rd street but root growth in Red River. Root biomass was higher in high performing genotypes and in sandy soils, and better measured by the direct approach than inferences based on coring. Total soil C and N stock decline precipitously with depth, and soil carbon is primarily found in the mineral associated fraction beyond the top 20 cm. Soil carbon turnover time ranged from ~3000 yrs in silt loam soils to >4000 yrs in sandy 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.

Exometabolomics of Switchgrass-Microbe Interactions in natural and fabricated ecosystems

Kateryna Zhalnina[1]* (kzhalnina@lbl.gov), Eden Tepper,[1] Yuan Wang,[2] Erin Nuccio,[3] Nameer Baker,[4] Katerina Estera,[4] Na Ding,[2] Lauren Hale,[5] Javier Ceja Navarro,[1] Malay Saha,[2] Benjamin P. Bowen,[1] **Jizhong Zhou,[5] Kelly Craven**,[2] **Jennifer Pett-Ridge,[3] Mary Firestone**,[1,4] **Eoin Brodie**[1,4] and **Trent Northen**[1]

[1]Lawrence Berkeley National Laboratory, Berkeley, California; [2]Noble Research Institute, Ardmore, Oklahoma; [3]Lawrence Livermore National Laboratory, Livermore, California; [4]University of California, Berkeley; [5]The University of Oklahoma, Norman

Project Goals: Our project studies molecular mechanisms driving beneficial plant–microbial interactions in superior switchgrass genotypes growing under a range of resource limitations. The genomics and chemistry of plant-microbe interactions have been 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.

Plants exude a large amount of photosynthesis-derived carbon (C) into rhizosphere soil. These exuded, small, organic molecules (metabolites) shape rhizosphere community composition and potentially attract beneficial microbes that improve plant nutrition, reproduction, drought tolerance, and other stress responses. Additionally, these interactions partially define the future fate of root C, specifically whether it is respired to the atmosphere or stabilized in soil. In this study, we combined mass spectrometry-based metabolomics with genomic analyses of rhizosphere isolates to uncover key mechanisms of switchgrass-microbe interactions in soil. Using this integrated approach, we are studying the release of exudates by switchgrass under different nutrient limitations and metabolite exchange between switchgrass and rhizosphere microorganisms. These data will provide a detailed understanding of the substrate preferences of the rhizosphere microbiome as it is linked to dynamic root exudation and ultimately how these interactions impact C persistence in soil.

Metabolite profiles of the switchgrass rhizosphere in field soil. Switchgrass roots deposit large quantities of metabolites into soil and these metabolites likely shape the rhizosphere community. We have developed protocols for metabolite extraction and we have identified soil metabolite profiles in two field experiments in Oklahoma in which switchgrass was cultivated in marginal soils. Metabolite profiles from the two soils differed in composition. Switchgrass rhizosphere soils from the Red River field had proportionally higher amounts of nitrogen containing compounds, such as amino acids and nucleotides; while soils from the Third Street field had a higher content of carboxylic organic acids. Currently we are analyzing metabolite profiles at the main stages of switchgrass establishment and at the different root depths. To identify how metabolites released by roots shape the rhizosphere communities, we will link metabolite profiles to the taxa identified in the soils from the field.

Synthetic rhizosphere community of switchgrass. To dissect how different microbial taxa interact with switchgrass and potentially improve plant biomass yield, we isolated and identified 300 heterotrophic bacteria from the switchgrass rhizosphere. We determined how abundance of each isolate changes during plant development by mapping 16S rRNA of isolates to OTU-representative sequences in the field where switchgrass was cultivated. We selected 40 representative isolates for further genome sequencing and for designing synthetic rhizosphere communities to be used to test potential mechanisms defining switchgrass-microbe interactions.

Switchgrass exudation during development and under nutrient stress. We hypothesize that switchgrass exudes specific molecules to select microbial taxa to mitigate stress and improve its nutrition. We identified signature molecules released by switchgrass under nitrogen and phosphorus limitation. We are currently identifying microbial taxa that metabolize these molecules.

Fabricated laboratory switchgrass ecosystems. To dissect plant-soil-microbe interactions in the lab we are using 3D printing technologies to create 'controlled laboratory ecosystems' (EcoFABs) developed in the Northen Lab. Under simplified, controlled, sterile conditions we combine switchgrass with isolated bacteria to collect and identify metabolite profiles under different environmental stresses and to visualize bacteria and root architecture.

This research is based upon work supported by the U.S. Department of Energy Office of Science award DE-SC0014079 to UC Berkeley, Noble Research Institute, University of Oklahoma, Lawrence Berkeley National Lab, and Lawrence Livermore National Lab.

Identifying microbial traits driving microbial succession and C transformation in the rhizosphere of *Avena sp.*

Heejung Cho[1,2*] (beeologai@berkeley.edu), Kateryna Zhalnina,[2] Ulas Karaoz,[2] Shengjing Shi,[1,3] Benjamin P. Bowen,[2] Richard A. White III,[4] **Mary S. Lipton,[4] Trent R. Northen,[2] Mary K. Firestone,[1,2]** and **Eoin L. Brodie[1,2]**

[1]University of California, Berkeley; [2]Lawrence Berkeley National Laboratory, Berkeley, California; [1,3]Lincoln Science Centre, AgResearch Ltd, Christchurch, New Zealand; [4]Pacific Northwest National Laboratory, Richland, Washington

Our project is designed to advance our understanding of the complex interactions controlling C flow in the rhizosphere by addressing two fundamental topics: 1) How multi-trophic interactions control soil C dynamics, and 2) How changing precipitation regimes alter these interactions, and thus impact flow and fate of soil C.

Plant-soil-microbial interactions strongly impact the dynamics of soil organic carbon (SOC). Plants exude a variety of compounds, supplying rhizosphere microorganisms with readily assimilable substrates and driving microbial succession in response to developing plant roots.

In this study, we examined succession of rhizosphere bacteria throughout *Avena sp*. developmental stages using metagenomic approaches, isolation and metabolomics. To determine the changes in bacterial abundance during root growth, metagenome and 16S rRNA gene reads from rhizosphere and bulk soils during *Avena* developmental stages were mapped to the genomes of thirty-nine bacterial isolates that are phylogenetically representative of Hopland soils and numerically abundant, and to fifty-five moderately complete metagenome-derived genome bins. We classified these bacteria into three different groups based on response in abundance to root growth (positive, negative and no response). We used comparative analyses of these genomes to identify bacterial traits that may contribute to higher abundance in rhizosphere.

Each genome was analyzed for key traits related to C transformation and survival in rhizosphere, such as minimum generation times, polymer degrading enzymes and transporters for different substrates (amino-, fatty-, organic-acids, sugars, nucleotides and plant hormones). We determined that genomes present in bacteria responding positively to root development were slower growers compared to the bacteria with negative responses. Positive responders demonstrated higher numbers of organic and amino acid transporters present in their genomes and fewer genes involved in decomposition of polymeric C. Additionally, multiple functional genes involved in sugar and fatty acid transport, bacterial secretion systems, sugar, amino acid, and organic acid metabolism, motility, pilus formation, and ecdyson metabolism were more abundant in bacteria with positive response to root growth.

We compared closely related *Bradyrhizobial* isolates and genome bins that presented positive, negative and no responses to root development. Genome comparisons among *Bradyrhizobium* isolates and genome bins showed that a *Bradyrhizobium* isolate with higher abundance in rhizosphere had sugar, plant cell-wall driven polysaccharides, amino acids, organic acids and lignin metabolism genes, bacterial secretion system genes, antibiotics exporter genes, and papain-like cysteine protease genes that are absent in *Bradyrhizobial* isolates and genome bins with negative or no response to root development. The presence of these functional genes may play a role in niche optimization and improved fitness, thus shaping rhizosphere bacterial succession.

To investigate the relationship between *Avena* exudates and rhizosphere bacteria and to link identified rhizosphere traits to root chemistry we first analyzed exudate composition of hydroponically-grown plants using LC-MS/MS based metabolomics. We then designed a medium to simulate plant exudates and using this medium we examined the substrate preferences of rhizosphere isolates.

The major fraction of plant exudates was found to be composed of amino- and carboxylic acids, sugars, nucleosides, released at vegetative growth stages and quaternary amines and plant hormones exuded at latter stages of *Avena* growth. Amino acids, sugars and nucleosides were consumed by all analyzed isolates. However, isolates that were preferentially stimulated by plant growth, positive responders, revealed substrate utilization preferences towards aromatic organic acids, such as salicylic, cinnamic, nicotinic, indole-acetic while those not responding to growing roots did not utilize these compounds.

To confirm the genome-predicted trait of rhizosphere-negative isolates more efficiently utilizing polymeric C compared to the positive responders, we evaluated polymeric substrate preferences and enzymatic activities of different isolates using nanostructure initiator mass spectrometry-based enzyme activity assay (Nimzyme) and secretome analysis. Rhizosphere-negative isolates demonstrated higher ability to degrade polymers compared to the rhizosphere-positive bacteria, which revealed monomeric molecules as the preferred source of C.

Together these approaches are being used to predict and test how key traits of soil bacteria interact with root metabolism and soil organic matter to impact C flow 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 Numbers DE-SC0010570 and DE-SC0016247 to UC Berkeley. Part of this work was performed at Lawrence Berkeley National Laboratory under the auspices of the U.S. Department of Energy under Contract DE-AC02-05CH11231 and a portion at the Environmental Molecular Sciences Laboratory (EMSL) sponsored by OBER and located at PNNL under Contract DE-AC05-76RLO 1830.

Plant-microbe and microbe-microbe interactions mediate switchgrass sustainability: following rhizosphere microbial communities during switchgrass establishment

Yuan Wang[1], Erin Nuccio[2], Travis Simmons[1], Katerina Estera-Molina[3], Christina Fossum[3], Jialiang Kuang[4], Colin Bates[4], Lauren Hale[4], Na Ding[1], Josh Barbour[1], Nameer Baker[4], Javier Ceja Navarro[5], Abelardo Arellano[5], Kateryna Zhalnina[5], Eoin Brodie[5], **Trent Northen**[5], Wolf Scheible[1], Michael Udvardi[1], **Jennifer Pett-Ridge**,[2] Malay Saha[1], **Jizhong Zhou**[4], **Mary Firestone**[3] and **Kelly Craven**[1]* (kdcraven@noble.org)

[1]Noble Research Institute, Ardmore, Oklahoma; [2]Lawrence Livermore National Laboratory, Livermore, California; [3]University of California, Berkeley, California; [4]University of Oklahoma, Norman, Oklahoma; [5]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 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. Thus, SG is a model for low-input agriculture as well. A major barrier to large-scale cultivation of SG under low-input management is seedling establishment, somewhat of an Achilles heel for all or most perennial plants. Further, little is known regarding how SG responds to abiotic and biotic stressors, nor what beneficial relationships the plant may forge to deal with these types of stresses. We hypothesize that successful establishment and sustainable cultivation of SG in marginal soils is in part enabled by beneficial plant-microbial interactions. We are investigating the succession of rhizosphere microbial communities associated with high- and low-performing SG plants grown in nutrient-limited soils at two Oklahoma fields. 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 microbiomes, plants and ecosystems, and ultimately guide efforts for translation into agronomic row crops.

In the soils surrounding roots (rhizosphere), biotic, chemical and physical drivers enrich for specific bacterial and fungal communities. These organisms can play multiple roles, and some may benefit plant productivity through assisting in nutrient acquisition, water uptake and/or pathogen suppression. We are investigating the composition, succession and function of rhizosphere microbial communities during SG cultivation in an effort to better understand the plant-microbiome interactions that enable plant survival and adaptation under stressed conditions, such as drought and nutrient deficiency.

To study plant-soil microbiome characteristics of SG growing in 'marginal' nutrient or water-limited soils, we selected two research farms, both remnants of the Dust Bowl Era in Oklahoma. Red River Farm lies near the border of Oklahoma and Texas, and has sandy loam soil that is low in NO_3-N and organic matter. The other one, Third Street Farm, has a silt loam soil and relatively low phosphorus availability. Five hundred genetically distinct Alamo seedlings were planted into each field in May-June of 2016. Other than hand weeding during the summer, no management, water, or nutrients were supplied to the fields. Thirty plants were randomly selected at the onset for monitoring of rhizosphere community succession, plant performance and soil physio-chemical characteristics including: gravimetric moisture, pH, and NO_3/NH_4. Rhizosphere and bulk soil of these plants were sampled over the first growing season (Year 1) at 5 time points (T1-T5): early and late vegetative growth, reproductive growth, maximal growth, and senescence. Although plant biomass exhibited large variation within plots due to plant genetic

diversity, we believe the selected thirty plants are representative of the whole plots in terms of biomass production. We are investigating microbial community profiles by amplicon sequencing of marker genes specific to bacteria, fungi and other soil eukaryotes. To date, we have finished 16S amplicon sequencing for T2, T3 and T4 samples and here present some of the preliminary data. Proteobacteria, Actinobacteria, Acidobacteria and Firmicutes are dominant in all the soil samples, accounting for ~75-80% of the microbial communities. Third Street (clay soil) has more Verrucomicrobia and Firmicutes present when compared to Red River (sandy soil). Proteobacteria is significantly more abundant in the rhizosphere soil (directly around the root surface) relative to bulk soil (no close proximity to SG plants) in both sites. Proteobacteria abundance appeared to decrease when plants transitioned from active growth to senescence. Overall, microbial communities of Red River soil exhibits higher alpha-diversity than those found at Third Street. As expected, rhizosphere microbial communities are less diversified than the corresponding bulk soil at both sites, indicating a selective effect of root exudation. Significant differences in soil microbial communities were observed between both sites as well as between different habitats (rhizosphere vs. bulk soil). However, site effect appears to be greater than the rhizosphere effect. Intriguingly, Third Street soil differences between rhizosphere and bulk soil are less pronounced than those found at Red River. Community shifts in rhizosphere soils were observed between most time points at both sites, suggesting there is indeed a succession pattern of rhizosphere microbial communities over growth season.

In order to dissect the changes of rhizosphere communities *between* growth seasons, we performed our rhizosphere community survey of these thirty plants in Year 2 (year of 2017). To enable this additional research, an “in-growth core strategy” was adopted, and used to collect roots and the rhizosphere soil produced in Year 2. To maintain consistency with Year 1, soil samples were collected at three time points: vegetative growth, reproductive growth, and senescence. In addition, 40 plant genotypes from the plots were selected based on plant performance, to establish replicated field experiments. This collection includes 12 high- and eight low-performing genotypes originating from each of Red River and Third Street plots. The new plots of clonal plants have been planted and established at both sites adjacent to the Year 1 plots in May 2017. These clonal plots will allow us to investigate the relationship between host plant genetic diversity and their associated rhizosphere communities. Rhizosphere and bulk soil of plant clones of 18 genotypes (12 high-performing and 6 low-performing genotypes) were sampled in Sep. 2017. DNA extraction and marker gene sequencing of these soil samples is ongoing.

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.

Stable isotope-enabled metagenomics reveal phages active in rhizosphere soil

Evan Starr[1]* (starr.evan@berkeley.edu), Alexa Nicolas[1], Shengjing Shi[2], Sneha Couvillion[3], Alexander Probst[4], Donald Herman[1,6], Steve Blazewicz[5], **Mary Lipton[3]**, Jillian Banfield[1,6], and **Mary Firestone[1,6]**

[1]University of California, Berkeley, California; [2] Lincoln Science Centre, AgResearch Ltd, Christchurch, New Zealand; [3]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA; [4] Biofilm Center, University of Duisburg-Essen, Essen 45141, Germany; [5] Lawrence Livermore National Laboratory, Livermore, California; and [6]Lawrence Berkeley National Laboratory, Berkeley, California

Project goals: The flow and fate of carbon from roots into soil is controlled by a myriad of interactions. Our project explores the impact of phages on the rhizosphere community and ultimately on the fate of carbon in soil. We are tracing carbon movement, through stable isotope labelling, from plant fixation to exudation in the root zone, uptake into the microbial community, and finally through replication and lysis into active phages. The combined approach of stable isotope probing with genome resolved metagenomics gives us insight into the flow of carbon, novel phage diversity, and phage activity. Additionally, we are employing metaproteomics, Nanoscale Secondary Ion Mass Spectrometry (NanoSIMS), and fluorescence imaging of viral like particles to quantify and follow phage activity and dynamics. This combination of approaches will allow us to better understand the impact of phages on carbon flow from roots into soil, microbial community dynamics and the soil food web, all with implications for soil health and agricultural productivity.

The zone of influence around plant roots, the rhizosphere, is of great importance to plant health and carbon cycling. There have been numerous experiments exploring plant-bacteria interactions, but phages have been largely overlooked in rhizosphere studies. We hypothesize that phages have an effect on the rhizosphere community and ultimately on the fate of soil carbon. The impact of phages in soil may be analogous to the "phage shunt" or "biological pump" from marine systems in which phage lysis either keeps carbon in the dissolved organic carbon pool which is quickly metabolized by microbes and respired back to the atmosphere or phage lysis causes organic carbon to be sorbed onto mineral surfaces, creating persistent soil organic carbon. To begin to address our larger hypotheses we need a fundamental understanding of our system. Phages have been understudied in soil, not because of their perceived lack of impact, but because of the difficulty of phage research and the abiotic and biotic diversity of soil more generally. Therefore, we needed to determine how best to study phage activity, diversity, and dynamics in soil. We combined stable isotope probing (SIP) and genome resolved metagenomics (DNA sequencing) to find and analyze recently active phage populations in the rhizosphere. In this study we grew *Avena fatua*, common wild oat, in $^{13}CO_2$ and collected rhizosphere soil, and bulk soil (soil not associated with roots) at weeks 0, 6, and 9. The DNA extracted from these samples underwent density-gradient centrifugation, yielding DNA samples with a range of ^{13}C label: unlabeled, partially labeled, and heavily labeled DNA. DNA from the separated fractions were then sequenced, and the sequence data were assembled, binned for host genomes, and phage contigs were identified.

We are thus able to identify active phages in the rhizosphere. The populations of phages that incorporated the label into their genomes were different from the non-labelled phages, indicating that the development of rhizosphere-competent bacterial consortia enabled the production of a new (different) phage population. We were also able to link several phages with their hosts and show that there were differences in phage-host abundances between samples, indicating a dynamic system. Our discovery of active phages is supported by metaproteomic analyses which identified expressed phage proteins in the rhizosphere soil. In preparation for future, more targeted studies, we have improved our quantitative methods including enumeration of phage-like particles using fluorescent microscopy and assessment of phage C isotope content using NanoSIMS.

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 to MF at UC Berkeley and DE-SC10010566 to JB at UC Berkeley. AJP was supported by the German Science Foundation under DFG1603-1/1. Work at LLNL was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344 and work at the Environmental Molecular Sciences Laboratory (EMSL located at PNNL) under Contract DE-AC05-76RLO 1830.

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, Oklahoma, [2] University of California, Berkeley, California, [3] Lawrence Berkeley National Laboratory, Berkeley, California, [4] Lawrence Livermore National Laboratory, Livermore, California

Project Goals: Our project works towards a fundamental understanding of the key molecular mechanisms driving beneficial plant-microbial interactions in superior switchgrass genotypes adapted to a range of resource limitations. Plant-microbe interactions are examined during establishment to gain insight into how symbiotic and associative microbes improve plant performance and carbon stabilization in marginal soils. We will combine focused (single plant-microbe pairing) and 'community' systems biology approaches to examine the complex interplay among plants, microbes, and their physio-chemical environment.

In the rhizosphere, root exudation is a key process for C transfer into the soil, influencing the role of soil microbial communities in the decomposition of organic matter and in nutrient cycling. Root exudates have been shown to 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, potentially 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 transcriptomes and exudates shape rhizosphere microbial communities, metagenomes, and metatranscriptomes and how these interactions are affected by abiotic stresses. The first group of experiments focused on the Alamo SG transcriptome, metabolome and physiology under a range of P_i supply conditions to explore the gene-to-metabolite networks responsible for coping with P starvation. The second experimental approach used Alamo SG clones growing 1-m soil profiles recreated in the greenhouse and applying $^{13}CO2$ stable-isotope labeling to trace SG photosynthate into fresh root exudates and the metagenomes and metatranscriptomes of the microbial communities that consume root exudates and debris.

Sand-based: Alamo SG seedlings were grown in sand culture over a 4-week period with nutrient media containing KH_2PO_4 concentrations ranging from 20µM to 600 µM (control). Plants growing at 200 µM KH_2PO_4, accumulated only 64.7% of the biomass present in plants grown at 600 µM, whereas plants grown in the presence of 20 µM only had 3.8% of the control biomass. Severe P limitation (20 µM) did not inhibit primary root growth, as has been frequently reported (e.g. Arabidopsis thaliana) and is considered a typical root system architectural change resulting from inhibition of primary root elongation by P limitation. At the transcriptional level, SG showed expected (previously known) and novel responses to P limitation. The number of gene transcripts and the strength of their response increased with the severity of P limitation both in shoot and root tissues. RNA-Seq data were analyzed with MapMan software to identify coordinated, system-wide changes in metabolism. During P limitation, a large number of gene transcripts related to lipid degradation, glycolipid biosynthesis as well as tryptophan synthesis

were induced. For example, NPC4, encoding nonspecific phospholipase C, was markedly induced upon P limitation and is known to play an important role in the supply of phosphate from membrane phospholipids during P limitation. SG under P stress also showed large changes in the expression of genes involved in secondary metabolism and photosynthesis. It is surprising that a suit of genes related to light reaction were sensitively down-regulated in shoots, but also in roots experiencing P limitation. Analyses of metabolites confirmed that P limitation led to a shift towards the accumulation of sugars and organic acids in roots relative to shoots. Shoots and roots show distinct adaptation patterns at the molecular and metabolic levels towards P limitation, suggesting that distinct P starvation response strategies are used for different plant organs in response to a shortage of P. Lipid remodeling is known to be a dramatic metabolic response to P starvation. The abundance of phospholipid species (PC, PE, PS) was significantly decreased in shoot under 20μM treatment; however, this response was not observed during more moderate P limitation (60μM and 200 μM). Membrane glycolipids, such as MGDG and DGDG, accumulated in roots during P limitation as compared to control.

Soil-based: Alamo SG clones were grown in mesocosms containing three horizons of Oklahoma pasture soil packed into 1m columns, to which one of five treatments (+N, +P, +N/+P, $-H_2O$, and control) was applied. These treatments had a significant ($p<0.05$) effect on plant height, shoot biomass, and relative root biomass, with the greatest height and biomass observed in the +N/+P treatment and the least in the $-H_2O$ treatment, as expected. The +N/+P treatment also had a significant ($p<0.05$) positive effect on exopolysaccharide (EPS) abundance in top horizon bulk soil, though further analysis is required to determine if this EPS is plant or microbial in origin. After nine weeks, a subset of SG plants of each treatment were labeled with 99 atom-percent $^{13}CO2$ for 6 days, and after 14 and 18 weeks, another two subsets were labeled for 12 days each. Labeled mesocosms were cored or destructively harvested to recover roots, rhizosphere soil, and bulk soil at different depths within the rooting profile. Bulk soil chemistry, density fractions, and microbial metagenomes are being used to determine how growing SG alters bulk soil characteristics with depth. Rhizosphere microbe metagenomes and metatranscriptomes, root exometabolites, and root transcriptomes extracted from these samples will be used to determine how SG-microbe interactions in the rhizosphere are delineated and how these networks respond to abiotic stress.

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.

Molecular interactions of the plant-soil-microbe continuum of bioenergy ecosystems

Kirsten Hofmockel[1]* (kirsten.hofmockel@pnnl.gov), Allison Thompson[1], Sheryl Bell[1]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

Project Goals: The overall goal of this project is to identify the factors regulating plant-microbe-soil interactions to determine the extent to which biofuel feedstock production can confer environmental benefits. We aim to test if plant-microbe interactions are limited to influencing the rate of C accrual, while mineralogy regulates the sink capacity of biofuel cropping systems. To accomplish this goal, we are (1) identifying how microbial community composition and function relate to the production of microbial necromass that contribute to soil C accumulation, (2) characterizing microbial dynamics that contribute to necromass accumulation in situ in response to crop selection and edaphic factors and (3) generating long-term, cross-site data that can be used to model C cycling in bioenergy cropping systems under different soil conditions.

Abstract text. Crop selection and soil texture influence the physicochemical attributes of the soil, which structures microbial communities and influences 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. By integrating lab to field studies we aim to identify the molecules, organisms and metabolic pathways that control the formation of molecules that contribute to long term organic matter stabilization in bioenergy soils.

We investigated the relative influence of crop and site selection on microbial community composition, enzyme activity and soil chemistry at the Great Lakes Bioenergy Research experiment. Sites in WI and MI, USA have been in corn and switchgrass cropping systems for a decade, resulting in pronounced differences in soil structure. By comparing soil aggregate ecology across sites and cropping systems we are able to test the relative importance of plant, microbe, and mineral influences on soil organic matter dynamics. Soil bacterial and fungal community structure (Bray-Curtis) differ with crop selection, and are even more strongly influenced by site selection. Crop and site effects in microbial community structure were consistently apparent in all aggregate fractions and treatment effects were more pronounced in large compared to small aggregate fractions. Similarly crop and site selection influenced potential enzyme activity with strongest site effects in enzymes that liberate carbon molecules (betaglucosidase, betaxylosidase) and more pronounced crop effects in enzymes that release compounds containing carbon and nitrogen (aminopeptidase, nagase). FT-ICR MS revealed differences in the molecular composition of water-soluble fraction of soil organic matter. Similar to microbial activity, we found that bioavailable substrate pools were more strongly influenced

by plant inputs compared to soil origin. For example, we observed more molecules classified as amino sugars, carbohydrates, proteins, and lipids in switchgrass plots, while condensed hydrocarbons and lignins were more prevalent in the corn cropping systems. These data are consistent with cropping system effects on potential enzyme activity, where betaglucosidase, betazylosidase, aminopeptidase and nagase activity was greater in soils under switchgrass compared to corn. These results suggest that although soil origin influences the bacterial and fungal community structure, plants have a stronger influence on microbial activity which is consistent with the soluble soil chemistry. While cycling of organic compounds is reflective of bioenergy cropping systems, the differences in microbial community structure may have long lasting effects on the stabilized carbon remaining in the soil. Ongoing research aims to characterize the persistent organic compounds that are hypothesized to reflect microbial necromass more than recent plant inputs.

This research was supported by an Early Career Research Program award to K 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.

Climate adaptation and sustainability in switchgrass: exploring plant-microbe-soil interactions across continental scale environmental gradients

Thomas E. Juenger,[1,*] (tjuenger@austin.utexas.edu), Katherine Behrman,[1] Jeremy Schmutz,[2] Tanja Wolke,[3] Felix Fritschi,[4] Laura Bartley,[5] Julie Jastrow,[6] Roser Matamala,[6] Sean Watson,[7] Denise Costich,[8] Alina Zare,[9]

[1]University of Texas, Austin;[2] HudsonAlpha Institute for Biotechnology, Huntsville, Alabama; [3]DOE Joint Genome Institute, Walnut Creek, California; [4]University of Missouri, Columbia;[5] University of Oklahoma, Norman; [6]Argonne National Lab, Argonne, Illinois; [7]Ladybird Johnson Wildflower Center, Austin, Texas; [8]International Maize and Wheat Improvement Center; [9]University of Florida

https://sites.cns.utexas.edu/juenger_lab/switchgrass

Project Goals: Our collaborative project is focused on understanding switchgrass genetic diversity and adaptation across continental scale environmental gradients. Our goal is to improve the sustainability of switchgrass as a biofuel by gaining insight into the interaction of switchgrass diversity with its associated microbiome and environmental conditions. Our approach involves 1) the collection and characterization of new switchgrass germplasm from across the species range, 2) the development of a genetic association mapping panel and extensive common gardens to evaluate switchgrass performance, 3) a detailed characterization of the switchgrass microbiome, 4) studies of the impact of switchgrass stands on ecosystem processes, and 5) extensive multiscale modeling to define conditions of sustainability and identify key tradeoffs between genetic diversity, productivity, and ecosystem services. Ultimately, these studies will identify critical plant-microbe-soil traits that may be manipulated, through breeding or agronomic management, to improve the sustainability of biofuel feedstocks.

Less carbon-intensive energy sources are needed to reduce greenhouse gas emissions and their predicted role in climate change. There is growing interest in the potential of biofuels for meeting this need. A critical question is whether large-scale biofuel production can be sustainable over the time scales needed to mitigate our carbon debt from fossil fuel consumption. The carbon balance and ultimately the sustainability of biofuel feedstock production is the result of complex climate-coupled interactions between carbon fixation, sequestration, and release through combustion. The long-term productivity of biofuels depends on the genetic and environmental factors limiting plant growth. These factors are often related to soil resources which involve complex interactions at the plant-microbe-soil interface impacting their availability and cycling.

Our collaborative project addresses sustainable switchgrass (*Panicum virgatum*) production through a detailed characterization of plant growth and performance in both individual spaced and stand plantings. The project represents an unprecedented field-based experimental system for a bioenergy grass. We bring together diverse skill sets from plant and microbial genetics and genomics, physiology and ecosystem modeling. An underlying theme of the research is the use of locally adapted plant material to explore plant function, to understand the mechanistic basis of

Environmental interactions, and to discover the plant genes important for adaptation and sustainability in the face of climate change. To this end, we have been collecting new genetic diversity in switchgrass from natural populations across the species range in North America. These new genotypes are being characterized through genome resequencing and clonally propagated for inclusion in experimental gardens at 14 field locations. This material will provide a detailed population genomic characterization of switchgrass along with resources for association mapping and genomic selection for future breeding programs.

Our plant-microbiome project will fully characterize the microbial communities associated with switchgrass at our planting locations using genomic tools. Specifically, we will sample field plantings of switchgrass for leaf and root microbial communities with 16S iTAG and metegenomic sequencing in collaboration with the DOE Joint Genome Center. Analyses will center on quantifying the relative importance of switchgrass host genotype, the planting environment, or their interaction on microbial community composition. Ultimately, these data will be linked with plant trait information to evaluate drivers of plant-microbiome interactions and their impacts on ecosystem processes.

Our ecosystem processes research focuses on carbon cycle responses at the ecosystem level using stand plantings of switchgrass diversity. We will couple plant growth and physiology measurements with measurements of key carbon pools and fluxes that integrate carbon dynamics across different temporal and spatial scales. Finally, our modeling will define conditions of a sustainable biofuel system and identify key tradeoffs between genetic diversity, productivity, and ecosystem services. Here, we highlight the background resources for our project, our progress establishing our new experiments, and plans for the coming field season.

This research is supported by funding through the Office of Biological & Environmental Research within the Department of Energy Office of Science and through the DOE Joint Genome Institute Community Science Program.

Towards durable resistance to *Septoria* stem canker and leaf spot: a molecular understanding of resistance.

Jared M. LeBoldus[1] (your email: jared.leboldus@science.oregonstate.edu), Wellington Muchero[2], Jin-Gui Chen[2], and Steven Strauss[3].

[1] Department of Botany and Plant Pathology & The Forest Engineering, Resources and Management Department, Oregon State University, Corvallis, OR, USA.

[2] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA.

[3] Department of Ecosystems and Society, Oregon State University, Corvallis, OR, USA.

Fuels developed from lignocellulosic biomass offer a potential renewable and clean alternative to conventional fossil-based energy sources. *Populus* is one of DOE's "flagship" plant species that is of special interest as a biofuel feedstock. *Septoria* canker is the major limiting factor in the use of *Populus* as a biomass feedstock in the central and eastern United States. An effective disease resistance-breeding program has not been developed due to an absence of information on the genetic basis of resistance. To identify resistance alleles, we combine the re-sequenced *P. trichocarpa* genome-wide association population with our robust disease resistance phenotyping platform. This enables us to: (1) identify and analyze alleles conferring resistance to *Septoria* stem canker and/or leaf-spot disease; (2) use a combination of stable and transient expression systems in *Populus* and *Nicotiana* to validate the function of the encoded proteins of the identified genes; and (3) field test resistant genotypes in order to validate their performance under changing environmental conditions.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018196.

Identification of Adaptive Fungal Pathogen Resistance Loci in Switchgrass

David B. Lowry[1,*] (dlowry@msu.edu), Shawn C. Kenaley[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 will: 1) Characterize the pathogens associated with disease in switchgrass and quantify their geographic distributions. 2) Discover genetic loci for effective switchgrass disease resistance across different geographic locations. 3) Validate 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.

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.

Our research will utilize new and powerful genetic mapping populations to identify genomic regions responsible for divergence in disease resistance between northern upland and southern lowland switchgrass ecotypes. The mapping populations include 1) a northern upland X southern lowland four-way, pseudo-testcross F2 tetraploid genetic mapping population for Quantitative Trait Locus (QTL) mapping and 2) a new large panel of ~400 resequenced population accessions for genome-wide association studies (GWAS). These mapping populations have been planted at an unprecedented geographical scale, spanning ten common garden field sites distributed over 17 degrees of latitude in the central United States. Therefore, these populations are ideal for identification of regionally effective disease resistance loci (at one sites) as well as globally effective loci (across multiple sites).

To understand the geographic context for which specific disease resistance loci are effective, we are currently quantifying disease severity and the geographic distribution of pathogens across field sites. Overall, our research will achieve an unparalleled understanding of the genetic basis of disease resistance in switchgrass to multiple pathogens with yield-reducing potential. We will focus on three major goals:

Goal 1. Characterize the pathogens associated with disease in switchgrass and quantify their geographic distributions. We are currently in the process of identifying the specific

fungal pathogens associated with disease at each of the ten field sites, using a combination of microscopy and molecular methods. Analyses of pathogen distribution and environmental factors will be used to develop predictive models for pathogen presence across space.

Goal 2. Discover genetic loci for effective switchgrass disease resistance across different geographic locations. Disease severity of fungal pathogens is being quantified across mapping populations at all ten field sites. We will conduct QTL mapping and GWAS to identify regionally and globally effective disease resistance loci.

Goal 3. Validate QTLs and genes for pathogen resistance through controlled experiments. We will conduct controlled greenhouse experiments to confirm that QTLs identified in the field are effective against fungal pathogens present at particular field sites.

The proposed research builds upon many years of genomic resource, germplasm, and methods development for switchgrass by the collaborators. This strong foundation, combined with the collaborative synergy of the research team, positions this research to make rapid progress in identifying loci involved in disease resistance in this key bioenergy crop.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0017883.

Spatially Resolved Rhizosphere Function for Elucidating Key Controls on Below-ground Nutrient Interactions

James Moran[1*] (james.moran@pnnl.gov), Elizabeth Denis[1], Peter Ilhardt[1], Joshua Rosnow[1], Vivian Lin[1], Jamie Nunez[1], Ryan Renslow[1], Colin Brislawn[1], Abigail Tucker[1], Nicholas Huggett[1], Sarah Jenson[1], Eirik Krogstad[1]

[1]Pacific Northwest National Laboratory, Richland, Washington

http://www.pnl.gov/biology/programs/fsfa/people/moran.stm

Project Goals: This project seeks to elucidate key microbiological and geochemical controls on nutrient exchange through the rhizosphere and the role that spatial organization within the root-rhizosphere-soil continuum plays in nutrient transfer. Spatially-resolved understanding of nutrient exchange across the rhizosphere will identify key variables amenable to manipulation as part of an effective rhizosphere management program targeting enhanced plant productivity. Our aims are directed towards 1) spatially tracking plant-derived organic carbon contributions to soil, 2) identifying key microbial membership distribution within zones of high nutrient transfer, and 3) evaluating whether directed geochemical and/or microbiological modifications can be used to stimulate nutrient exchange to foster improved plant biomass productivity.

The central hypothesis we are testing is that spatially focused regions funnel a disproportionate amount of nutrients to a plant root. Further, we hypothesize that the location of these resulting nutrient exchange hotspots are not stochastically distributed throughout the rhizosphere but, rather, that they are controlled by microenvironmental conditions resulting from a combination of local microbiological communities in conjunction with host soil geochemistry.

To begin testing these hypotheses, we constructed microcosms using switchgrass (variety Cave-in-Rock) and soil harvested from the Kellogg Biological Station (Hickory Corners, MI). We used ^{13}C-labeled carbon dioxide (CO_2) to track photosynthetic fixation and subsequent migration of labeled organic compounds through the

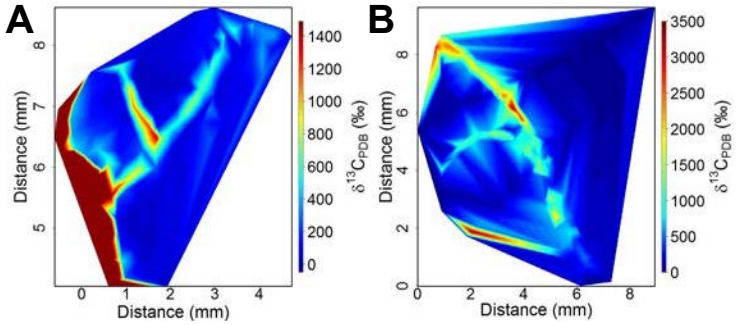


Figure 1: Tracking photosynthate into roots and the rhizosphere. A) Significantly higher $\delta^{13}C$ in one root versus another reveals selective distribution of photosynthate into different roots. B) We are assessing the spatial extent of root exudation into proximal soil using a series of analyses along and parallel to the root, as well as perpendicular transects over the root:soil surface plane. The LA-IRMS approach is minimally destructive such that the sample analyzed above can also be used for elemental analysis (Figure 2).

roots and into the rhizosphere by applying laser ablation isotope ratio mass spectrometry (LA-IRMS). LA-IRMS allowed us to quantify the amount of ^{13}C-labeled, recent photosynthate at specific locations harvested from the soil microcosms (Figure 1). We were able to quantify the variable distribution of photosynthate to different roots and identify specific branch points within the root architecture where roots receiving high versus low photosynthate allocations diverged. Spatial analysis of the soil proximal to the roots revealed the extent of root exudation that would be available for microbial activity. We developed a laser induced breakdown spectroscopy (LIBS) technique to enable coupling of the elemental content of the rhizosphere with the localized extent of exudate additions (revealed via LA-IRMS). LIBS permitted suitable spatial resolution to identify the elemental content of distinct mineral grains within the KBS soil and map elemental content over the root:soil interface (Figure 2). We are currently exploring specific regions within the rhizosphere that contain anomalous elemental distributions and highlight a possible expansion of the root influence into the surrounding soil.

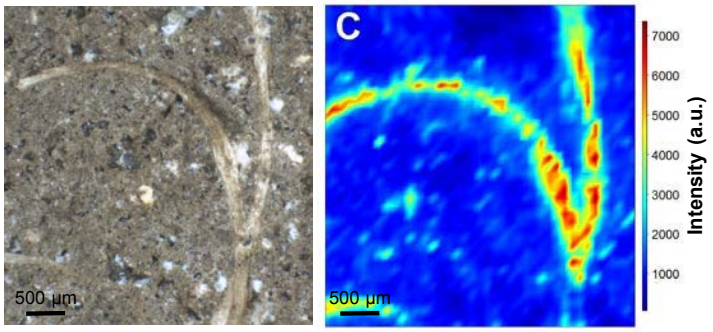


Figure 2: Spatially resolved elemental analysis of a root in soil. A visual microscopy image (left) shows the basic root architecture in the sample extracted from the microcosm. We mapped this visible surface to reveal spatial distributions of a suite of elements including carbon (shown right) as well as various macro- and micro-nutrients and soil constituents.

To help overcome the high spatial heterogeneity within soil and the rhizosphere, we are developing a proteomics-based method for elucidating the spatial variability of soil and rhizosphere-associated microbial communities from a taxonomic and functional perspective. The preparation technique we are using is non-destructive to enable time series sample analysis and provides spatially-resolved proteomics analysis (approximately 1 mm resolution) suitable for basic community taxonomic identification.

Taken together, our developments are allowing us to spatially track photosynthate from leaves, through the roots, and into the rhizosphere, and subsequently characterize the microbial and elemental composition of specific locations that show enhanced carbon deposition.

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).

Genomic divergence in pattern recognition receptors and its implications on endophytic microbial associations in the genus *Salix*

Carly Shanks[1], Timothy B. Yates[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: Identify and characterize species-divergence of immune-suppressing pattern recognition receptors in *Salix spp.*

Recruitment of microbial symbionts by long-lived perennial plants is a key evolutionary strategy to deal with constantly changing environmental conditions as well as cope with limitations in nutrient availability over decades of life time. In these plants, some G-type pattern recognition receptors (PRRs) mediate the highly specific recognition of microbial symbionts to facilitate colonization by triggering suppression of the host defense machinery. Despite evidence to show strong species-specificity in symbiotic interactions, divergence of these PRRs across closely related host species remains poorly understood. In this study, we sampled root endospheres from ten *Salix* genotypes representing seven species, *S. purpurea*, *S. viminalis*, *S. udensis*, *S. integra*, *S. koriyanagi*, *S. alberti*, and *S. suchowensis*, in a fully replicated field site in Geneva, NY. Of these, six were also sampled across multiple replicates in Morgantown West Virginia. Microbial community profiling was performed using 16S rDNA community profiling to establish species-divergence in endospheric microbial communities within the same environment as well across the two sampling sites. To complement this study, divergence in genomic composition of PRRs is being characterized using targeted genotyping of whole-gene deletion events involving PRRs, de novo transcriptomes of the same root samples used in 16S profiling and long-read Nanopore sequencing and assembling of the ten *Salix* genomes. Patterns of genomic divergence in PRR composition and differences in endospheric microbial communities will be presented.

Acknowledgement: This work was funded by the Department of Energy Office of Science under the Early Career Research Program.

Climate Predicts the Dominant Form of Plant-Microbe Symbiosis in Forests at a Global Scale

Brian Steidinger[1], Thomas Crowther[2], Michael Van Nuland[1], Gijsbert Werner[4], Devin Routh[2], Jingjing Liang[3], and **Kabir Peay*[1]** (kpeay@stanford.edu)

[1]Stanford University, CA; [2]ETH Zurich, Switzerland; [3]West Virginia University, WV; [4]University of Oxford, UK

https://mykophile.com/plant-microbe-interactions/mutualistic-niche/

Project Goals: The goal of this research project is to understand how microbial symbioses can be used to maximize the potential of forest trees, such as those in the genus *Populus*, for use in the production of biofuels across a wide range of climatic and environmental conditions. As a first step in guiding this work we aim to establish whether global scale patterns exist in the dominant forms of plant-microbial root symbioses, and identify the climate and environment factors that drive these patterns.

While microbes have long been viewed as agents of disease, recent explorations of the microbiome have led biologists to recognize that beneficial microbes play an equally vital role in maintaining the health of plants and animals. Perhaps the most ubiquitous form of beneficial interaction in terrestrial ecosystems occurs between fungi and plant roots, known as mycorrhizas. Despite the emerging consensus that plant-root symbioses play a critical role in determining the current and future status of forested ecosystems, little progress has been made in mapping their distributions. In one of the earliest efforts to map the functional biogeography of symbiosis, Read (1) categorically identified major biomes by their perceived dominant mycorrhizal associations and proposed that slower decomposition rates at high latitudes favors microbial symbionts with stronger decomposition abilities that enable them to compete directly for organic nitrogen. The existence of a latitudinal gradient in symbiosis would represent one of the major biogeographic patterns on the planet, on par with the latitudinal diversity gradient. Yet definitive support for the Read hypothesis does not yet exist.

As the first step in a multi-component research project to determine how climate, soil environment, and mycorrhizal interactions determine the growth potential of trees in the genus *Populus* we are leading a collaborative, global mapping project of forest root-symbioses. To do this, we have begun analyzing data from a global forest database including 1.2 million plots and 8,000 species, covering all major forested biomes. We conducted a literature search to classify all tree species in the database for their ability to form major types of microbial symbioses (arbuscular mycorrhizal, ectomycorrhizal, ericoid mycorrhizal, nitrogen-fixers, non-mycorrhizal) and used this to estimate the relative abundance of each symbiosis type across the global plot network. By linking the identity of forest trees with their associated symbionts, we generate the first spatially-explicit map of tree symbiosis functional types at a global scale.

Our map shows that there are strong latitudinal gradients in the dominance of major microbial symbiosis types, with arbuscular mycorrhizal fungi dominating lower latitude forests and ectomycorrhizal symbiosis dominating higher latitude forests. We also explored the relationships between these functional groups and broad climate variables using a random forests machine learning approach. Our analyses show that these latitudinal patterns in symbiosis are strongly tied to seasonality of temperature and precipitation.

Future experiments will provide mechanistic understanding of the patterns in these biological communities by directly testing the physiological factors that provide a competitive advantage to different mycorrhizal symbioses using *Populus*, a genus of tree which can associate with both ectomycorrhizal and arbuscular mycorrhizal fungi. By doing so we can provide important basic insights into the way beneficial interactions shape the natural world and also have a direct impact on predicting the suitability of particular sites for bioenergy projects in the light of current and future climate variability.

References

1. Read, D.J. (1991). Mycorrhizas in ecosystems. Experientia *47*, 376-391.

This project is funded by Early Career Award DE-SC0016097 from the Office of Biological and Environmental Research in the DOE Office of Science.

How Does Precipitation Impact the Taxonomic and Functional Diversity of the *Populus trichocarpa* Soil Microbiome?

Michael E. Van Nuland[1] and **Kabir G. Peay**[1]

[1]Stanford University, Stanford, CA

https://mykophile.com; https://www.michaelvannuland.com

Project Goals: Plant-microbe relationships are vital to plant health and nutrition, with strong geographic patterns in their form and function. Our primary goal is to test quantitative predictions about the climatic, environmental, and historical variables that best control *Populus*-microbe relationships (including symbioses between *Populus* and mycorrhizal fungi). We recently collected soil and root samples beneath *P. trichocarpa* across a strong rainfall gradient in the Pacific Northwest, and preliminary data show that ectomycorrhizal fungal colonization declines with precipitation. From this, we are actively pursuing three main research questions with amplicon and metagenomic sequencing.

One of the first global maps of mycorrhizal distributions hypothesized that ectomycorrhizal fungi (EMF) dominate in northern biomes–where plant carbon (C) to nitrogen (N) ratios are high and N mineralization is low–based on their ability to acquire N directly from organic matter (Read 1991). In contrast, arbuscular mycorrhizal fungi (AMF) were thought to have limited ability to decompose organic matter. For three decades, this hand-drawn map largely shaped ecologists' thinking on the causes and consequences of EM- versus AM-dominated forests. However, new data have challenged some of the central assumptions in this early hypothesis, raising questions about the accuracy of these categorical maps and the mechanisms that determine the dominance of the various symbiosis forms on the landscape.

Our goal is to test quantitative predictions about the climatic, environmental, and historical variables that best control *Populus*-microbe symbioses and their impacts on ecosystem function. *Populus* is an ideal candidate for this research because: (a) it is ecologically relevant as an important component of natural forests in the northern hemisphere; (b) *Populus* species vary naturally in their degree of EM symbiosis; and (c) there are tremendous *Populus* genetic resources available due to its commercial importance and potential bioenergy feedstock. We recently conducted a field survey of soils and root samples from sites dominated by *P. trichocarpa* that differ markedly in annual precipitation (range = 213-1674 mm yr^{-1}; **Fig. 1a**). Preliminary data from these sites show that EM colonization declines with precipitation (**Fig. 1b**). As these samples undergo amplicon (16S & ITS2) and metagenomic sequencing, we will focus on three main research questions:

1. How does the taxonomic and functional of soil and root microbial communities vary with precipitation? Because we surveyed across a continuous climate gradient, does diversity respond non-linearly such that we might identify thresholds or tipping points in microbial responses to precipitation?
2. Are there other climate or edaphic factors that structure metagenomic diversity by amplifying or dampening the effect of precipitation (i.e., potential counter-gradients)?

3. Is precipitation associated with changes in the structure/topology of co-occurence networks, and are there connections between the soil and root microbiomes? Is there a "core" belowground microbial network that is associated with *P. trichocarpa*?

First, we expect to see significant turnover associated with precipitation in the community composition and function of soil and EM-enriched metagenomes (*Question 1*). Second, we will use a combination of PCoA and multiple regression approaches to test the strength and direction of additional variables (bioclim variables and soil C, phosphorus (P), organic N, NH_4, pH, cation exchange capacity, and moisture) that might further influence microbial diversity (*Question 2*). These results will provide an important baseline from which we can: (a) predict how soil microbiome diversity and function will differ for other *Populus* species based on assumptions about the strength and importance of environmental gradients underlying their distributions, and (b) compare plant-microbial symbioses among *Populus* species that occur across gradients of different factors and sizes. Third, we will test how microbial network structure varies with precipitation (*Question 3*) using random matrix theory co-occurrence models derived from four site replicates of three discrete precipitation categories: High (>1300 mm yr^{-1}), Medium (<1300 and >1000 mm yr^{-1}), and Low (<1000 mm yr^{-1}). Importantly, this will reveal co-occurrence patterns between specific prokaryotic and fungal taxa that may be indicative of facilitative or antagonistic relationships between these groups. Finally, we will identify any relationships among common microbial taxa that are consistent across all sites, creating a candidate "core" belowground microbial network for *P. trichocarpa*.

Funding provided by DOE BER Early Career Award #DE-SC0016097 (awarded to PI Peay).

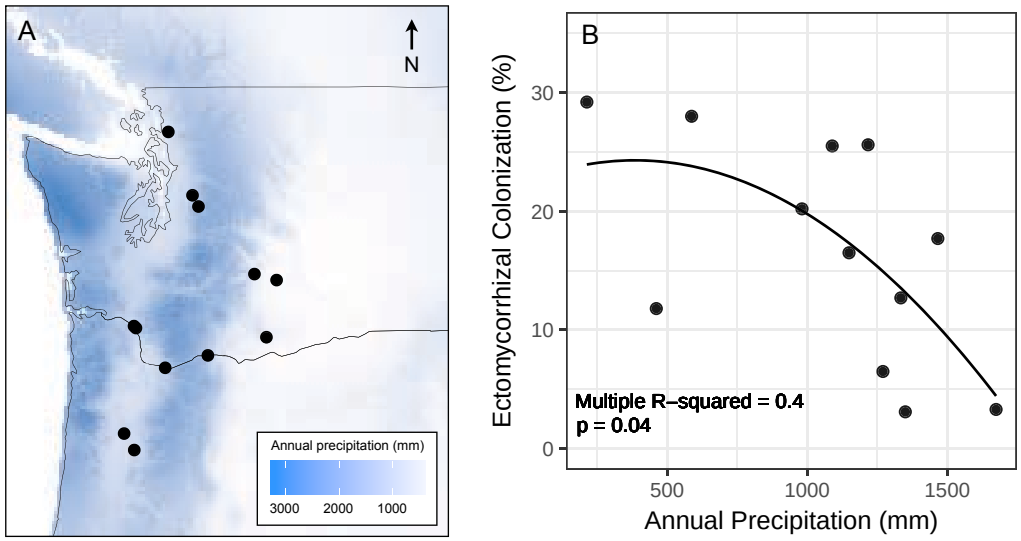


Figure 1. *Populus trichocarpa* sampling locations and ectomycorrhizal colonization across the precipitation gradient. (**A**) We recently sampled soils and roots from *P. trichocarpa* (taking care to trace roots back to focal trees) across 12 sites that vary in the amount of annual precipitation. (**B**) Preliminary data from these sites shows that ectomycorrhizal colonization is negatively related to precipitation.

Using Metabolomics to Uncover Metabolic Pathways Utilized in Nitrogen Stress and Plant-Microbe Interactions in Energy Sorghum.

Amy M. Sheflin[1*] (Amy.Sheflin@colostate.edu), Jessica E. Prenni[1], Rebecca Bart[3], Thomas Brutnell[3], Dawn Chiniquy[9], Asaph Cousins[4], Jeffrey Dangl[5], Ismail Dweikat[2], Andrea Eveland[3], Stephen Kresovich[7], Indrajit Kumar[3], Peng Liu[8], Ellen Marsh[2], Allyn Pella[2], Bala Sonawane[4], Susannah Tringe[9], Chaohui Yuan[8] and **Daniel P. Schachtman[2]**

[1]Colorado State University; [2]University of Nebraska-Lincoln; [3]Donald Danforth Plant Science Center; [4]Washington State University; [5]UNC-Chapel Hill; [6]Boyce Thompson Institute; [7]Clemson University; [8]Iowa State University; [9]Joint Genome Institute-DOE

https://sorghumsysbio.org

Project Goals:

1. **Conduct deep census surveys of root microbiomes concurrent with phenotypic characterizations of a diverse panel of sorghum genotypes to define the microbes associated with the most productive lines under drought and low nitrogen.**
2. **Utilize metabolomics to define the metabolic phenotype of a diverse panel of sorghum genotypes to explore changes at the molecular scale occurring in plants as a response to stress (e.g. drought, nitrogen availability) or plant-microbe interactions.**
3. **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.**

Methods: In the first year of our project, a preliminary field study was performed in which 9 sorghum genotypes were grown under conditions of both low and high nitrogen. Root samples (collected at 2 time points: July and September 2015) were harvested for metabolomics analysis. Molecular profiles were characterized using a combination of targeted and non-targeted mass spectrometry platforms to increase coverage of both primary and secondary metabolites. Targeted LC-MS analyses included key energy substrates in carbon and nitrogen metabolic pathways and phytohormones important in the plant stress response. Non-targeted GC-MS and LC-MS analysis was also performed for broad capture of metabolic changes across various metabolic pathways.

Preliminary data: Results of this analysis reflect alterations within pathways related to the metabolic storage of nitrogen in root cells, an expected result given that plants receiving high nitrogen treatment would likely be storing available reserves. Metabolomics results also suggest incomplete flux through the Shikimate pathway in plants grown under low nitrogen conditions. Complementary investigation of the rhizosphere microbiome revealed a bloom of Pseudomonas in samples from the July harvest that was absent in samples collected in September. Integration of metabolite and microbiome data revealed interesting correlations that also point to incomplete

Shikimate pathway metabolism, co-occurring with the Pseudomonas bloom. Since metabolic end products of this pathway are important precursors for lignification and plant defense hormones, availability could compromise plant defenses. Salicylic acid was significantly decreased with low versus full nitrogen, further supporting that reduced flux through the Shikimate pathway may be altering the plant defense response. Levels of other phytohormones (in roots) with known roles in plant defense did not vary by nitrogen treatment, but were increased in July compared to September. However, it is unclear if defense hormones are elevated in July as a result of interaction with *Pseudomonas* or as a normal progression in plant development over the season.

In summary, the molecular results from this pilot experiment indicate the value of metabolite analysis in understanding mechanistic responses to stress and changes in the soil microbiome. The 2016 field study was designed with increased biological replication (to improve statistical power) as well as the inclusion of a water stress environment. Analysis of these samples is underway and will enable a closer evaluation of these pathways and integration of the molecular results with both microbiome and phenotypic field data.

This project is funded by DOE BER Sustainable Bioenergy Research Program, Award DE-SC0014395, and was also supported by DOE JGI Community Science Program.

Dissecting Genetic Resistance to Willow Leaf Rust (*Melampsora* spp.) in Shrub Willow (*Salix* spp.)

Lawrence B. Smart[1,*], Dustin Wilkerson[1], Craig H. Carlson[1], Chase Crowell[2], and Christine D. Smart[2]

[1]Horticulture Section and [2]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Geneva, NY 14456 USA

http://willow.cals.cornell.edu

Project Goals: The goals of this project are to characterize the genetic diversity of the willow rust pathogen, *Melampsora* spp., and utilize the diversity of *Salix* to identify genes for rust resistance that can be exploited in breeding improved cultivars for bioenergy.

Willow leaf rust (*Melampsora* spp.) is a devastating plant pathogen that can cause considerable premature defoliation in shrub willow (*Salix* spp.) grown as a biomass energy crop, significantly reducing yield. In order to map resistance to willow leaf rust and other traits of interest including yield components, insect resistance, and physiological traits, we have developed three mapping resources, a small association panel planted on three sites, an F_2 mapping population of *Salix purpurea* with 485 progeny planted in Geneva, NY, and eight species hybrid F_1 mapping populations with a common parent of *S. purpurea*. Each of these F_1 hybrid families has either *S. purpurea* 94006 as the female parent (the reference genome) or *S. purpurea* 94001 as the male parent crossed to individuals of *S. suchowensis, S. viminalis, S. udensis, S. integra, S. koriyanagi,* or *S. alberti*. These mapping populations were planted in adjacent field trials in Geneva, NY with 88 to 150 individuals in each family in four randomized complete blocks. Each individual was genotyped by genotyping-by-sequencing and genetic maps developed from segregating SNP markers. During the summers of 2015 and 2017, ratings were collected for damage from imported willow leaf beetle (*Plagiodera versicolora*), potato leafhopper (*Empoasca fabae*), and willow leaf rust severity, as well as stem diameters, height, and specific leaf area measurements. Characterization of QTL controlling these traits will be presented. We have also collected *Melampsora* isolates from across the Northeast and have used GBS to characterize the diversity of the rust population. PacBio genome sequencing of one *Melampsora americana* isolate in underway at JGI.

Funding statement.

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 2015-67009-23957.

Diverse sorghum microbiome discovery and characterization in nitrogen- and water-limited soils for improved biomass production

Dawn Chiniquy,[1] Susannah Tringe[1]* (sgtringe@lbl.gov), Jinglie Zhou,[1] Ellen Marsh, [2] Allyn Pella,[2] Jeff Dangl,[3] and **Daniel Schachtman**[2]

[1] DOE Joint Genome Institute, Walnut Creek, CA; [2] University of Nebraska, Lincoln, NE; and [3]University of North Carolina, Chapel Hill, NC

Project Goals:

- **Generate microbial profiles of diverse sorghum genotype panels.**
- **Examine shifts in microbial community composition over the growing season.**
- **Identify microbes associated with the most productive and efficient sorghum genotypes under both nitrogen- and water- limited conditions.**

Plant roots harbor microbial communities selected from the environment, some of which improve nutrient uptake and growth. *Sorghum bicolor* is a genetically diverse, promising biofuel feedstock with high biomass yield under water and nutrient limitation. Our goal is to establish a systems-level understanding of plant, microbial, and environmental interactions for improved sorghum growth through microbial and plant adaptations to nitrogen- and water-limited environments. We have examined the soil, rhizosphere, and root microbial communities of 30 genetically diverse sorghum genotypes in four sorghum fields (well watered, drought, low nitrogen, full nitrogen) using 16S ribosomal sequencing and to determine the microbial composition. Preliminary results indicate significant differences in the microbial community composition between genotypes, as well as a large seasonal variation. This research will widen knowledge of the genetic and physiological mechanisms involved in plant interactions that shape microbial communities and will lead to strategies for enhancing the nutrient and water use efficiency to create sustainable sorghum biofuel feedstock systems on marginal lands.

https://sorghumsysbio.org/

This project was funded by the DOE BER Sustainable Bioenergy Research Program, Award DE-SC0014395, 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.

Determining the genetic and environmental factors underlying mutualism within a plant-microbiome system: insights from genome sequencing, mass spectrometry imaging and exometabolite characterization

David J. Weston[1*] (westondj@ornl.gov), Alyssa Carrell[1], Dale Pelletier[1], Jeremy Schmutz[2], Dušan Veličković[3], Rosalie K. Chu[3], Christopher R. Anderton[3], Katherine B Louie[2,4], Trent R. Northen[2,4]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]Joint Genome Institute, Walnut Creek, CA; [3]Environmental Molecular Sciences Laboratory Pacific Northwest National Laboratory, Richland, WA; [4]Lawrence Berkeley National Laboratory, Berkeley, CA.

Project Goals: To characterize the *Sphagnum*-diazotroph symbiosis by incorporating plant host *Sphagnum* and microbial genetic variation, variable climatic drivers, and complex communities that scale across biological organizations to regulate peatland carbon and nitrogen cycling.

The importance of plant-microbiome systems on terrestrial carbon and nitrogen processes is perhaps most pronounced in *Sphagnum* dominated ecosystems, which occupy 3% of the Earth's land surface yet store approximately 25% of terrestrial carbon as recalcitrant organic matter (i.e., peat). The foundation plant genus *Sphagnum* is responsible for much of the primary production in peatland ecosystems and produces recalcitrant dead organic matter. Together with associated N_2-fixing microorganisms, *Sphagnum* contributes to substantial peatland nitrogen inputs. *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 are initiating a 5-year project using synthetic communities, genotype-to-phenotype associations, and metabolic characterization to address two overarching hypotheses, 1) *Sphagnum* host and diazotroph genetic variations 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.

The first year of this project is centered on developing the genetic and analytical resources necessary for synthetic community and transfer community experimentation. Through a JGI CSP, we now have draft genomes of 15 *Sphagnum* species and (re)sequencing for a 200-member pedigree is currently underway. On the microbial side, 72 *Sphagnum* associated heterotrophic bacteria strains, along with 12 cyanobacteria and 30 putative methanotrophs have been isolated on multiple medium types including N free. The JGI collaboration includes exometabolite characterization from a cross feeding experiment among *Sphagnum*, a cyanobacterium and a fungal partner. A pilot optimization experiment confirms that our synthetic community approach is amenable to exometabolite characterization with the identification of over 65 key metabolites. A complementary collaboration with EMSL is now adding spatial characterization of target metabolites among the tri-partite members through different phases of symbiosis using matrix assisted laser desorption/ionization (MALDI) mass spectrometry imaging (MSI) along with

liquid extraction surface analysis (LESA). This approach allows for accurate metabolite structural information from LESA and high spatial resolution from MALDI. Equipped with these resources, our team is now initiating experimentation to address the quantitative genetics of symbiosis, metabolite exchange and codependency, and ultimately how environmental perturbations interact with plant and microbial genetics to form and break symbiosis.

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

Successional Dynamics of Grassland Microbial Communities in Response to Warming, Precipitation Alternation, and Clipping

Xue Guo[1], Qun Gao[1], Mengting Yuan[1], Zhou Shi[1], Xishu Zhou[1], **Liyou Wu[1]*(lywuou.edu)**, **Zhili He**[1], Joy D. Van Nostrand[1], Lauren Hale[1], Daliang Ning[1], Aifen Zhou[1], Jiajie Feng[1], Ying Fu[1], Chang-Gyo Jung[1], Tong Yuan[1], Xuanyu Tao[1], **Jim Cole**[2], **James M. Tiedje**[2], **Edward A.G. Schuur**[3], **Konstantinos Konstantinidis**[4], **Yiqi Luo**[1], and **Jizhong Zhou**[1]

[1]Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK; [2]Center for Microbial Ecology, Michigan State University, East Lansing, MI; [3]Center for Ecosystem Sciences and Society, Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ; [4]Center for Bioinformatics and Computational Genomics and School of Biology, Georgia Institute of Technology, Atlanta, GA

http://www.ou.edu/ieg.html

Project goal: The overall goal of this project is to advance systems-level predictive understanding of the feedbacks of belowground microbial communities to multiple climate change factors and their impacts on soil carbon (C) cycling processes. The specific objectives are to: (i) reveal the responses of soil microbial communities to climate warming and soil moisture alteration in both tundra and temperate grassland ecosystems; (ii) determine temperature sensitivity of recalcitrant C decomposition and characteristics of the microbial degraders; and (iii) develop integrated bioinformatics and modeling approaches to scale information across different organizational levels.

As a part of the integrated project, here we present results from; i) field experiments established in 2008 in a temperate grassland of central Oklahoma to reveal the influence of elevated temperature, altered precipitation and plant biomass clipping on long-term succession of plant and microbial communities, and ii) soil respirations under long-term warming.

Long-term succession of microbial communities. To determine successional dynamics of microbial communities in response to warming, clipping, altered precipitation and their combinations, 264 annual soil samples from 2009 to 2014 were analyzed by sequencing of 16S rRNA genes for bacteria and archaea, ITS regions for fungi, and 18S rRNA genes for other micro-eukaryotes excluding fungi, and by functional gene arrays (GeoChip 5.0). Our analyses indicated that global change factors including warming, clipping, half precipitation, double precipitation and most of their combinations differently shifted the temporal successional patterns of the taxonomic composition and phylogenetic structure of different microbial populations. Importantly, among these global change factors, climate warming played a dominant role in accelerating divergent succession of all soil microbial communities as evidenced that experimental warming enhanced microbial temporal divergences under the context of various global changes. Secondly, our results also showed these global change factors and most of their interactions significantly ($P < 0.05$) changed species-time relationships (STRs) of different soil microbial populations including bacteria, fungi and micro-eukaryotes. And climate warming significantly ($P < 0.05$) promotes temporal scaling rates (STR exponent) of all microbial populations even under the context of various global changes. All of these results indicated that warming plays a dominant role in accelerating temporal succession rates of soil microbial communities.

In general, both deterministic (i.e. species traits, interspecies interactions and environmental conditions) and stochastic (i.e. birth, death, extinction, and speciation) processes simultaneously influence the assembly of local communities. To quantify the relative importance of deterministic vs. stochastic processes in

shaping soil microbial communities under climate change treatments, stochastic ratios were calculated based on taxonomic and phylogenetic metrics. Our results indicated that the stochastic processes contributed the majority of community variations across all treatments in taxonomic (48.2% ~ 68.1%) and phylogenetic (54.1% ~ 86.5%) levels across 6 years of manipulated experiment. Secondly, warming significantly ($P < 0.05$) decreased the relative importance of stochastic processes (4.6-17.6%) in controlling bacterial and fungal community composition, and the similar results were also obtained in many comparisons between combined warming treatments and their corresponding controls. Furthermore, the roles of stochastic processes in all microbial populations under single warming and most combined warming treatments decreased substantially along with time, indicating that deterministic processes play more important roles over time in controlling microbial communities in response to climate warming. These findings have important implications for predicting ecological consequences of climate changes. Because warming reduced stochasticity over time, the communities could converge quickly to less variable states, and hence the future community states could be more predictable under warmed climate if we have enough knowledge on the successional trajectories of the contemporary microbial communities.

Network analysis of microbial temporal successions. Soil microorganisms coexist in complex arrays in which interactions among members are essential for community assembly and ecosystem. However, most of studies in the last decades examined the responses of ecological communities to climate changes by quantifying diversity, but whether and how climate changes affect ecological community organization and the interactions among members of ecological communities, particularly microbial communities, remains elusive. Our network analysis revealed that warming predominantly led to larger and more complex bacterial and fungal networks along time under the context of various global change factors, as indicated that the warmed soil networks significantly increased in size ($r^2 = 0.836$, $P = 0.011$) and connectivity ($r^2 = 0.916$, $P = 0.003$) over time. Secondly, more and larger modules with more positive and negative links were found in the warmed soil networks, suggesting that more mutualistic and competitive interactions may occur under climate warming. Thirdly, we identified more putative keystone taxa including module hubs, connectors and network hubs in the warmed soil networks. Almost all of these keystone taxa had low relative abundances (0.002% ~ 2.593%), suggesting low-abundance taxa may significantly contribute to soil microbial function. Intriguingly, no network hubs were identified in any of non-warmed soil networks, but one network hub was detected only in the last year of warmed soil network, which were assigned to the typical oligotrophic phylum *Acidobacteria* and exhibited 91% identity to an isolate of *Acidobacteria* Gp16. These results indicated that oligotrophic taxa may play more important roles than those copiotrophic taxa in the warmed soil communities.

Soil respiration under long-term warming. The warming-induced stimulation of soil carbon release through respiration could amplify future increases in atmospheric CO_2 levels and associated climate warming. However, the magnitude and persistence of the warming-induced stimulation of soil carbon release under future climates, including more frequent climate extremes, are poorly understood. In the field experiment, ecosystem C fluxes and soil respirations were monitored monthly from 2009 to 2016, during which rainfall greatly changed in different years, containing one of the driest years and the wettest year on the historic record. Therefore, we obtained an opportunity to evaluate the magnitude and persistence of the warming-induced stimulation of soil carbon release under future climates, including more frequent climate extremes. Our results showed, after an eight-year field study of a native, tall-grass prairie ecosystem under warming and control, that although the warming response of autotrophic respiration is differentially shifted under drier (-36%) and wetter (+8%) conditions, the stimulation of heterotrophic respiration by warming is consistent and persistent under these extreme conditions (18%). Integrated metagenomic analyses indicated that warming-induced shifts in microbial community structure and function were consistent and did not significantly vary under drier or wetter conditions. More genes for soil carbon decomposition and nitrogen-cycling processes were consistently ($P < 0.05$) stimulated by warming under dry, normal and wet conditions. Furthermore, Mantel test revealed that very strong correlations were observed between community structure and ecosystem functional processes, such as plant GPP and total soil respiration, and the functional

gene groups involved in C degradation, N cycling, P utilization and S cycling ($P < 0.05$). Remarkably, soil heterotrophic microbial respiration specifically showed strong correlations with the genes involved in degrading recalcitrant C, including vanillin, lignin and cellulose, suggesting that the increased heterotrophic respiration may contribute to the degradation of recalcitrant C. All of these results indicated that the dynamic succession of microbial community greatly contributed to the consistent warming-induced stimulation of heterotrophic respiration under climate extremes. Our findings indicate that microbial dynamic succession may be important for understanding the magnitude and persistence of ecosystem-scale responses to global change.

Funding statement: *Our work is supported by the U.S. DOE Office of Science, Biological and Environmental Research Division (BER), Genomic Science Program, Award No. DE‐SC0010715.*

Plant-Microbe Interfaces: Genetic mapping and genomic resequencing identify a lectin receptor-like kinase as a regulator of *Populus-Laccaria bicolor* interaction

Jay Chen,[1*] (chenj@ornl.gov) Wellington Muchero,[1] Jessy Labbé,[1] Xiaohan Yang,[1] Jennifer L. Morrell-Falvey,[1] Olaf Czarnecki,[1] Juan Wang,[1] Xiaoping Wang,[1] Anthony C. Bryan,[1] Kaijie Zheng,[1] Yongil Yang,[1] Meng Xie,[1] Jin Zhang,[1] Dongfang Wang,[1] Sara Jawdy,[1] Lee Gunter,[1] Wendy Schackwitz,[2] Joel Martin,[2] Erika Lindquist,[2] Ting Li,[1] Zhihao Zhang,[1] Priya Ranjan,[1] Daniel A. Jacobson,[1] Timothy J. Tschaplinski,[1] Kerrie Barry,[2] François Le Tacon,[3] Jeremy Schmutz,[2,4] and **Gerald A. Tuskan**[1]

[1] Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2] Department of Energy Joint Genome Institute, Walnut Creek, CA; [3] Institut National de la Recherche Agronomique (INRA) et Université de Lorraine, Labex ARBRE, UMR 1136, Champenoux, France; [4]HudsonAlpha Institute for Biotechnology, Huntsville, AL

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) developing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Populus species, as keystone members of boreal and temperate ecosystems, interact with a wide variety of microbes. The *Populus-Laccaria bicolor* system has emerged as an excellent system of choice for studying plant-ectomycorrhizal interactions aided by the availability of both *Populus* and *Laccaria* reference genomes and genetic tools. Modes of action and molecular mechanisms underlying ectomycorrhizal interactions are poorly understood in this and all other plant-mycorrhizal systems though progress has been made in recent years. Active recruitment and acceptance of mycorrhization have been proposed to occur in a species-specific manner and *L. bicolor* has been found to preferentially colonize *P. trichocarpa* over *P. deltoides*. We therefore hypothesized the existence of distinct genetic loci that are present in *P. trichocarpa* but absent in *P. deltoides* and harbor high-fidelity recognition mechanisms for *L. bicolor.*

Using genetic mapping and re-sequencing, we identified a whole-gene deletion event in *Populus* that was associated with a decrease in colonization by the ectomycorrhizal fungus *L. bicolor*. This locus contains a gene encoding a lectin receptor-like kinase, designated as PtLecRLK1. The role of PtLecRLK1 mediating mycorrhization was validated via heterologous expression in a non-host species for *L. bicolor*, *Arabidopsis*, conveying the ability in the transgenic plants to accept interstitial hyphal growth and Hartig net-like structure formation by *L. bicolor*. Expression and metabolomics analyses indicated that plant defense-related genes and metabolites were down-regulated in *Arabidopsis* plants expressing *PtLecRLK1*. These results uncover an important molecular step in the establishment of symbiotic plant-fungal associations and provide a molecular target for rational design of mycorrhizal symbiosis in economically important crops to enhance water and nutrient acquisition in marginal lands.

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.

Time-Dependent Tradeoff In Microbe-Plant Interactions Under Drought: Short-Term Stress Dampening Versus Long-Term Accelerated Mortality

Danielle E. Marias[1] (daniellem@lanl.gov), La Verne Gallegos Graves[1], Sanna Sevanto[2], **John Dunbar**[1]

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos
[2]Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos

http://www.lanl.gov/org/padste/adcles/bioscience/bioenergy-biome-sciences/soil-carbon.php

Project goal:
Identify soil microbial community traits that influence ecosystem sensitivity to climate stress.

Abstract

The LANL Genomic Science 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.

The work presented here addresses BER Grand Challenge 4.1: *characterize biogeochemical exchanges driven by plant-microbe interactions and evaluate their process-level impacts and sensitivity to climate change*. We are attempting to identify microbial community traits that influence plant function because the production and longevity of plant biomass carbon in terrestrial ecosystems is a key component of climate feedbacks. Discovery of microbial-plant interactions that influence cycling of C and N is a step towards inclusion of microbial processes into earth system models (e.g. BER Grand Challenge 4.4).

Although soil microbial communities can improve plant productivity and survival, the interaction is not well-understood. A major knowledge gap is how soil microbial communities alter plant responses to drought. To address this gap, we examined plant responses to drought as a function of soil microbial community composition. We hypothesized that microbial community composition will affect plant functions including germination, growth, drought tolerance, photosynthesis, stomatal conductance, and wilting.

To test this hypothesis, seeds of a fast-growing C4 grass, blue grama (*Bouteloua gracilis*), were planted in sterilized sand (the control treatment) or in sterilized sand inoculated with soil microbial communities from 15 geographically distinct soils. The 15 soil microbial communities were selected to inoculate plants because they exhibited functional differences in carbon cycling. After substantial growth (14 weeks), drought was imposed. Measurements of plant physiology and soil microbial communities were taken before, during, and after drought.

One month after planting, germination and shoot height were significantly greater in inoculated plants compared to controls. During the initial stage of drought, inoculated plants were more productive (indicated by greater growth and rates of photosynthesis) than controls because the

inoculated plants maintained higher soil moisture. This delayed the onset of drought but also caused inoculated plants to have lower drought tolerance than controls. As drought persisted and soil moisture declined to zero, inoculated plants were more susceptible to drought as indicated by significantly lower stomatal conductance, greater wilting, and faster mortality compared to controls. Among the inoculated plants, plant functional differences (plant height and drought tolerance) were linked to the composition of the soil communities used for inoculation.

These data suggest that soil microbes promote *B. gracilis*' opportunistic growth strategy—i.e., fast plant growth when water is available and dieback when water is scarce. **In this system, microbes initially enhanced plant productivity and dampened drought stress, but accelerated mortality over longer timescales of persistent stress, suggesting tradeoffs linked to patterns of stress oscillation.** This work provides an unexpected insight into plant-microbe-soil interactions. The underlying mechanisms are relevant to understanding responses of terrestrial ecosystems to climate change. Investigation of microbial community traits linked to differences in plant susceptibility to drought is ongoing.

Funding Statement: The information presented in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2015LANLF260) to Los Alamos National Laboratory, the Los Alamos National Laboratory Center for Space and Earth Science Chick Keller Post-Doc Fellowship to DM, and LDRD-ER project number 20160373ER.

Plant-Microbe Interfaces: Developments in integrated omics to link microbial metabolism to community structure/function in plant/microbial systems

Robert Hettich,[1] (hettichrl@ornl.gov) Sanjeev Dahal,[2] Dale Pelletier,[2] Greg Hurst,[1] Nancy Engle,[2] Timothy Tschaplinski,[2] and **Mitchel Doktycz**[2]

[1]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The availability of new mass spectrometric platforms for higher performance and enhanced measurement speeds can be combined with optimization of new sample preparation protocols to dramatically enhance the impact of omics measurements for plant-microbe research. For example, a traditional deep proteome measurement (LC/LC-MS/MS) might require more than 20 hours of MS measurement time. Thus, there is a strong need for development of methods to dramatically increase measurement throughput while maintaining measurement depth. To this end, we have installed and demonstrated a new proteome measurement scheme based on a high-performance LC-MS/MS instrument (ThermoFisher Q-Exactive Orbitrap mass spectrometer). This instrument provides measurement speeds of up to 12 Hz, enhanced dynamic range measurement, mass accuracies in the low ppm range, and mass resolutions exceeding 100,000 (FWHM). This provides the ability to execute multi-dimensional LC/LC-MS/MS experiments in a higher throughput fashion (up to 5-6 complete runs per day), with equal or superior mass accuracy and measurement depth. By utilizing autosamplers for automated sample injection campaigns, we have been able to dramatically improve the precision and reproducibility of proteome measurement, which greatly enhances differential proteome measurements. These advances increase proteome measurement throughput by 6-10X overall, thereby opening new avenues for the larger sample campaigns expected for both microbial and plant studies. To extend the range of experimental information possible for metabolomic studies, we have undertaken implementation of LC-MS measurement capabilities for a range of semi-volatile and non-volatile metabolites to complement our existing GC-MS capabilities. To this end, we have implemented a high performance reverse phase LC-MS/MS approach (based on Q-Exactive Orbitrap MS) for bacterial and plant metabolomic studies.

This instrument is configured with an autosampler/high performance LC system to permit fairly large-scale metabolomic measurement campaigns with high mass accuracy, deep dynamic range, and tandem mass spectrometry fragmentation. These advances should help transform the capacity for executing detailed "integrated omics characterization campaigns," such as the study discussed below.

We have applied these new tasks and procedures to understand the microbial metabolism of higher order salicylates. One of the defining characteristics of *Populus* is the production of these secondary metabolites, which are involved in host defense and signaling mechanisms. Two interesting bacterial strains, *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16, are involved in the degradation of simple phenolic compounds such as salicin and salicyl alcohol respectively through uncharacterized mechanisms. We hypothesize that these two strains can utilize pathways to cooperatively grow on salicin. We acquired both metabolomic and proteomic data for mono- and co-cultures of *Rahnella* sp. OV744 and *Pseudomonas* sp. GM16. The hypothesis tested with this experiment was that in a co-culture of salicin-containing medium, OV744 would metabolize salicin to salicyl alcohol, which would then be used as a carbon source by GM16. Culture experiments were designed to test this hypothesis, as well as to potentially provide information on the underlying enzymes and pathways employed by the two species in metabolizing salicylates. When monitoring a microbial co-culture in salicin using qPCR strain specific primers, growth of both OV744 and GM16 was observed, suggesting a cross-feeding of salicyl alcohol leading to growth of both strains. By using various quantitative and multi-omics (proteomics and metabolomics) approaches, we demonstrated that the individual pathways for salicin and salicyl alcohol metabolism are present in OV744 and GM16, respectively. Furthermore, we establish that in salicin co-culture, these two strains utilize both of these pathways to cross-feed salicyl alcohol. The research presents one of the potential mechanisms for microbial transformation of salicin and potentially other more complex HOS, and demonstrates the intricate interactions that occur within the plant microbiome.

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: AI-GWAPA, explainable AI-based approaches to genome wide phytobiome association

Piet Jones,[1,2]* (jonespc@ornl.gov) Benjamin Garcia,[1] Wellington Muchero,[1] Jay Chen,[1] **Gerald Tuskan**[1] and **Daniel Jacobson**[1,2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN

http://pmi.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The phytobiome consists of the plant, organismal communities, and their environment. The interactions between these have significant effects on observable measurable traits that have potential economic and sustainability implications. A better systems-level understanding of the beneficial and antagonistic relationships between these components will enhance our capacity to influence these systems to produce desirable and impactful traits.

In Artificial Intelligence – Genome Wide Association Phytobiome Analysis (AI-GWAPA), a project that utilizes machine learning, deep learning and general artificial intelligence (AI) techniques, the goal is to elucidate the interactions between microbial and viral constituents of the 1000 member *Populus Trichocarpa* population arrayed in common gardens in the Pacific Northwest. Metatranscriptome samples from leaf, xylem and root along with approximately 28 million SNPs called across the population allows us to associate host genetic variants to microbial/viral constituents. Sample-specific networks, which is a machine learning approach that allows us to model contributions that a genotype has to putative pathogenic-mutualistic relationships between taxa, is being used. Furthermore, we utilize an AI approach by training a deep learning neural network to estimate putative phytobiome-derived protein interactions

among the host proteome. Together these approaches allow us to improve our fundamental understanding of the relationships between the plant and its phytobiome.

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: Phytobiome and transcriptional adapation of *Populus deltoides* to acute progressive drought and cyclic drought

Benjamin J. Garcia,[1] Jessy L. Labbé,[1] Piet Jones,[1,2] Paul E. Abraham,[3] Ian Hodge,[1,4] Sharlee Climer,[5] Sara Jawdy,[1] Lee Gunter,[1] Gerald A. Tuskan,[1] Xiaohan Yang,[1] Timothy J. Tschaplinski,[1] and **Daniel Jacobson**[1,2]* (jacobsonda@ornl.gov)

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN; [3]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [4]Department of Computer Science, Stanford University, Standford, CA; [5]Department of Computer Science and Engineering, Washington University, St. Louis, MO.

http://pmi.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Plant drought stress causes systematic changes to photosynthesis, metabolism, and it's phytobiome. Additionally, drought affects plants in both a species-specific and water deficit driven manner, causing the response to drought to be dependent both on how much drought is being experienced and on any adaptation to prior drought exposure. As such, to understand the effect of drought on plants requires assessing drought response in multiple conditions, such as progressive acute drought and recurrent cyclic drought, and at different levels of severity. In this study, we have utilized RNA sequencing to identify changes to the plant transcriptome and the phytobiome during both acute progressive drought and cyclic drought at multiple severities. We have identified that the drought response ranges from increased transcripts related to photosynthesis and metabolic activity in mild acute drought to decreased transcripts related to photosynthesis and metabolic impairment in severe drought. Moreover, while water deficit is a main driver of transcriptional responses in severe drought, there are increases in reactive species metabolism and photosynthetic transcripts in cyclic severe drought compared to acute severe drought, independent of water deficit. Lastly, the phytobiome is more separated by the cyclic or

acute nature of the drought rather than the severity of the drought, with drought-resistance and plant growth promoting organisms at higher abundances in cyclic drought.

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: Altered root metabolome composition impacts microbiome composition in *Populus PdKOR1* RNAi plants

Udaya C Kalluri* (kalluriudayc@ornl.gov), Jessy Labbe, Allison Veach, Daniel Yip, Amber Bible, Jennifer Morrell-Falvey, Nancy Engle, Dale Pelletier, Timothy Tschaplinski, Christopher Schadt, **Gerald Tuskan**, and **Mitchel Doktycz**

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

In the present study, we examined the effect of altered carbon partitioning and allocation in the *Populus* host due to interactions with individual beneficial microbes as well as on the overall root endosphere and rhizosphere associated microbiome. The altered secondary metabolism host type used in this study are transgenic *Populus PdKOR1* RNAi plants that are downregulated in an endo-β-1,4-glucanase gene family member. Gas chromatography-mass spectrometry profiles of *PdKOR1* plants showed a higher phenolic and salicylic acid content, and reduced lignin, sugars, shikimic acid and maleic acid content relative to non-transgenic control. Co-culture with the fungal mutualist, *Laccaria bicolor*, showed enhanced mycorrhization rate and improved biomass production in *PdKOR* plants (Kalluri et al. 2016). This suggested strong potential for impact on the broader microbial community that the plant interacts with in field settings. In contrast, the colonization rate of a previously characterized Gammaproteobacterial isolate, *Pantoea* YR343, was lower in *PdKOR1* RNAi plants.

To test whether altered root metabolome has an effect on the microbiome associated with roots under fields settings, we collected root samples from independent ramets of field-grown *PdKOR1* RNAi and control plants and performed Illumina MiSeq 16S rRNA gene sequencing. Bacterial community composition, as measured by Bray-Curtis dissimilarity, differed between *PdKOR1* RNAi and control rhizospheres and roots. *Actinobacteria*, and the family *Micromonosporaceae*, were significantly more abundant, whereas *Nitrospirae* were reduced in *PdKOR1* RNAi plant

rhizosphere. These findings from single isolate co-culture experiments as well as field-based microbiome analyses show the relevance of host carbon partitioning and metabolome composition, including phenolic, sugar, amino acid and fatty-acid composition, on concomitant alterations in root-associated microbial communities. We are currently conducting metagenomics of leaf, stem and root and soil samples to capture the genetic diversity of the microbial communities (bacterial and fungal) associated with specific plant tissue type/niche, and examining the differential molecular pathways underlying the differential association of microbes via RNA-Seq analysis and microbial isolate sequencing approaches (Kalluri et al. 2018).

In conclusion, our study shows the significance of plant metabolome composition on shaping the associated microbiome and is prompting new hypotheses and experiments that will address the cascading effects of host genotype, root tissue environment, root exudate composition on interactions with soil microbiome.

References

Kalluri et al. Frontiers in Plant Sciences. 2016. doi: 10.3389/fpls.2016.01455.
Kalluri et al. FY18 DOE Joint Genome Institute Community Sequencing Project.

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: The key fungal lineage of the Russulaceae, a new resource for untangling and linking beneficial plant-fungal associations and ecosystem functions

Brian Looney,[1] Marek Piatek,[2] Piet Jones,[2] Alan Kuo,[3] Kerrie Barry,[3] Igor Griegoriev,[3] Francis Martin,[1] Dan Jacobson,[2] **Mitchel Doktycz,[2]** and **Jessy Labbé[2*]** (Labbejj@ornl.gov)

[1]INRA, Tree-Microbe Interactions, UMR1136, Champenoux, France; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [3]Joint Genome Institute, Walnut Creek, CA

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The *Populus* root microbiome harbors a diverse community of ectomycorrhizal fungi (EMF) that significantly increases nutrient uptake and acquisition by the plant host while also providing protection against antagonistic parasites. Over 30 genera of EMF are known to associate with *Populus*, including many groups of mushroom-producing families including Amanitaceae, Boletaceae, Cortinariaceae, Tricholomataceae, and Russulaceae. The family Russulaceae is considered an iconic lineage of mostly mushroom-forming basidiomycetes due to their importance as edible mushrooms in many parts of the world and their ubiquity as ectomycorrhizal symbionts in both temperate and tropical forested biomes. While much research has been focused on this group, a comprehensive or cohesive synthesis by which to understand the functional diversity of the group has yet to develop. Interest in ectomycorrhizal fungi, of which Russulaceae is a key lineage, is prodigious due to the important roles they play as plant root mutualists in ecosystem functioning, global carbon sequestration, and a potential role in technology development toward environmental sustainability. As one of the most species-diverse ectomycorrhizal lineages, the Russulaceae has recently been the focus of a dense sampling and genome sequencing initiative with the Joint Genome Institute to untangle their functional roles and test whether functional niche specialization exists for independent lineages of ectomycorrhizal fungi. Here we present the reference genomes produced by this project, some comparative analyses, along with the development of co-culture systems, which are a

tremendous and promising resource for the identification of potential genetic controls for ectomycorrhizal association and decomposition.

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: Characterization of IAA biosynthesis pathways in *Populus*-associated microbes

Jennifer Morrell-Falvey[1,3,6]*(morrelljl1@ornl.gov), David Garcia,[2] Kasey Estenson,[3] Miriam Land,[1] Xiaolin Cheng,[4] Gregory Hurst,[5] Robert Standaert,[1,6] Amber Bible,[6] and **Mitchel Doktycz**[1,2]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]The Bredesen Center for Interdisciplinary Research and Graudate Education, University of Tennessee, Knoxville, TN; [3]UT-ORNL Graduate School of Genome Science and Technology, University of Tennessee, Knoxville TN, [4]Medicinal Chemistry and Pharmacognosy, The Ohio State University, Columbus, OH; [5]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge TN; [6]Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville TN

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Populus deltoides (poplar) hosts a diverse microbiome that influences its growth and productivity. Many plant-associated bacteria have the ability to produce phytohormones, such as indole-3-acetic acid (IAA). However, elucidating the pathways associated with secondary metabolite production in microorganisms is an ongoing challenge. Multiple IAA biosynthetic pathways have been described in microbes, most of which require the precursor tryptophan. The tryptophan-dependent pathways include the indole-3-acetonitrile (IAN) pathway, the indole-3-acetamide (IAM) pathway, the tryptophan side-chain oxidase (TSO) pathway, the indole-3-pyruvate (IPA) pathway, and the tryptamine pathway. We are pursuing the use of genetic knockouts and –omics measurements along with cell-free metabolic engineering coupled with bioinformatic searches of genome databases and protein-ligand docking simulations in order to identify the proteins most likely to be involved in a metabolic pathway. In particular, genomic

analysis was used to predict that *Pantoea* sp. YR343 synthesizes IAA using the indole-3-pyruvate (IPA) pathway. This prediction was tested using a combination of proteomics, metabolomics and genetics. To better understand IAA biosynthesis and the effects of IAA exposure on cell physiology, we characterized proteomes of *Pantoea* sp. YR343 grown in the presence of tryptophan or IAA. These data indicate that indole-3-pyruvate decarboxylase (IpdC), a key enzyme in the IPA pathway, is upregulated in the presence of tryptophan and IAA. Metabolite profiles of wildtype cells showed the production of IPA, IAA, and tryptophol, which is also consistent with an active IPA pathway. Finally, we constructed a mutant in *Pantoea* sp. YR343 in which the *ipdC* gene was deleted. This mutant was unable to produce tryptophol, consistent with a loss of IpdC activity, but was still able to produce IAA (20% of wildtype levels). This result suggests the possibility of an alternate pathway or the production of IAA by a non-enzymatic route. To examine this possibility and to aid in the assignment of candidate enzymes in the pathways, we employed protein-ligand docking simulations. The resulting computationally predicted set of enzymes were then expressed and are being tested in cell free systems for their ability to produce IAA.

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.

A Mycorrhizal Helper Bacteria Increases Mycorrhization in Aspen Seedlings by Modulating Expression of Defense Response Genes in Roots

Peter E. Larsen[1,2*](plarsen@anl.gov), Shalaka Shinde[1], Sarah Zerbs[1], Frank R. Collart[1], Jonathan Cumming[3], and **Philippe Noirot**[1].

[1] Argonne National Laboratory, Biosciences Division, IL; [2] University of Illinois at Chicago, Department of Bioengineering, Chicago IL; [3] West Virginia University, Biology Department, Morgantown

Project Goals: Use transcriptomic analysis of a laboratory rhizosphere community to identify the molecular mechanisms in aspen roots by which the mycorrhizal helper bacteria SBW25 induces increased mycorrhization of aspen seedling roots by the ectomycorrhizal fungi Laccaria.

Beneficial rhizosphere communities of symbiotic fungi and bacteria protect plants from a variety of biotic and abiotic stresses. In return, plants provide photosynthetically fixed carbon to rhizosphere community members, making these subsurface communities not only crucial for the health and stability of terrestrial ecosystems, but also important components of the carbon cycle. One of the beneficial roles bacteria can play in the rhizosphere community is that of Mycorrhizal Helper Bacteria (MHB). MHB facilitate positive symbiotic interactions between sol fungi and plant roots. While the positive effects of MHB on mycorrhizal interactions have been well characterized, the specific molecular mechanisms by which MHB enhance mycorrhizal interactions are less well established. We have developed a sand-pot, tripartite laboratory community of aspen seedlings, ectomycorrhizal fungi *Laccaria bicolor* (Laccaria), and the Plant Growth Promoting (PGP) [1] bacteria *Pseudomonas fluorescens* SBW25 (SBW25), suitable for HTP omics analysis. Here, we utilize this laboratory community model system to propose the specific molecular mechanisms in aspen roots that drive MHB activity.

Our experimental design is comprised of four biological conditions: aspen seedlings, aspen with Laccaria, aspen with SBW25, and aspen with Laccaria and SBW25. Rhizosphere communities were cultured for 63 days, at which time aspen seedling phenotype data and community transcriptomes were collected. We found that the presence of SBW25 significantly increased mycorrhization of aspen seedling roots by Laccaria by almost 2-fold, demonstrating that SBW25 is a MHB under our experimental conditions. The analysis of community transcriptomic data revealed that both Laccaria and SBW25 mRNA were detectable at statistically significant levels in the sand pot rhizosphere after 63 days. Using computational and modeling methods we have previously developed for the analysis of rhizosphere community interactions [2,3], we identified six clusters of co-regulated genes is aspen seedling roots across biological conditions. One gene cluster, significantly enriched for anti-fungal defense response genes, was found to be highly up-regulated when Laccaria is present alone, but down-regulated when both Laccaria and SBW25 are present. This suggests that the inhibition by SBW25 of fungal defense-response genes in aspen roots is a key component to SBW25's MHB activity. However, this finding does not implicate the specific regulatory mechanisms by which fungal defense response genes are controlled.

By considering the expression pattern of each gene cluster as a conditional statement, it becomes possible to arrange the clusters of differentially-expressed co-regulated genes as a logic circuit (**Figure 1**), highlighting potentially causal relationships between rhizosphere community composition, patterns of gene regulation, and observed MHB activity. The mechanism that enables MHB activity in this predicted network is a XOR logic gate that down-regulates global-regulation genes when both Laccaria and SBW25 are present in the rhizosphere community, but not when Laccaria or SBW25 are present individually. Statistically significant enrichment for genes with the annotation 'pollen recognition' indicate excellent

candidates for the molecular mechanism of the predicted XOR gate that down-regulates either the global gene regulation or the fungal defense response gene cluster in the gene regulation logic circuit diagram.

The results of this analysis propose three key findings: (i) SBW25 is a MHB under our experimental conditions, (ii) increased mycorrhization by Laccaria when SBW25 is also present occurs in response to the down-regulation of aspen root fungal-defense response genes, and (iii) the molecular mechanisms by which aspen roots detect the rhizosphere community involve sensors with homology to 'pollen detection' sensors. Future analyses will further characterize these sensors and identify the specific ligands present in the rhizosphere that aspen roots used to collect information about the composition of the rhizosphere community.

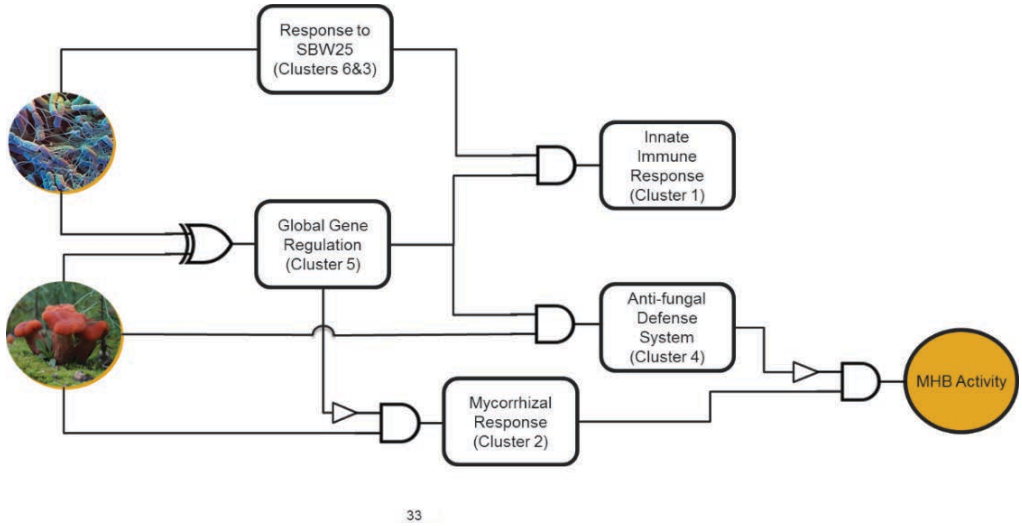


Figure 1. Predicted aspen root gene regulatory network for MHB activity as a logic circuit diagram. The proposed regulatory interactions between co-expressed gene clusters is represented as a logic circuit diagram. In the network, circles to left indicate presence or absence of Laccaria (red mushrooms) or SBW25 (blue microscopy image). Rectangles are co-regulated gene clusters identified by K-means clustering and gene cluster function descriptions are taken from statistically enriched functional annotations within gene clusters. An edge between nodes indicate a predicted causal relationships between gene clusters, inferred from observed patterns of gene regulation.

References

[1] S Shinde, JR Cumming, FR Collart, PH Noirot, PE Larsen (2017). Pseudomonas fluorescens Transportome Is Linked to Strain-Specific Plant Growth Promotion in Aspen Seedlings under Nutrient Stress. *Frontiers in Plant Science* 8.

[2] PE Larsen, A Sreedasyam, G Trivedi, S Desai, Y Dai, LJ Cseke, ... (2016). Multi-omics approach identifies molecular mechanisms of plant-fungus mycorrhizal interaction. *Frontiers in Plant Science* 6.

[3] PE Larsen, FR Collart, Y Dai (2015). Predicting ecological roles in the rhizosphere using metabolome and transportome modeling. *PloS One* 10 (9), e0132837.

Funding Statement: This contribution originates in part from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory. The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357.

Plant-Microbe Interfaces: Characterization of natural products from the *Populus* microbiome

Patricia M. Blair, **Dale A. Pelletier*** (pelletierda@ornl.gov), Miriam L. Land, Tse-Yuan Lu, and **Mitchel J. Doktycz**

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

http://pmiweb.ornl.gov/

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The *Populus* root microbiome is an incredibly diverse community, comprising organisms from across plant, animal, oomycete, fungal, viral, archaeal and bacterial taxa. Bacteria from the soil are known to harbor many gene clusters encoding complex natural products that can act as signaling molecules, antibiotics, and antifungals. We set out to characterize the natural product potential of bacteria from a plant root community in order to understand its biosynthetic diversity as well as begin to determine keystone members and associated molecules that regulate community structure and plant health. The model plant system used was *Populus*, the first fully genome-sequenced tree species having an already well-characterized root metagenome.

We first considered metagenomic samples collected from the roots of *P. deltoides* and examined the overall bacterial diversity and natural product variety, comparing to other plant and human microbiomes to show the predicted species and natural product richness. The diversity of bacteria in the plant microbiome is greater than the well-studied human gut microbiome, and the organisms within this community have greater biosynthetic potential as well. We next utilized the fully sequenced genomes of over 400 bacterial isolates, representing the four major bacterial phyla in the metagenome, connecting molecules to genomes and surveying the overall natural product potential. While some species harbor greater numbers of clusters, especially Actinobacteria of the genus *Streptomyces*, we found over 10 clusters per organism on average, with over 4000 predicted clusters. Comparison to known natural product gene clusters revealed that only 1% of clusters produced an already-characterized secondary metabolite, revealing the great potential to discover compounds with novel structures and possibly novel activities.

About 15% of the predicted clusters could not be connected to known natural products classes, revealing the potential to discover structurally novel metabolites. Of the remaining clusters, many grouped within classes known to produce molecules with antibiotic or antifungal properties. Ribosomally synthesized and post-translationally modified peptide natural products were both prevalent in the collection and divergent from previously characterized molecules. These natural products, which are peptides that have been modified by additional enzymes, were the most abundant class of natural product identified, being more common than even nonribosomal peptide and polyketide clusters. Lactones and siderophores, molecules known to be important for quorum sensing and iron acquisition, respectively, were prevalent in the genomes and suggest a high level of communication as well as pressure to compete for resources. These molecules are involved not just in microbe-microbe interactions, but have consequences in signaling to the plant to enable colonization and nutrient exchange.

While the diversity and richness of natural product gene clusters within the genome-sequenced fraction of the *Populus* microbiome reveals an additional layer of complexity in the community, the presence of a gene cluster does not necessarily mean that the compound will be produced. Thus, we set out to determine if genome-sequenced *Streptomyces* isolates were capable of producing compounds with antifungal and antibiotic activity under laboratory conditions. *Streptomyces* sp. OK461 and *Streptomyces* sp. OK006 each contain an identical lasso peptide gene cluster with known Gram-positive antibacterial activity. Both isolates inhibited the growth of Gram-positive isolates, and methanolic extracts containing the lasso peptide replicate this activity. Thus, complex molecules such as the lasso peptide are involved in interspecies signaling and communication, shaping community structure and therefore influencing the overall health and growth of the host plant.

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.

How Plant-Microbial Interactions Shape Resource Allocation in Perennial Bioenergy Grasses

Christine Hawkes[1]* (chawkes@austin.utexas.edu), Elise W. Connor[1], Hannah Giauque[1], Kelly Craven[2], **Erin Nuccio[3], Jennifer Pett-Ridge[3]**

[1]Microbial and Plant Ecology, University of Texas at Austin, Austin, TX; [2]Noble Research Institute, Ardmore, OK; [3]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.

Microbe-host interactions have long shaped the biosphere and geosphere and can fluctuate between mutualistic, commensal and parasitic. In bioenergy systems, microbial symbionts may have direct effects on crop productivity, resilience, or stress tolerance of the feedstock crop. We hypothesize that foliar endophytes regulate a diverse set of genes putatively related to C use, stress, and secondary metabolites. In the first stage of our newly-funded SFA Plant Bioenergy focus area, we will determine how foliar fungal endophytes influence belowground carbon cycling and feedstock robustness during water and nutrient stress by altering aboveground foliar processes.

To understand the effects of foliar endophytes on stress tolerance in bioenergy plants, we have identified fungal endophytes in the leaves of *P. virgatum* and *P. hallii* exposed to drought (5% vs. 15% soil moisture). These isolates represent a broad spectrum of symbionts, ranging from mutualists to antagonists. The transpiration efficiency of *P. hallii* infected by these endophytes varied from 0.1x to 25x, and wilt resistance from 4 to 22 days, which can scale up to enormous effects at the ecosystem level both directly and via constraints on C available for transfer belowground. Plant physiological outcomes were largely predicted by independent fungal traits related to osmotic stress tolerance, and results were generally consistent between *P. virgatum* and *P. hallii*. Based on *P. hallii* transcriptomics for plants grown under drought with one of two mutualists or three antagonists, we found no genes stimulated by all five fungi, and antagonists

shared more genes than mutualists under drought. These preliminary results suggest there are likely multiple genetic pathways by which endophytes benefit plants and fewer pathways by which they cause harm. Identifying the genetic and ecological mechanisms that drive fungal endophyte effects on host plants will provide both a basic biological understanding of these symbioses, and could ultimately guide use of such biotic mutualisms to improve bioenergy plant production.

In our future work, we will determine if there are functional categories of belowground C allocation that are generalizable to different types of symbioses (antagonists to mutualists), and the robustness of these strategies to drought. While endophytic fungi are ubiquitous plant symbionts and play important roles in plant physiology and responses to environmental stress, we have little understanding of their belowground impacts. We hypothesize that both foliar fungal endophytes will affect belowground C allocation during drought, with expected trade-offs for plant stress tolerance vs. biomass or stress and nutrient transfer. The degree and type of impact belowground will depend on the specific fungus and the mechanism by which they affect drought stress tolerance/resilience in the plant. We will initially test our hypothesis in a simplified sterile system with our model *Panicum* grass, *P. hallii*, grown individually with each of our well-characterized 20 foliar fungal endophytes in C-free sand soils with low inorganic nutrient addition. Our goal is to predict these interspecies exchanges under deficient conditions to support sustainable and predictable 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 Bioenergy SFA, FWP SCW1039. The work was supported by funding from the Texas Ecolab program to HG and CVH and from the University of Texas Green Fee and the University of Texas Plant Biology Graduate program to EWC.

Plant Microbe Interfaces: Emerging analytical techniques for controlling and monitoring architectural changes in developing multi-kingdom systems

Jayde A. Aufrecht,[1,2] Amber Bible,[2] Damith E. W. Patabadige[1,3], **Jennifer Morrell-Falvey**,[1,2,3] **Dale A. Pelletier,** [1,2,3] **Scott T. Retterer,** [1,2,3] ***(rettererst@ornl.gov), and Mitchel J. Doktycz[1,2,3]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN; [3]Center for Nanophase Materials Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these events manifest themselves within the spatial, structural, and temporal complexity of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) studying model microbial systems to elucidate *Populus*-microbial interactions at the molecular level and dissect the signals and pathways responsible for initiating and maintaining microbial relationships and, 3) developing metabolic and genomic models of these interactions to aid in interpreting the molecular mechanisms that shape the *Populus*-microbial interface.

The structure and development of natural systems depend on complex physical and chemical signal exchanges over space and time. The ephemeral nature of these signals and the dynamically changing host and community phenotypes associated with these events requires the use of complementary imaging, chemical, and genetic sampling strategies that can be used to capture and correlate these properties over time. In this work, we present a suite of techniques and emerging technologies to monitor and interpret the drivers of host development and colonization. A combination of microfluidic and 3D printed plant-microbe co-culture platforms enabled the characterization of root colonization patterns by plant-growth promoting bacteria isolated from *Populus*. These platforms have the potential to provide insights into how physical and chemical features of the host shape colonization, niche formation, and microbe-microbe interactions within the rhizosphere. Likewise, the influence of constructed communities on a dynamic host phenotype can be quantified. Ultimately, this suite of multi-modal sampling and measurement techniques will allow us to understand the influence of biochemical transport and exchange on the functional composition and structure of these dynamic plant-microbe systems in tractable model environments that capture critical features of their natural habitats.

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: The contribution of host chemotype as a driver of rhizosphere microbial community structure

Allison M. Veach, Timothy J. Tschaplinski, Nancy L. Engle, Zamin K. Yang, Daniel Z. Yip, Reese H. Morris, Melissa A. Cregger, Gerald A. Tuskan, **Christopher W. Schadt*** (schadtcw@ornl.gov)

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serve as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships, and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

A diverse community of bacteria and fungi are closely associated with the roots of plants in the rhizosphere. Beyond carbon exchange between plant roots and their microbiota, the degree to which rhizosphere microbiomes are shaped by other plant host properties is relatively unexplored. *Populus* spp. trees produce characteristically high levels of salicylic acid that is complexed into higher-order salicylate conjugates that are actively involved in host defense, but also may be utilized as a nutritive source by exogenous soil microbial taxa or be inhibitory to others. We conducted a greenhouse-based experiment using 12 *Populus trichocarpa* genotypes that vary in higher-order salicylate concentration and composition (i.e., total phenolics, catechin, salicylic acid, salicortin, salicin, α-salicyloylsalicin, tremuloidin, and populin) to better understand how the genotype and chemotype of *P. trichocarpa* influence rhizosphere microbiome assembly and composition. We planted genotypes in soils originating from 2 locations: Corvallis and Clatskanie, Oregon that differed in their soil physical properties and nutrient concentrations. To assess the relative importance of host genotype vs. soil origin, we planted 5 replicate cuttings per genotype (N = 120) in these 2 soil types (2:1 sterile sand:soil inoculum) and allowed plants to grow for 4 months under greenhouse conditions. At the end of the experiment, leaf chlorophyll content, leaf growth (number of new leaves since transplant), and net photosynthetic rate differed across genotypes and soil type ($P < 0.01$). *P. trichocarpa* chemotypes were confirmed for roots from experimental samples via GC-MS analysis, and

showed host total salicylate concentrations varying from 1221 – 10610 μg/g FW, tremuloidin varying from 17 – 225 μg/g FW, and populin ranging from 0 – 8.9 ug/g FW plant tissue, depending on plant genotype. Total ectomycorrhizal colonization of root tips differed between genotypes ($P = 0.01$), but not soil type. Specifically, the genotype with the lowest root salicortin and salicin concentrations had significantly greater ectomycorrhizal colonization (10 ± 1.5% colonized) compared to three other genotypes with greater higher-order salicylate concentrations. Bacterial diversity within rhizospheres differed among genotypes and soil origin ($p < 0.01$), whereas fungal diversity did not differ among genotypes, but did differ by soil origin ($p < 0.01$). Bacterial diversity decreased with root tremuloidin concentrations ($p = 0.04$) and increased with populin concentration ($p = 0.02$), whereas fungal diversity was not correlated with any higher-order salicylate ($p > 0.09$). Regarding microbial community composition within rhizospheres, soil origin accounts for the majority of variation for bacteria (perMANOVA: $R^2 = 0.52$, $p = 0.001$) and fungi (perMANOVA: $R^2 = 0.40$, $p = 0.001$) with genotype secondarily influential (perMANOVA: bacteria $R^2 = 0.09$, $p = 0.003$; fungi $R^2 = 0.09$, $p = 0.02$). Furthermore, based on distance-based redundancy analyses, bacterial community composition was influenced by both tremuloidin and populin, whereas fungal community composition was influenced by salicortin and salicylic acid ($p < 0.05$). These results suggest that not only do salicylic acid metabolites produced by *Populus trichocarpa* roots directly impact belowground microbial community structure, but also that that bacteria and fungi respond differentially to derivatives of salicylic acid, specifically populin and tremuloidin.

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: Dissecting the compatibility and diversity of the mycobiome of *Populus trichocarpa*

Alejandro Rojas*(alejo.rojas@duke.edu) Khalid Hameed,[1] Posy Busby,[1,5] Gregory Bonito[6], Hui-Ling Liao[7], Brian Looney,[2,4] Jessy Labbé,[2] Jessica Velez, [3] Christopher Schadt, [1,3] Claire Veneault-Fourrey,[4] Aurélie Deveau,[4] Annegret Kohler,[4] Francis Martin,[4] and **Rytas Vilgalys[1]**

[1]Duke University, Durham, NC; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [3]Graduate School of Science and Technology, University of Tennessee, Knoxville, TN; [4]INRA and the Université de Lorriaine, Labex ARBRE, UMR 1136, Champenoux, France; [5]Oregon State University, Corvallis, OR, [6]Michigan State University, Plant Soil and Microbial Sciences, East Lansing, MI, and [7]North Florida Research and Education Center, University of Florida, Quincy, FL

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The *Populus* root microbiome harbors a diverse community of endophytic and ectomycorrhizal fungi that promote nutrient acquisition and plant health. In collaboration with the JGI, we have isolated and sequenced the genomes of over 50 ectomycorrhizal and endophytic fungi isolated from *Populus* roots. These genomes are beings used to interrogate the metagenomics of North American soil fungal communities with other cottonwood species *P. nigra* collected from France in order to identify core groups of fungi associated with *Populus*. *Populus* genotypes often vary in their ability to form symbioses with different root-associated fungal taxa, and one aim of our studies is to identify the genetic determinants which underly host-specificity. A diverse collection of root associated fungi are being tested to evaluate their compatibility with different *Populus* genotypes representing different geographical ecotypes of *P. trichocarpa*. Selected species of *Populus*-associated EMF including *Lactarius, Hebeloma, Cenococcum, Laccaria* and *Paxillus* were inoculated on eight *P. trichocarpa* genotypes to address plant-fungal compatibility and function of the *Populus* mycobiome by in vitro synthesis of ectomycorrhiza and by using split-root systems with stable isotope tracing. The systems are under evaluation and will facilitate the understanding of *Populus*-fungal associations assessing the effects on the plant host using isotope methods, transcriptomics and assessment of plant health and fungal colonization. By using different split-root and in vitro systems in combination with genetic

tools and isotopic tracing methods, these studies will provide insight into *Populus*-fungal associations and their development in this model tree genetic 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.

Plant-Microbe Interfaces: Developing a synthetic community system to test preferential allocation to nitrogen-fixing bacteria

Dave Weston,[1]*(westondj@ornl.gov) Dale Pelletier,[1] Alyssa Carrell,[1] Collin Timm,[1] Yasuhiro Oda,[2] Carolyn Harwood,[2] Sara Jawdy,[1] Lee Gunter,[1] Shen Lu,[1] and **Mitchel Doktycz**[1]

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [2]Department of Microbiology, University of Washington, Seattle, WA

http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Plant-microbiome function results from complex interactions among microbial members, host plant genetics/physiology and surrounding environmental conditions. Once the plant – microbiome is established (after initial colonization events), a key question is whether the host plant can actively discriminate among mutualists by 'rewarding' beneficial members through preferential allocation of carbon. To begin to address this question, we developed and characterized reference microbiomes consisting of 10 bacterial strains representing abundant and functionally diverse orders that were isolated and genome-sequenced, from natural *Populus* microbiota. These include potential diazotrophic strains. Subsequently we investigated the ability of five diazotrophs to colonize and function in a *P. trichocarpa* host. After three weeks of co-culture conditions, our results showed strain-specific preferences for plant organs and tissues as indicated by CFUs and qPCR analyses. *Rahnella* sp. OV588 was determined to be a robust colonizer of *Populus* tissues and a *nifH* deletion mutant was generated for that strain for further functional characterization. Using a germ-free magenta box system with calcined clay substrate, *P. trichocarpa* genotype 819 was either uninoculated or inoculated with wild-type OV588 or a *nifH* deletion mutant. All experimental combinations were provided with either Hoagland's complete nutrient medium (with N) and without N. In the no N condition, plants cultured with OV588 showed a 48% increase in total plant dry weight relative to uninoculated plants or plants with the *nifH* mutant strain. Furthermore, acetylene reduction assays of whole *Populus* plants

showed ~5-fold increase in ethylene production when colonized by wild-type OV588 compared to uninoculated or *nifH* mutant-inoculated plants. There were no significant differences in ethylene production, total dry weight or chlorophyll concentration when N was included in the growth medium regardless of the bacterial inoculum. Experiments using $^{15}N_2$ gas in the plant growth chamber are underway. Our work here suggests that N_2 is being fixed by *Rahnella* sp. OV588, which contributes to enhanced plant growth under N-limiting conditions. Future studies will use a dual label of $^{13}CO_2$ and $^{15}N_2$ within a split root system to address questions of preferential allocation within a community context.

Microbiome diversity and assembly in the phyllosphere of perennial bioenergy crops

Keara L Grady[1], Nejc Stopnisek[1], Jackson W Sorensen[1]*(sorens75@msu.edu), **Ashley Shade**[1]

[1]Michigan State University, East Lansing, Michigan

https://www.glbrc.org/

http://ashley17061.wixsite.com/shadelab

Project Goals: Our overarching objective is to understand the interactive dynamics of bioenergy crop phyllosphere microorganisms with each other, their hosts, and their environment. Toward this objective, we have two project goals. (1) Characterize the intra-annual temporal dynamics of phyllosphere microbial communities of switchgrass and miscanthus. (2) Quantify the contributions of soil microorganisms to switchgrass and miscanthus phyllosphere assembly.

The aerial surface of plants, known as the phyllosphere, makes up a large portion of the terrestrial microbial environment (an estimated 10^8 km^2 globally). Phyllosphere microbes have can influence ecosystem services by moderating biogeochemical fluxes of greenhouse gasses. They also can benefit their hosts by improving plant stress tolerance or promoting the production of plant growth hormones. However, we have limited information about the composition, dynamics and functions of bioenergy crop phyllosphere microbiota. Here, we characterize the diversity and assembly of microbial communities present in the phyllosphere and nearby soil of two perennial cellulosic feedstocks for biofuel - switchgrass (*Panicum virgatum*) and miscanthus (*Miscanthus x giganteus*) over the course of the 2016 growing season at the Great Lakes Bioenergy intensive cropping sites (Kellogg Biological Station, Hickory Corners, MI). We sampled both nitrogen-fertilized and unfertilized sub-plots to assess the effect of N-fertilization on phyllosphere community structure and function. We used 16S rRNA gene sequencing to assess microbiome diversity and structure. We found that community membership of the phyllosphere was distinct from that of soils. The leaves harbored relatively simple communities of less than 100 taxa, and had a core microbiome (persistent and abundant taxa) that included several Proteobacteria. There were directional changes in the phyllosphere communities over the growing season, suggesting that these consortia were fit for life on the leaf surface rather than vagabonds randomly assembling from the air or soil. The highest numbers of taxa were observed early and late in the growing season, and the lowest at mid-summer. Miscanthus and switchgrass phyllosphere communities became more differentiated over time, indicating an influence of host on the late-season phyllosphere community. There was no strong influence of fertilization on the phyllosphere structure. Understanding the dynamics of community assembly, structure, and function of phyllosphere microbiota is a first step toward managing these communities to benefit their hosts. In the case of bioenergy crops, we interrogate these communities for functions that may improve crop productivity or wellness, or enable efficient cellulose degradation from plant material.

This work was funded in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and the DOE OBP Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830)

Plant and Biogeochemical Controls on the Switchgrass Microbiome: Perspectives from a fine-scale time series

Sarah Evans[1,*](evanssa6@msu.edu), Jim Cole,[1] Maren Friesen, [1] Steven Gougherty, [1] Lisa Tiemann,[1]

[1]Michigan State University

www.rhizosphere.msu.edu

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

Nitrogen (N) is the most commonly limiting nutrient for plants, especially in marginal lands. These lands are unsuitable for food crops because of low productivity and vulnerability to environmental stress. The introduction of perennial bioenergy cropping systems (PBCS) in marginal lands can improve whole system N use efficiency and N retention, while also contributing to energy sustainability without competition with food. However, little is known overall about N-cycling and associated microbial function in marginal land biofuel cropping systems. As part of a project studying Microbial-Mediated Perennial Rhizosphere Nitrogen Transformations (MMPRNT), we have begun to characterize N-cycling microbial communities and associated plant and soil biogeochemical properties in six marginal land sites in Michigan and Wisconsin. Sites are part of the DOE Great Lakes Bioenergy Research Center (GLBRC), and including different cropping systems (switchgrass, prairie, control) and fertilized and unfertilized plots. A unique aspect of this study is the temporal resolution at which we measured properties; at our focal site, we looked at these properties in 2-week time intervals, at another site on a monthly basis, and at all sites 1x/season.

We found that overall, site was the strongest factor explaining microbial and biogeochemical dynamics, but that microbial communities and soil nitrogen pools varied widely on relatively short temporal scales. For instance, microbial community composition varied as much over time as it did in fertilized and unfertilized plots in a single site. Fertilizer affected soil and microbial characteristics after being applied in spring, but we saw surprisingly few long-term effects of this treatment on soil or plant traits. An improved method for measuring free-living N fixation revealed that N-fixation is occurring in switchgrass, and may be especially prominent near senescence. These field data will complement other lab and greenhouse mesocosms and field manipulations in our project, which will be used to parse out mechanisms for many of these patterns.

References

1. Bowsher, A, Sarah Evans, Lisa K. Tiemann, Maren L. Friesen. 2017. Effects of soil nitrogen availability on rhizodeposition in plants: a review. *Plant and Soil.* https://link.springer.com/article/10.1007/s11104-017-3497-1

Funding statement.

This work is funded by the Department of Energy Genomic Sciences program DE-FOA-001207

Notes on abstract:

- Note the placement of superscripts in the authors and affiliations.
- URL above should be specific to the project. More than one URL is permitted.
- **References** can be **Publications** instead, if needed. Use any common style for these citations.

From Leaves to Roots to Microbes: How Sorghum Responds to Drought

Peggy G. Lemaux (lemauxpg@berkeley.edu)[1*], Jeffrey 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; [2]University of California, Kearney Ag Research & Extension Center, Parlier; [3]USDA-ARS Plant Gene Expression Center, Albany CA; [4]University of California, West Side Research & Extension Center, Five Points; [5]EMSL-Pacific Northwest National Laboratory, Richland WA; and [6]DOE Joint Genome Institute, Walnut Creek CA.

Project Goals: Transcriptomic and epigenetic control mechanisms during temporal and spatial responses to water-limiting conditions are being studied in leaves and roots of field-grown, pre-flowering and post-flowering drought-tolerant *Sorghum bicolor* (L.) Moench varieties. Changes in bacterial and fungal communities, associated with drought-stressed sorghum, are also being studied in bulk soil, rhizosphere and roots. Our goal is to understand roles transcriptomic and epigenetic signals play in acclimation to and recovery from pre- and post-flowering drought, revealed through transcriptional networks and molecular profiles *in planta*, using RNA-Seq, ChIP-Seq, BS-Seq, proteomics, metabolomics, and histone profiling. Also, impact of microbial populations is being inferred from metagenomics, metatranscriptomics and metabolomics. Ultimately, we will identify genes and molecular markers to develop genetic strategies for improving drought tolerance in sorghum and other crops. Cumulative data will be used to devise models to better predict and control the roles and interactions of transcriptional regulation, epigenetics and the microbiome in sorghum's response to drought.

Based primarily on phenotype, classical breeding and mutagenesis have focused mainly on changing a plant's DNA to modify desired traits. Increasing published data, however, show that environmental responses and plant development are also mediated by epigenetics that does not involve a change in DNA sequence. Critical to EPICON research, both transcriptomic and epigenetic changes play major roles in regulating drought responses. Thus, impact of our studies will increase given the likelihood that frequency and severity of drought will increase with climate change, posing major challenges for world agricultural productivity.

Plant exposure to abiotic stresses triggers cascades of transcriptomic and epigenetic changes. EPICON efforts focus on discovering the temporal and spatial influences that transcriptomic, epigenetic, proteomic, metabolomic and microbial signals play in acclimation to and recovery from drought. Our studies focus on *Sorghum bicolor* (L.) Moench, a widely cultivated cereal noted for drought and flood tolerance, that also offers advantages as a bioenergy feedstock because of its relatively reduced environmental footprint and its flexibility in bioenergy uses. For our experimental design, drought conditions are imposed in the field in California's Central Valley, where summer rainfall is rare. In two years of field experiments, one pre-flowering and one post-flowering drought-tolerant sorghum cultivars was planted in a replicated split plot design. Both varieties were subjected to three watering treatments: normal watering and pre- and post-flowering drought treatments. Phenotypic measurements, like grain and biomass yields and flowering times and growth rates, were taken from early June to late September. Analyses of years 1 and 2 data indicate that, compared to control plants, those exposed to pre- and post-flowering drought stress had notable differences in drought effects, particularly comparing the pre-flowering drought tolerant variety to the post-flowering drought-tolerant variety.

For molecular analyses, leaf and root samples, collected weekly, were used to track spatiotemporal changes in transcriptomic, epigenetic, metabolomic and proteomic footprints. From transcriptional profiling of year 1 samples, widespread adaptations were seen in all developmental stages, along with rapid transcriptional changes after watering of droughted plants. Also, imposing drought after flowering also caused similarly rapid transcriptional changes. Genes, with developmental changes in transcription during pre-flowering drought, were classified into distinct sets of temporal patterns, revealing varied reactions to pre-flowering drought. Ongoing work is attempting to identify biological function of these responses and their relation to other epigenetic and phenotypic properties of the plant. Analysis of such responses to post-flowering drought is underway. Initial analysis of BS-Seq data in leaves revealed many regions where methylation changes are in concert with plant development, including differences between the two varieties. Drought effects on such methylation patterns is under investigation.

Corresponding metabolomic and proteomic analyses are in progress, employing a method where proteins and metabolites are extracted from the same samples. Global proteomic analyses with extensive multiplexing and fractionation enable deep proteome coverage; metabolites from the same samples are also being analyzed. LC-MS was used to analyze histones, purified in untargeted fashion, enabling discovery of novel drought-related histone posttranslational modifications. A comprehensive molecular map of soil organic matter was generated to assess differences in chemical composition across the field and rhizosphere-mediated processes.

Using soil, root and rhizosphere samples collected as above, microbiome changes were studied, following drought and re-watering treatments. Gene function was inferred fro m shotgun metagenomic and metatranscriptomic analyses. Using year 1 bacterial data, pre-flowering drought was shown to lead to rapid changes in community composition, with relative enrichment of most Gram-positive bacterial lineages. This enrichment is reversible, leading to reversion, within one week after re-watering, to a state dominated by Gram-negative lineages. The enrichment in Gram-positive lineages was accompanied by increases in transcriptional activity, specifically for gene functions related to carbohydrate and amino acid transport and metabolism. Through metabolomic analyses, drought-treated roots were shown to be enriched in many of the same carbohydrate and amino acid metabolites, suggesting interplay between plant metabolism and bacterial community activity. From year 1 fungal data, both pre-flowering and post-flowering drought exerted significant effects on fungal diversity and community composition. Pre-flowering drought induced an ~100-fold enrichment of an *Acremonium* fungus in roots; post-flowering drought had an ~2-fold enrichment of a *Gibberella* fungus. Rhizosphere fungal community largely followed patterns of the root fungal community, but, the soil fungal community was not substantially affected. The leaf fungal community was not affected by pre-flowering drought, but it was affected following post-flowering drought. Symbiotic arbuscular mycorrhizal fungi were found in root, rhizosphere and soil; however, their diversity and community composition were not affected by drought.

Over the entire project period, collected data will provide a deeper understanding of the restructuring of the metabolic and regulatory landscape during drought, including impacts of microbes. We hope to identify key transcriptional and epigenetic regulators, controlling drought tolerance, and the relationship between sorghum and its associated microbes. Identified genetic targets, regulatory pathways and beneficial microbial symbionts will be used to improve growth and biomass production of sorghum and other crops under water-limiting conditions.

The EPICON project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Improving Nitrogen-Use Efficiency of Switchgrass Production: Past, Present, Future

Michael Udvardi[1*] (mudvardi@noble.org), Kelly Craven[1], Malay Saha[1] and **Gerald A. Tuskan[2]**

[1] Noble Research Institute, Ardmore, Oklahoma; [2] The Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN, USA

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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Nitrogen (N) is an essential macronutrient for plant growth, although low levels of available-N in most soils limit plant production. Use of synthetic N-fertilizers has largely eliminated this limitation in many agricultural systems, albeit at substantial economic and environmental cost. Moreover, approximately 1-2% of the world's fossil fuel consumption drives industrial fertilizer-N production. Sustained production of plant biomass for biofuels will require constant inputs of N into agricultural systems. To minimize the energy and environmental footprints of biofuel production, we seek plant genotypes that produce high biomass with low fertilizer-N and other chemical inputs. Switchgrass, a highly-productive perennial C_4 grass, is adapted to diverse environments across North America and has been targeted for development as a biofuel feedstock crop. We aim to increase the nitrogen-use efficiency of switchgrass plants and production systems, using a variety of strategies. One strategy is to identify plant genotypes that not only produce large amounts of biomass per unit of fertilizer-N added to the system, but also remobilize a large fraction of the N from shoots to roots during annual plant senescence prior to harvest. This conserves N in the plant for use in the next growth season and reduces the amount of additional N-fertilizer required to sustain production. We found substantial natural variation for N-remobilization efficiency among switchgrass accessions, ranging from 20-61% of shoot-N (Yang et al., *Bioenerg Res* (2009) 2:57–266). Interestingly, some of the most efficient genotypes were also the most productive in terms of biomass, making them ideal parents for breeding programs. Another strategy that we are pursuing is the use of nitrogen-fixing bacteria/endophytes of switchgrass that we plan to engineer to deliver substantial amounts of fixed-N to plants, which would also reduce the need for industrial N-fertilizer.

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.

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

Rebecca Bart[1*](rbart@danforthcenter.org), **David Braun**[2], **Noah Fahlgren**,[1] **Malia Gehan**,[1] and **Wilfred Vermerris**[4]

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

Project Goals: The primary goal of this project is to understand the effect of altered cell wall and sugar content on plant susceptibility phenotypes. To accomplish this goal, sorghum and *Xanthomonas* will be established as a model pathosystem.

High-biomass-yielding crops may harbor modifications to cell walls, which are a major barrier to pathogen entry, and to the tissue distribution of sugars, which are the pathogen's food source; hence they are likely to present previously unseen challenges for disease resistance. *Xanthomonas* is a known pathogen of sorghum (*Sorghum bicolor* (L.) Moench), though the incidence and impact of the disease has historically been low. We are working to establish the sorghum – *Xanthomonas* pathosystem as a model for deducing how latent microbial pathogens might exploit key biofuel crop traits. Our approach will be to quantitatively model the disease triangle that describes sorghum, pathogenic bacteria, and the environment. Field and laboratory experiments will be 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 will provide a holistic view of this pathosystem within variable environments. Further, transcriptomics will be 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 will be analyzed during pathogen invasion. This research will reveal the mechanisms underlying tolerance to pathogens that must be maintained during biofuel trait optimization. The proposed research will yield a detailed understanding of the impact of bioenergy relevant traits on pathogen susceptibility. This is a necessary first step towards the development of novel routes for disease control that can be deployed in parallel with targeted alterations to sugar and cell wall composition during bioenergy crop improvement and breeding efforts.

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

Trait components of whole plant water use efficiency are defined by unique, environmentally responsive genetic signatures in the model C_4 grass *Setaria*

Ivan Baxter[1,3]* (ivan.baxter@ars.usda.gov), Max Feldman[1],Asaph B. Cousins[2], Patrick Z. Ellsworth[2]and **Tom Brutnell**[1]

[1]Donald Danforth Plant Science Center, St. Louis, MO 63132, [2]Washington State University, Pullman, WA 99163, [3]USDA-ARS, Plant Genetics Research Unit, St. Louis, Missouri, USA_`

url: http://foxmillet.org

Project Goals: Objectives of this project are to 1) identify the gene regulatory and metabolic networks important for adaptation to low water availability and high-density plantings and to understand the physiology underlying these adaptations and 2) develop technologies to precisely control gene insertion and replacements events for large scale engineering of pathways in model and target feedstocks.

Abstract: Plant growth and water use are interrelated processes influenced by the genetic control of both plant morphological and biochemical characteristics. Improving plant water use efficiency (WUE) to sustain growth in different environments is an important breeding objective that can improve crop yields and enhance agricultural sustainability. However, genetic improvements of WUE using traditional methods have proven difficult due to low throughput and environmental heterogeneity encountered in field settings. To overcome these limitations the study presented here utilizes a high-throughput phenotyping platform to quantify plant size and water use of an interspecific *Setaria italica* x *Setaria viridis* recombinant inbred line population at daily intervals in both well-watered and water-limited conditions. Our findings indicate that measurements of plant size and water use in this system are strongly correlated; therefore, a linear modeling approach was used to partition this relationship into predicted values of plant size given water use and deviations from this relationship at the genotype level. The resulting traits describing plant size, water use and WUE were all heritable and responsive to soil water availability, allowing for a genetic dissection of the components of plant WUE under different watering treatments. Linkage mapping identified major loci underlying two different pleiotropic components of WUE. This study indicates that alleles controlling WUE derived from both wild and domesticated accessions of the model C_4 species *Setaria* can be utilized to predictably modulate trait values given a specified precipitation regime.

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

Modulating hemicellulose to improve bioenergy crop

Sang-Jin Kim[1*] (sjkim@msu.edu), Starla Zemelis-Durfee[1], Jacob Krüger Jensen[1,4], Curtis G. Wilkerson[1,3,4], Kenneth Keegstra[1,2,3] and **Federica Brandizzi**[1,2,4]

[1]Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI; [2]MSU-DOE Plant Research Lab, Michigan State University, East Lansing, MI; [3]Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI; [4]Department of Plant Biology, Michigan State University, East Lansing, MI.

https://www.glbrc.org/

Project Goals: The goal of Bioenergy Plant Design team in Great Lakes Bioenergy Research Center (GLBRC) is to increase the quantity and quality of bioenergy crop biomass per hectare of land, which is crucial for the sustainable and economically viable production of lignocellulosic-derived bioproducts.

The plant cell wall (CW) is composed of polysaccharides and lignin, which have specific roles during plant's growth. Importantly, CW constitutes the majority of the biomass destined for conversion. Therefore, we want to achieve our goal by improving quantitative and qualitative CW traits such as elevating the production and accumulation of cell wall polysaccharides, including Mixed linkage (1,3;1,4)-β-glucan (MLG), a low-recalcitrance glucose polymer. Thus, characterization and engineering MLG synthases are also required to produce MLG with high efficiency.

To achieve the goals, we used *Brachypodium distachyon* as a model grass species. MLG is one of the major components of cereal grains, and occupies up to 80% of cell walls of the Brachypodium endosperm. The MLG biosynthesis depends on the biochemical activity of membrane spanning glucan synthases encoded by the CSLH and CSLF cellulose synthase-like gene families. However, relatively little is known about their topology with respect to the biosynthetic membranes and requirement for producing two different linkages in the MLG glucan backbone. As the first step on the project, we have demonstrated the topology of CSLF6 protein derived from Brachypodium (BdCSLF6) using heterologous expression systems. Using live cell imaging and immuno-electron microscopy analyses of tobacco epidermal cells expressing BdCSLF6, we demonstrate that a functional YFP fusion of BdCSLF6 is localized to the Golgi apparatus and that the Golgi localization of BdCSLF6 is sufficient for MLG biosynthesis. By implementing protease protection assays of BdCSLF6 expressed in the yeast *Pichia pastoris*, we also demonstrated that the catalytic domain, the N-terminus and the C-terminus of the protein are exposed in the cytosol. Furthermore, we found that BdCSLF6 is capable of producing MLG not only in tobacco cells but also in *Pichia*, which generally does not produce MLG.

To further investigate *in-vivo* MLG synthesis in its native environment, we have performed immuno-localization analyses with the MLG-specific antibody in Brachypodium and in barley. We found MLG present in the Golgi, post-Golgi structures and in the cell wall. Accordingly, analyses of a functional fluorescent protein fusion of CSLF6 stably expressed in Brachypodium demonstrated that the enzyme is localized in the Golgi as we have seen in tobacco. We also established that overproduction of MLG causes developmental and growth defects in Brachypodium as also occur in barley. Our results indicate that MLG production by BdCSLF6 occurs in the Golgi similarly to other cell wall matrix polysaccharides, and support the broadly applicable model in grasses that tight mechanisms control optimal MLG accumulation in the cell wall during development and growth.

References

1. Kim, S.-J., Zemelis, S., Keegstra, K. and Brandizzi, F. (2015) The cytoplasmic localization of the catalytic site of CSLF6 supports a channeling model for the biosynthesis of mixed-linkage glucan. *The Plant Journal*, 81, 537-547.
2. Kim, S.-J., Zemelis, S., Jensen, J.K., Wilkerson, C.G.,Keegstra, K. and Brandizzi, F (2018) In the grass species *Brachypodium distachyon*, the production of mixed linkage (1,3;1,4)-β-glucan (MLG) occurs in the Golgi apparatus. *The Plant Journal*.

The Great Lakes Bioenergy Research Center is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Title: Center for Advanced Bioenergy and Bioproducts Innovation - CABBI

Evan H. DeLucia[1]* (delucia@illinois.edu), Matthew Hudson[1], Madhu Khanna[1], Stephen P. Moose[1], Donald R. Ort[1], Christopher V. Rao[1], Vijay Singh[1], Wendy H. Yang[1], Huimin Zhao[1]

[1]University of Illinois, Center for Advanced Bioenergy and Bioproducts Innovation, Urbana, IL 61801 USA

https://cabbi.bio/
http://sustainability.illinois.edu/
https://www.igb.illinois.edu/

Project Goals:

The mission of CABBI is to meet a major challenge facing the world: how to provide sustainable sources of energy for societal needs as the population continues to grow and global change accelerates. CABBI will develop efficient ways to grow, transform, and market biofuels and other bioproducts. The vision of CABBI is to integrate recent advances in genomics, synthetic biology, and computational biology to increase the efficiency, sustainability and value of biomass crops. This holistic approach will help reduce our nation's dependence on fossil fuels thereby increasing sustainability and national security.

Abstract Text:

The Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) will develop efficient ways to grow, transform, and market biofuels, generating new products directly from biomass — thus reducing our nation's dependence on fossil feedstocks, increasing sustainability and decreasing dependence on overseas oil. CABBI will conduct innovative research in biofeedstock development, conversion, and sustainability that integrates recent advances in genomics, synthetic biology, and computational biology to increase the efficiency and value of biomass crops. The Center represents a transformative research model designed to accelerate bioproduct development while retaining the flexibility to assimilate new disruptive technologies.

CABBI is founded on the "plants-as-factories" paradigm (Figure 1), in which biofuels, bioproducts, high-value molecules, and foundation molecules for conversion are synthesized directly in plant stems. This plants-as-factories approach represents value added to efforts of developing efficient lignocellulose deconstruction methods, while retaining residual biomass for deconstruction by traditional or emerging methods. CABBI will focus on sorghum, energycane, and Miscanthus, which are high-yielding throughout the rain-fed eastern U.S., including on marginal soils.

Foundation molecules produced in plants will be efficiently converted to diverse, high-value molecules such as biodiesel, organic acids, jet fuels, lubricants, and alcohols using technologies developed in a versatile and automated biofoundry for rapidly engineering microbial strains. Using the design-build-test-learn framework, CABBI research will overcome the challenges associated with driving biological systems to produce non-naturally occurring compounds.

The Center will employ a data-driven and integrated modeling framework to develop predictive capability on which feedstock combinations, regions and land types, market conditions, and bioproducts have the capacity to support the ecologically and economically sustainable displacement of fossil fuels. Key areas of emphasis will be to obtain a mechanistic understanding of the plant, soil, microbe, and climate interactions that underlie the productivity and delivery of ecosystem services of different feedstocks, and on investigating the technological and economic pathways to a sustainable and resilient bioeconomy. This work will result in an overarching framework for viewing the research through an environmental and economic lens — and for designing a closed-loop and integrated program for CABBI.

The Center for Advanced Bioenergy and Bioproducts Innovation, a U.S. Department of Energy Bioenergy Research Center is supported by DOE, Office of Science, Office of Biological and Environmental Research.

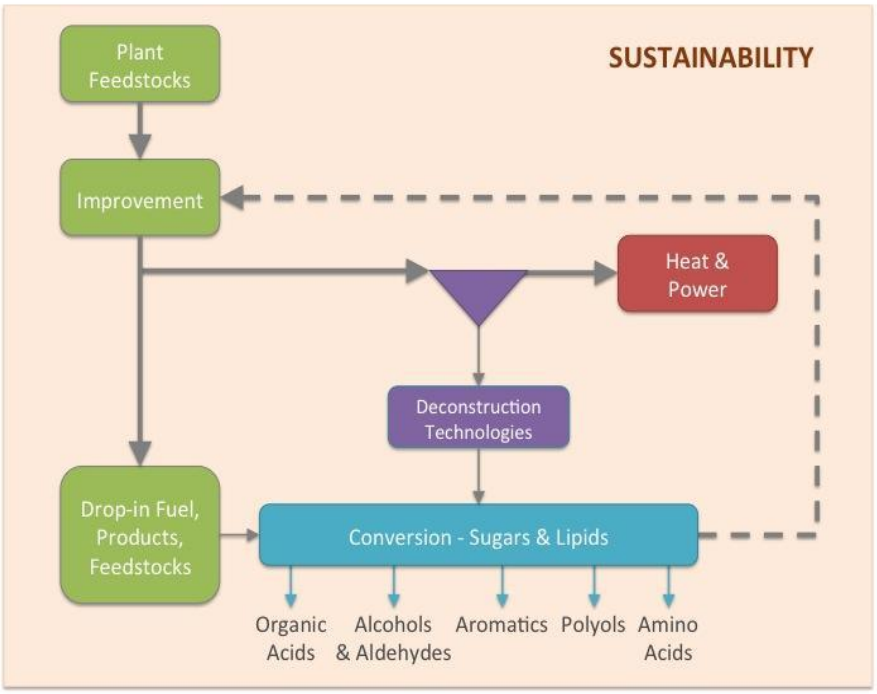


Figure 1. The Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) science model highlights the *plants as factories* concept, in which direct production of drop-in fuels and bioproducts through plant engineering using coupled systems and computational biology approaches employing microorganisms to convert either plant-derived foundation chemicals or sugars and lipids derived from existing deconstruction technologies to produce valued added products. *The CABBI biofoundry* is depicted by solid lines representing the flow of biomass and dashed line representing the application of new knowledge and techniques discovered by the Conversion Theme, to novel engineering approaches to improve plant production and efficiency.

Plant Sphingolipid Glycosylation And Its Role In Immunity And Cell Wall Organization

Lin Fang[1,2], Gosia Murawska[1,2], Toshiki Ishikawa[3], Emilie Rennie[1,2], Isabell Albert[4], Thorsten Nürnberger[4], Henrik Scheller[1,2], **Jenny C. Mortimer[1,2]*** (jcmortimer@lbl.gov)

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Environmental Genomics and Systems Biology, Biosciences, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Saitama University, Saitama, Japan; and [4]University of Tübingen, Tübingen, Germany

http://www.jbei.org

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

Glycosylinositol phosphorylceramides (GIPCs) are a class of glycosylated sphingolipids found in plants, fungi and protozoa. They are extremely abundant in the plant plasma membrane, estimated to form ~25-40 % of total lipids, but almost nothing is known about their function. GIPCs consist of a ceramide attached to a glycan headgroup via a phosphate group. Recently we have identified the first three Arabidopsis proteins involved in the headgroup biosynthesis - IPUT1 (a UDP-glucuronic acid glycosyltransferase), GONST1 (a GDP-mannose transporter) and GMT1 (a GDP-mannose glycosyltransferase. Plants lacking functional copies of these proteins are either pollen lethal (*iput1*) or have extreme developmental defects (*gonst*, *gmt1*), despite the lipid portion of the GIPC being unaffected. This implies a critical function for the GIPC glycan headgroup in membrane function. Here, we identify a new Golgi-localized protein involved in GIPC headgroup biosynthesis in both Arabidopsis and rice - GINT1 (GLUCOSAMINE INOSITOLPHOSPHORYLCERAMIDE TRANSFERASE1), and have described its role. We are now using our collection of GIPC biosynthetic genes as a toolbox with which to explore GIPC function. Two examples will be presented.

Funding statement.

This work was part of the DOE Joint BioEnergy Institute (http:// www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

A Dominant Approach to Reduce Xylan in Bioenergy Crops

Andrew G. Brandon*[1,2] (abrandon@lbl.gov), Devon Birdseye[2], **Henrik V. Scheller**[1,2]

[1]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA ; [2]Feedstocks Divison, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

http://www.jbei.org

Project Goals: Increase the C_6:C_5 sugar ratio and improve the quality of biomass by suppressing xylan biosynthesis in a wild-type background

Xylan is the most abundant non-cellulosic polysaccharide in plant biomass and one of the most abundant biopolymers on earth. The xylan backbone is a homopolymer of β-(1,4)-linked xylose, decorated at regular intervals with GlcA, 4-*O*-MeGlcA and acetyl groups. As a hemicellulose, xylan coats and crosslinks cellulose microfibrils, promoting their crystallinity. Indeed, xylan is critical for the overall health and mechanical strength of the plant. Xylan biosynthesis mutants are severely dwarfed due to cell wall collapse in the water-conducting xylem vessels. While important, the relatively high amount of xylan in plant biomass creates several problems for the development of advanced biofuels. Xylose, a 5-carbon sugar, is poorly utilized by microorganisms and strongly inhibits the fermentation of 6-carbon sugars like glucose. Additionally, the acetate released from the xylan backbone creates a toxic environment for microbial growth. Any way to reduce the amount of xylan in plant biomass will significantly reduce the cost and enhance the efficiency of conversion to biofuel and bioproducts. Since few mutants of important biomass crops exist, the ideal approach would act as a dominant suppressor of xylan biosynthesis. This has been accomplished by identifying potential catalytic residues in the xylan biosynthetic enzyme IRX10 and mutating them. Overexpression of the mutated IRX10 outcompetes the native form of the enzyme, suppressing the biosynthesis of the polymer.

Highly Resolved Genome-Wide Association Mapping (GWAS) Reveals Novel Variants for Engineering Increased Biomass Yield in *Populus trichocarpa*

Wellington Muchero,[1*] (mucherow@ornl.gov), Jin-Gui Chen,[1] Stephen P. DiFazio,[2] Luke Evans,[2] Gancho Slavov,[2] Eli Rodgers-Melnick,[2] Priya Ranjan,[1] Joel Martin,[3] Wendy Schackwitz,[3] Jeremy Schmutz[3], Kathy Haiby, [4] Richard Shuren,[4] Brian Stanton,[4] and **Gerald A. Tuskan**[1]

[1] The Center for Bioenergy Innovation and BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, [2]Department of Biology, West Virginia University, Morgantown, West Virginia, [3]DOE Joint Genome Institute, Walnut Creek, California, [4]GreenWood Resource Inc., Portland, Oregon

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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Undomesticated and outcrossing plant species such as *Populus trichocarpa* have a surprising ability to maintain high-levels of defective alleles among genotypes across their species ranges. As such, high-resolution genome-wide association mapping (GWAS) provides an unparalleled opportunity to resolve loci with large allelic effects on complex traits like biomass yield. Here, we summarize results of a GWAS study leveraging a panel of >8.2 million single nucleotide polymorphisms (SNPs) segregating in population of 1,084 *P. trichocarpa* genotypes established in replicated field sites in California and Oregon. Multi-year measurements of biomass indicators -- height, diameter and sylleptic branching, were collected between 2009 and 2017. Genotype-to-phenotype correlations were performed using mixed liner models with kinship as a covariate. This analysis revealed twelve reproducible associations that were statistically significant after correcting for multiple testing using the Bonferroni approach. Notably, these loci harbored genes including a DELLA family transcription factor previously implicated in gibberellic acid signaling, cell proliferation and internode elongation and expansion, a Cell Division Control protein 48-homolog with established roles in cell cycle regulation, and an XBAT35, a gene with known roles in lateral root production via ethylene biosynthesis. Allelic variants of these genes conferring increased biomass yield are being incorporated in genomic selection algorithms to expedite domestication of high yielding *Populus* genotypes and cis-genic validation of the molecular basis of their function is also underway.

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.

Metabolic Reconstruction, Analysis and Design of Metabolic Pathways

Charles J. Foster[1], Debolina Sarkar[1], Satyakam Dash[1], Saratram Gopalakrishnan[1], Lin Wang[1], Joshua S.H. Chan[1], Costas D. Maranas[1*] (costas@psu.edu) and **Gerald A. Tuskan**[2]

[1]Department of Chemical Engineering, The Pennsylvania State University of State, University Park, PA 16802, [2] Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee

https://cbi.ornl.gov
http://www.maranasgroup.com/

Project Goals: The Center for Bioenergy Innovation (CBI) vision is *to accelerate domestication of bioenergy-relevant, non-model plants and microbes to enable high-impact innovations at multiple points in the bioenergy supply chain.* CBI will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Metabolic models provide a mechanistic description of an organism's metabolic repertoire, by combining reaction stoichiometry information with enzyme availability and metabolite pool levels to form a structured framework that can be analyzed computationally to understand, as well as predict, cellular metabolism. In this study, we reveal that incorporation of kinetic information improved predictive capabilities in a cellulolytic microbe, *C. thermocellum*. We further describe the limitations of existing parametrization procedures for kinetic model development and methods to overcome such limitations using *E. coli* as a model system. We also introduce a recently developed framework (i.e., novoStoic) [1] that seamlessly blends known transformations with reaction rules to construct pathways towards targeted chemicals. This framework is repurposed for prospecting pathways from lignin monomers to value-added chemicals. Finally, we extend our modeling framework from microbes to develop a multi-tissue plant (poplar) model which captures nutrient flow and growth across its tissues.

For *C. thermocellum*, we constructed a core kinetic model "k-ctherm118" [2] to capture the regulatory impact of changes in metabolite pools on reaction fluxes using the Ensemble Modeling paradigm. k-ctherm118 was parameterized by using fermentation yield data in major fermentation pathways for 19 measured metabolites, 19 distinct single and multiple gene knockout mutants along with 18 intracellular metabolite concentration data for a *Δgldh* mutant and ten experimentally measured Michaelis-Menten kinetic parameters. k-ctherm118 captures metabolic perturbations caused by 1) nitrogen limitation leading to increased yields for lactate, pyruvate and amino acids and 2) ethanol stress triggering an increase in intracellular ammonia and sugar phosphate concentrations due to upregulation of cofactor pools. Overall, the *C. thermocellum* case study demonstrates that the developed kinetic model (k-ctherm118) provides greater insight into metabolic pathways and regulations than the stoichiometric model.

We also expand the scope of the kinetic models by incorporating a transcriptional regulatory layer which refines enzyme levels based on a linear combination of log-normalized changes in growth rate (global) and select intracellular metabolite pool (specific) levels. A major computational bottleneck in the kinetic parameterization process is the lack of a fast and efficient algorithm to identify the optimal set of kinetic parameters using MFA-derived steady-state flux distributions. The implementation of a gradient-based

procedure is limited by slow numerical integration. To this end, we have implemented an algorithm [3] to compute steady-state flux distributions for a given set of kinetic parameters that overcome the limitations of numerical integration. In conjunction with an efficient gradient-based scheme for updating kinetic parameters, we have developed a fast and automated algorithm for parametrization of a kinetic model that includes allosteric and transcript regulations. We apply flux datasets from a recent MFA study on *E. coli* (wild-type and 22 knockouts of enzymes in the upper part of central carbon metabolism) and the parametrization procedure to construct a genome-scale kinetic model of *E. coli* containing 787 model reactions, 674 metabolites, and 618 substrate-level regulatory interactions.

The need to identify atom and energy optimized pathways from substrates to terminal products motivates us to go beyond the known repertoire of enzymatic reactions by exploring hypothetical reactions predicted using reaction rules while simultaneously considering all design criteria such as complexity of pathway topology, mass conservation, etc. First, we track and codify all reaction centers as rules, using a novel prime factorization based encoding technique (rePrime). A MILP-based algorithm, novoStoic [1], then allows for the efficient integration of both reaction rules and reactions in the search for pathways that carry out the efficient conversion of the source to target molecules. We demonstrate the use of novoStoic in bypassing existing pathway steps through putative transformations, assembling complex pathways, and blending both known and putative steps from lignin monomers to value-added chemicals.

We also extend the metabolic modeling framework beyond microbes by developing whole-plant metabolic model of poplar (*P. trichocarpa*) to identify key genes responsible for controlling growth, yield, degradability, and biomass composition. Thus, a genome-scale model for poplar was constructed containing over 3,459 metabolites, 3,388 reactions, and 8 sub-cellular compartments. The GSM model acts as blueprint to explore organ-specific compartments and reaction networks spanning the root, shoot, and leaf tissues using omics data and ultimately capture carbon and nitrogen flows between tissues as a function of growth stage and sequester biomass in organ-specific ratios.

References

1. Kumar A., Wang L., Ng C.Y., and Maranas C.D., (2017) Pathway design using de novo steps through uncharted biochemical spaces. *Nature Communications*, in press.
2. Dash S., Khodayari A., Zhou J., Holwerda E.K., Olson D.G., Lynd L.R., Maranas C.D., (2017) Development of a core *Clostridium thermocellum* kinetic metabolic model consistent with multiple genetic perturbations. *Biotechnol Biofuels*, 10:108.
3. Gopalakrishnan S., Foster C.J., Dash S., and Maranas C.D., (2017) in preparation.

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

Poplar GWAS and eQTL Analyses Reveal a Key Transferase Gene, HCT2, as a Missing Link in Plant Growth-Defense Tradeoffs

Jay Chen,[1*](chenj@ornl.gov), Wellington Muchero,[1] Jin Zhang,[1] Yongil Yang,[1] Kaijie Zheng,[1] Meng Xie,[1] Kai Feng,[1] Sara S. Jawdy,[1] Lee E. Gunter,[1] Priya Ranjan,[1] Vasanth R. Singan,[2] Nancy Engle,[1] Erika Lindquist,[2] Kerrie Barry,[2] Jeremy Schmutz,[2,3] Nan Zhao,[4] Timothy J. Tschaplinski,[1] Jared LeBoldus,[5] and **Gerald A. Tuskan[1]**

[1]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]U.S. Department of Energy Joint Genome Institute, Walnut Creek, California; [3]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama; [4]Institute of Agriculture, University of Tennessee, Knoxville, Tennessee; [5]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon

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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Poplar is a fast-growing woody perennial with robust genetic tools and data. However, challenges remain in gaining functional information on its many genes and pathways. Secondary metabolite biosynthesis is a complex and precise process that is catalyzed by numerous enzymes that are under the control of complex transcriptional regulatory networks. The identification of key regulators in secondary metabolite biosynthesis remains restricted by low throughput techniques. We integrated genome-wide associated studies (GWAS) and expression-based quantitative trait loci (eQTL) studies in *Populus trichocarpa* to identify genetic elements controlling the abundance of *cis*- and *trans*-3-*O*-caffeoylquinic acid, key intermediates in lignin biosynthesis and important compounds with numerous therapeutic roles including antioxidant and antimicrobial activity. We found that the abundances of these metabolites were not only significantly associated with single nucleotide polymorphisms (SNPs) in a hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl transferase gene (*PtHCT2*), but were also correlated with the expression levels of the same gene. eQTL mapping revealed that *PtHCT2* expression was regulated by putative cis-acting elements, which coincided with GWAS SNP associations and were located in a W-box element, a binding site for WRKY transcription factors. Further analyses in co-expression networks, transcriptional response to infection by the fungal pathogen *Sphaerulina musiva*, and *in vitro* validation of transcriptional regulation suggested that *PtHCT2*

is involved in both caffeoylquinic acid biosynthesis and defense response, providing one example of the long sought after a mechanistic link in growth-defense tradeoffs.

The Center for Bioenergy Innovation, a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Switchgrass Field Experiments to Enable Rapid Domestication, Increased Sustainability, and a Genome-wide Association Study

Reginald J. Millwood,[1,2*] (rmillwood@utk.edu), Mitra Mazarei,[1,2] C. Neal Stewart Jr.,[1,2] and **Gerald A. Tuskan**[2]

[1]University of Tennessee Institute of Agriculture, Knoxville; The 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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Biomass production is the single most important factor impacting economic sustainability of biofuels and bioproducts. A diverse panel of 500 switchgrass accessions is planned to be planted in Knoxville in spring, 2018. This panel will consist largely of clonal replicates of lowland (tetraploid) switchgrass accessions provided by Dr. Thomas Juenger. The experimental common garden study in Knoxville, Tennessee will consist of six replicates of each switchgrass accession, using a randomized complete block experimental design, which includes a nitrogen fertility treatment.

For total biomass, the aerial parts of each genotype will be harvested in late fall, dried, and weighed. Tissue from selected lines will be used to assess lignin valorization potential. To dissect yield into its genetic components, we will record flowering time (50% of the first panicle emerged), plant height at flowering and at maturity, total biomass, leaf-to-stem ratio, lodging, and timing of spring regrowth using a combination of manual and automated phenotyping. Automated plant architecture measures will be acquired through the use of an unmanned aerial vehicle (UAV) using LIDAR-based methods.

The end-of-season shoot-to-root nitrogen will be measured to determine the nitrogen remobilization, which is a critical trait for long-term NUE. The nitrogen status of plants will be determined via elemental analysis at two growth stages: mid-season when plants are mature but still green and end-of-season when plants are senesced. We will employ UAV-based multispectral sensors to estimate the level of nitrogen in leaves and stems. Five spectral bands, including near-infrared, and various predictive algorithms will be used to estimate nitrogen

presence and use. These predictive algorithms will be “ground-proofed” against lab-based plant nitrogen analyses. Nitrogen uptake, utilization, and use efficiencies of each genotype will be determined.

We will also evaluate disease severity with focus on rust caused by fungal *Puccinia emaculata*, which is the most prevalent pathogen of switchgrass in the southern US. The plants will be examined for rust severity at different time points. The coverage of leaf surface with rust uredia will be visually assessed using the following scale: 0=0%, 1≤5%, 2≤10%, 3≤25%, 4≤40%, 5≤55%, 6≤70%, and 7≤100% of leaf area coverage with uredia. Ultimately, these data should also be collected by remote sensing using the UAV. All data will be incorporated in the GWAS analysis to determine the loci associated with rust resistance, which can be integrated into the genomic selection models.

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.

Increasing biomass in grasses by identifying genes involved in mixed-linkage glucan biosynthesis

Mingzhu Fan (mfan@msu.edu), [1,2] Jacob Krüger Jensen, [1,2] Starla Zemelis-Durfee, [2,4] Federica Brandizzi, [2,4] and **Curtis Gene Wilkerson**[1, 2, 3]

[1] Department of Plant Biology, Michigan State University, East Lansing, MI 48824;
[2] Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48824;
[3] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824;
[4] MSU-DOE Plant Research Lab, Michigan State University, East Lansing, MI 48824

Project Goals: The major goal of this project is to increase the quality and quantity of grass biomass. Specifically we aim to increase the amount of the easily digestible cell wall polysaccharide mixed-linkage glucan in grass cell walls. To achieve this goal, we have identified a transcription factor likely involved in mixed-linkage glucan biosynthesis that should allow us to better understand mixed-linkage glucan biosynthesis.

Mixed-linkage glucan (MLG) is a polysaccharide that is highly abundant in grass endosperm cell walls and at lower amounts in other tissues. We know that the enzymes produced by the genes *CSLF* and *CSLH*, members of the cellulose synthase-like gene families, are able to synthesize MLG but it is unknown if other genes participate in the production of MLG. We generated a large set of Brachypodium transcriptional profiling data that allowed us to identified a trihelix family transcription factor (*BdTHX1*) that is highly co-expressed with *BdCSLF6*. Using published data from other groups we were able to show that *THX1* and *CSLF6* are also co-expressed in wheat and maize [1, 2]. These co-expression data suggests that *THX1* is involved in the regulation of MLG biosynthesis.

Determining the genes regulated by this transcription factor could reveal more genes involved in MLG accumulation. To find such genes we conducted chromatin immunoprecipitation (ChIP)-seq experiments using immature Brachypodium seeds and an anti-BdTHX1 polyclonal antibody we produced. The ChIP-seq experiment identified the second intron of *BdCSLF6* as one of the most enriched sequences. The binding of BdTHX1 to the *BdCSLF6* intron was confirmed using electrophoretic mobility shift assays (EMSA). A gene encoding a grass specific glycoside hydrolase family 16 (GGH16-1) endotransglucosylase/hydrolase was also discovered in the ChIP-seq data and the binding was confirmed by EMSA. Such enzymes have been implicated in the incorporation of hemicelluloses in to the wall and as such could be involved in MLG accumulation. We have expressed GGH16-1 in *Pichia* and we are in the process of testing its activity on MLG. Motif analysis of the DNA regions shown to bind BdTHX1 by ChIP-seq and EMSA revealed that BdTHX1 binds to previously described elements bound by trihelix transcription factors termed GT-elements. GT-elements were also found in the introns of *CSLF6* gene in rice and maize.

Our work provides information on genes likely involved in MLG accumulation and

further characterization of these genes should allow for the improvement of biomass for biofuel production.

References

[1] Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, Mayer KFX, Olsen O (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345: 1–8

[2] Andorf CM, Cannon EK, Portwood JL, Gardiner JM, Harper LC, Schaeffer ML, Braun BL, Campbell DA, Vinnakota AG, Sribalusu V V., et al (2016) MaizeGDB update: New tools, data and interface for the maize model organism database. Nucleic Acids Res 44: D1195–D1201

This work was supported by the U.S. Department of Energy Great Lakes Bioenergy Research Center (grant no. BER DE-FC02-07ER64494 to C.G.W)

PHYSIOLOGICAL AND MOLECULAR-GENETIC CHARACTERIZATION OF BASAL RESISTANCE IN SORGHUM

Jennifer Kimball[1] , Thalita Tuleski[2,3] , Xinye Zhang[1], Yaya Cui[2], Catherine Espinoza[2] , Dongqin Chen[2], Rozalynne Samira[1], Gary Stacey[2] and **Peter Balint-Kurti***[1,4]

[1]North Carolina State University, Raleigh NC,; [2]University of Missouri, Columbia, MO [3]Federal University of Paraná, UFPR, Brazil; [4]US Dept. of Agriculture, Agricultural Research Service, Raleigh, NC, *Presenting author

https://sites.google.com/a/ncsu.edu/maize-disease/home

PROJECT GOALS

The objectives of the project are the following

- I. *Develop robust assays to measure the microbe-associated molecular pattern (MAMP) response and disease resistance in sorghum.*
- II. *Screen a set of diverse sorghum germplasm for variation in the MAMP response and disease resistance.*
- III. *Identify genes differentially regulated during the MAMP response in high- and low-responding sorghum genotypes.*
- IV. *Assess the effect of MTI on disease progression in sorghum*
- V. *Identify loci associated with variation in disease resistance and the MAMP response. Examine possible correlations between variation in the MAMP response and in disease resistance*

PROJECT SUMMARY

Plants recognize certain conserved microbial molecules (microbe-associated molecular patterns or MAMPs) and mount a basal defense response called MAMP-triggered immunity (MTI) that limits subsequent colonization. In many cases, the basal defense response is believed to be responsible for non-host resistance: the phenomenon whereby most plants are resistant to most microbial pathogens. Furthermore, there is some evidence that the MAMP response may be involved with quantitative disease resistance, resistance which although partial, tends to be extremely durable. While much is known about the MAMP response in model species, this is not the case for crop plants. Furthermore, naturally-occurring variation in the MAMP response within a species and its relationship to quantitative disease resistance is not well understood. This project builds on our work investigating the genetics controlling the Arabidopsis and soybean MAMP response and on characterizing maize quantitative disease resistance.

PROGRESS

Objective I. We have developed robust reactive oxygen species (ROS)-based assays for response to chitin and flagellin (flg22). We have also worked on assays measuring nitric oxide (NO) production in response to these PAMPs though this assay produces a very high background level in non-treated lines.

Objective II. We have measured the responses to chitin and flagellin in more than 500 diverse sorghum lines. We have identified several high responders to both MAMPs.

In roots we have identified an interesting biphasic response with peaks at about 5 minutes and 25 minutes after elicitation. This biphasic response segregates in the population. Our data suggests that the strength of the MAMP response in roots is not correlated with that in leaves.

We have examined the PAMP response over developmental time and have shown that it varies significantly with increased responses occurring about 45 days after planting

Objective III. We have isolated RNA from leaves and roots of 2 lines, Btx623 and SC155-14E, treated with flg22, chitin and water (as a control). We are currently performing RNAseq analysis on these samples. Results will be presented.

Objective IV. We have developed greenhouse disease assays for three fungal diseases (Target leaf spot, northern leaf blight, anthracnose leaf blight) and one bacterial pathogen (*Herbaspirillum rubrisubalbicans*, causal agent of mottled stripe disease). We have shown that elicitation of the PAMP response with either flg22 or chitin can confer resistnace to Herbaspirillum rubrisubalbicans. We have attempted to perform similar experiments with the fungal diseases but have not yet been able to obtain satisfactory data.

Objective V. We have assessed an association mapping population of 500 diverse lines for response to flg22 in two replications and have identified several associated SNPs and candidate genes. Several candidate genes are predicted to be involved in oxidative stress response, heavy metal detoxification and defense response. We have mapped resistance to sorghum target leaf spot in this population in one environment with two replications. We identified several QTL including one QTL that had been previously identified in other studies.

We have assessed the PAMP response and disease resistance of two recombinant inbred population (RIL) populations (BTx623 x SC155, 107 lines and BTx623 x BTx642,130 lines). We assessed them in replicated field trials in 2016 and 2017 for resistance to the pathogen *Bipolaris sorghicola* , the causal agent of target leaf spot and have identified several QTL. We also assessed the Btx623 x SC155 population for response to flg22 and resistance to *H. rubrisubalbicans* and we are currently performing similar assays for the BTx623 x BTx642 population. We have not identified any QTL shared between the disease resistance and PAMP response traits.

Under separate funding we have undertaken a similar project in maize. We anticipate that the results of our sorghum and maize projects will mutually inform the other. We will present the latest data from both these projects in our poster at the PI meeting

Funding statement.

This work is funded by DOE grant #DE-SC0014116 “Physiological and Molecular-Genetic Characterization of Basal Resistance in Sorghum”

Leaf carbon isotope composition in *Setaria*: genetic contribution and potential for high throughput screening for water use efficiency in C_4 plants

Asaph B. Cousins[1]* (acousins@wsu.edu), Patrick Z. Ellsworth[1], Max Feldman[2], Patricia V. Ellsworth[1], Ivan Baxter[2], and **Tom Brutnell**[2]

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

url: http://foxmillet.org

Project Goals: Objectives of this project are to 1) identify the gene regulatory and metabolic networks important for adaptation to low water availability and high-density plantings and to understand the physiology underlying these adaptations and 2) develop technologies to precisely control gene insertion and replacements events for large scale engineering of pathways in model and target feedstocks.

Abstract: Increasing whole plant water use efficiency (yield per transpiration; WUE_{plant}) through plant breeding can benefit the sustainability of agriculture and improve yield under drought. To select for WUE_{plant} a high throughput method of phenotyping must be developed, and the genetic architecture of traits such as transpiration efficiency (TE_i; rate of CO_2 assimilation relative to stomatal conductance) must be better understood. Leaf carbon stable isotope composition ($\delta^{13}C_{leaf}$) has been proposed as a high throughput proxy for TE_i, and there is a negative correlation between $\delta^{13}C_{leaf}$ and both WUE_{plant} and TE_i in the C_4 model grass species *Setaria*. Therefore, a water limitation experiment was conducted where transpiration, biomass, WUE_{plant}, and $\delta^{13}C_{leaf}$ were measured on a recombinant inbred line (RIL) population of *Setaria viridis* and *S. italica* to better define the genomic control of WUE_{plant} and TE_i. Three quantitative trait loci (QTL) for $\delta^{13}C_{leaf}$ were co-localized with transpiration and biomass, but not with WUE_{plant}. However, $\delta^{13}C_{leaf}$ was negatively correlated with WUE_{plant} when WUE_{plant} was calculated for allele classes based on the allele combinations of the three QTL for $\delta^{13}C_{leaf}$. This negative relationship suggests that variation in WUE_{plant} across these allele classes is in part due to differences in TE_i. In this C_4 grass population, multiple traits can influence WUE_{plant}; however, the analysis of $\delta^{13}C_{leaf}$ provides insights into how TE_i contributes to WUE_{plant}. The data presented here suggests that $\delta^{13}C_{leaf}$ can be used in marker-assisted breeding to select for TE_i and to better understand the genetic architecture of TE_i and WUE_{plant} in C_4 species.

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], Nathan Miller[2], Jiayang Xie[1], Charles Pignon[1], Gorka Erice[1], Timothy Wertin[1], Nicole Choquette[1], Maximilian Feldman[3], Funda Ogut[4], Parthiban Prakash[1], Peter Schmuker[1], Anna Dmitrievna[1], Dylan Allen[1], Elizabeth A. Ainsworth[5,1], Ivan Baxter[3], **Thomas P. Brutnell[3]**, Edward S. Buckler[6,7], Patrick J. Brown[1], Thomas E. Clemente[8], Stephen P. Long[1], Lauren M. McIntyre[4], Edgar Spalding[2]

[1]University of Illinois at Urbana-Champaign, Urbana, IL; [2]University of Wisconsin, Madison, WI; [3]Donald Danforth Plant Science Center, St Louis, MO; [4]University of Florida, Gainsville, FL; [5] USDA Agricultural Research Service, Urbana, IL; [6]USDA Agricultural Research Service, Ithica, NY; [7]Cornell University, Ithica, NY; [8]University of Nebraska, Lincoln, NE

http://foxmillet.org/

Genetically tractable model systems closely related to bioenergy grasses need to be developed to drive the crop improvement required for large scale, ecologically sustainable bioenergy production. *Setaria viridis* is an ideal candidate C_4 panacoid grass. The over-arching objectives of this large, collaborative project are to utilize genomic, computational and engineering tools to begin the genetic dissection of drought response in *S. viridis*. This will be achieved through: 1) Quantitative trait and association genetics; 2) novel controlled environment and field phenotyping combined with molecular and chemical profiling; 3) development of metabolic and gene networks; 4) development of transformation technologies; 5) reverse genetic testing of candidate genes.

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 and diminishing freshwater supply. The exchange of water and CO_2 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 such maize, sorghum and sugar cane. 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. To this end two novel phenotyping platforms were developed. First, a rapid method to assess stomatal patterning in three model C_4 species grown in the field – maize, sorghum and setaria has been implemented. Here the leaf surface is scanned in less than two minutes with a modified confocal microscope, 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 a training of a pattern-recognition neural network approach. Second, a thermal imaging capture strategy, to rapidly screen the kinetics of stomatal closure in response to light has been developed. We are gaining insight on the underlying genetics governing stomatal stomatal patterning through quantitative trait loci and genome wide association studies in addition to phenotypic evaluations of sorghum with transgenically

modified expression of stomatal patterning genes. These multifaceted approaches are complemented by a recently established field facility for comprehensive evaluation of leaf, root and canopy WUE traits under Midwest growing conditions.

This research was funded through Subaward No. 23009-UI, CFDA # 81.049 between University of Illinois and Donald Danforth Plant Science Center Under Prime Agreement No. DE-SC0008769 from Department of Energy.

Using systems approaches to improve photosynthesis and water use efficiency in sorghum

Thomas Brutnell*[1] (tbrutnell@danforthcenter.org), Ivan Baxter[2], Asaph Cousins[3], Jose Dinneny[4], Albert Kausch[5], Andrew Leakey[6], Todd Mockler[1], Hector Quemada[1], Sue Rhee[7], & Daniel Voytas[8].

[1]Donald Danforth Plant Science Center, St. Louis, MO; [2]USDA/Donald Danforth Plant Science Center, St. Louis, MO; [3]Washington State University, Pullman, WA; [4]Stanford University, Stanford, CA; [5]University of Rhode Island, West Kingston, RI; [6]University of Illinois, Urbana-Champaign, IL; [7]Carnegie Institution for Science, Stanford, CA; [8]University of Minnesota, St. Paul, MN

url: n/a

Project Goals:

1. Engineer photosynthesis for improved performance under water stress.
2. Optimize water relations to enhance drought tolerance and water use efficiency.
3. Develop a comparative GWAS pipeline for sorghum and Setaria.
4. Use metabolic network modeling to guide biomass engineering.
5. Manipulate plant gene expression through precision engineering.
6. Develop methods to improve transformation efficiencies in sorghum and establish a regulatory framework for deployment of engineered organisms.

Abstract:
The success of a bioenergy economy will depend on the development of second generation biofuel crops that can be grown in suboptimal conditions (e.g., hot, dry, marginal soil) and thus not compete with the food or feed markets. Sorghum is a naturally drought-tolerant, heat-resistant, low-cost crop, currently grown globally in Asia, Africa, Australia, and the Americas. To maximize the potential of sorghum as a biofuel crop, a new systems-oriented blueprint and re-design of sorghum is needed. Using both sorghum and the model grass green foxtail (*Setaria viridis*), this project will build a synthetic biology toolkit to accelerate development of elite energy sorghum varieties for production under marginal environments. This will involve the convergence of precision bioengineering technologies, genome-scale modeling, genome sequencing, and high-throughput analyses in laboratory and field settings. Using those computational and engineering tools, this research will enhance photosynthesis in sorghum to increase plant growth and biomass accumulation, while minimizing irrigation requirements by improving water acquisition through a redesigned root system and by reducing plant transpiration. With this multidisciplinary approach, this project will deliver stress-tolerant sorghum lines, addressing DOE's mission in the generation of renewable energy resources.

Funding statement:
This work is funded by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) award #DE-SC0018277

Using systems approaches to improve water use efficiency in sorghum by engineering root architecture

Thomas Brutnell[1], Ivan Baxter[2], Asaph Cousins[3], **Jose Dinneny[4]** (dinneny@stanford.edu), Albert Kausch[5], Andrew Leakey[6], Todd Mockler[1], Hector Quemada[1], Sue Rhee[7], & Daniel Voytas[8].

[1]Donald Danforth Plant Science Center, St. Louis, MO; [2]USDA/Donald Danforth Plant Science Center, St. Louis, MO; [3]Washington State University, Pullman, WA; [4]Stanford University, Stanford, CA; [5]University of Rhode Island, West Kingston, RI; [6]University of Illinois, Urbana-Champaign, IL; [7]Carnegie Institution for Science, Stanford, CA; [8]University of Minnesota, St. Paul, MN

url: n/a

Project Goals:

1. Engineer photosynthesis for improved performance under water stress.
2. Optimize water relations to enhance drought tolerance and water use efficiency.
3. Develop a comparative GWAS pipeline for sorghum and Setaria.
4. Use metabolic network modeling to guide biomass engineering.
5. Manipulate plant gene expression through precision engineering.
6. Develop methods to improve transformation efficiencies in sorghum and establish a regulatory framework for deployment of engineered organisms.

Abstract:
Grass species develop the majority of their root system through the adventious initiation of crown roots from the base of the shoot. Our recent work has shown that crown roots are particularly sensitive to the local availability of water at the soil surface in a response that is widely conserved across the Poaceae. Our data highlight the importance of crown root suppression in reducing the flux of water through the plant, which preserves soil water in a strategy known as water banking. While water banking may be a useful strategy for plants in natural environments, this strategy may have other negative consequences in an agricultural context including smaller root system size, reduced nutrient uptake and reduced carbon assimilation. For example, while Setaria viridis exhibits a severe reduction in crown root growth under drought, maize and sorghum inbreds show wide variation in this response with some inbreds able to maintain crown root growth under drought. Such variation suggests that breeding may have inadvertently "tuned" crown root responsiveness to drought to suit particular soil conditions and precipitation patterns. Understanding the relationship between environmental inputs (water, light, temperature) and crown-root response strategies will allow ideotypes appropriate to particular genotype-environment-management interactions. By modulating the growth of crown roots, we will change the rate at which water is taken up 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.

Funding statement:
This work is funded by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) award #DE-SC0018277

Discovery and characterization of disease resistance loci in poplar

Héloïse Bastiaanse,[1] Luca Comai,[2,3] Andrew Groover,[1,2*] Isabelle Henry,[3] George Newcombe,[4] Randall Rousseau[5]

[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

Project Goals

1. Identification of genomic loci that control response to two damaging poplar diseases, leaf spot and leaf rust. In field trials and controlled infection experiments, genome-wide scans will identify poplar loci underlying qualitative (allelic) and quantitative (indels) parental loci affecting disease susceptibility.
2. Characterization of transcriptional networks governing the response of poplar to leaf pathogens. We will characterize major QTLs through integration of transcriptome data with developmental responses. We will develop quantitative models of co-expression networks of poplar genes, which will enable hypotheses testing regarding molecular mechanisms.
3. Identification of genes whose manipulation could result in sustainable, long-term tolerance to the target pathogens. Candidate genes will be evaluated based on gene expression, dosage sensitivity, functional annotations, and correlations with disease-related processes in objective 2. Individual candidate genes will be functionally characterized by transgenic validation and/or genome editing.

Abstract

Pathogenic fungi that colonize poplar leaves and stems reduce yield and can cause failure of industrial bioenergy plantations. Despite extensive study of poplar pathosystems, no durable resistance genes have been characterized from poplar, underscoring that new approaches are needed. With funding from the USDA/DOE Plant Feedstocks Genomics for Bioenergy program, we have developed an efficient functional genomics resource based on gene dosage changes in a *Populus nigra* x *P. deltoides* elite biomass hybrid. In this ~800 clone population, gene dosage variation is created by large-scale deletions and insertions that tile each chromosome. In clonally replicated field trials, we have successfully mapped dosage QTLs for key bioenergy traits, and recently have identified preliminary QTL influencing susceptibility of poplar to pathogens, demonstrating the utility of this resource for identifying loci affecting disease responses. The goal of this proposal is to exploit this unique resource to enable genomic-guided strategies for achieving durable resistance.

The approach is to first perform genome-wide scans using the previously developed poplar functional genomics resource to identify dosage-sensitive loci associated with disease susceptibility. Using field trials in California and Missouri, we will survey the most important diseases of poplar: leaf rust, leaf spot and stem canker. Together, these diseases not only cause extensive reductions in yields, but also severely limit the geographic range of key, high-yielding poplar interspecific hybrids. Greenhouse-based controlled inoculation screens will be performed for leaf rust isolates using the same genotypes. In addition to identification of QTL, correlations will be determined among disease QTL and previously

identified bioenergy QTL and traits for the same genotypes. 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 use the information from the previous experiments to identify candidate genes for functional analysis by manipulation using CRISPR-Cas9. Such dosage-sensitive candidate genes with significant effects on disease susceptibility phenotypes could then be manipulated in breeding programs through selection of germplasm with naturally-occurring allelic variation or indels/copy number variation covering susceptibility or resistance loci.

This material is based upon work that is supported by the 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.

Genomic dissection of anthracnose resistant response in sorghum [*Sorghum bicolor* (L.) Moench]

Hugo E. Cuevas [1, *](hugo.cuevas@ars.usda.gov), Clara M. Cruet[2], Louis K. Prom[3], Joseph E. Knoll[4], Lauren Stutts[5], Wilfred Vermerris[6]

[1]USDA-ARS Tropical Agriculture Research Station, Mayaguez, PR; [2] University of Puerto Rico; Mayaguez Campus, Department of Biology. Mayaguez, P.R. 00680; [3]USDA-ARS Crop Germplasm Research Unit, College Station, TX; [4]USDA-ARS Crop Genetics and Breeding Research Unit, Tifton, GA; [5]University of Florida, Program in Plant Molecular & Cellular Biology and [6]Department of Microbiology & Cell Science, Gainesville, FL

Project Goals:

The goal of this project is to use a genomics-based approach to identify anthracnose resistance loci from diverse sorghum germplasm, to establish against which pathotypes these loci protect, and to determine the disease resistance mechanism of at least one of these genes. This information will provide plant breeders a tool kit that can be used to maximize levels of resistance in different areas of production.

The productivity and profitability of sorghum [*Sorghum bicolor* (L.) Moench] is reduced by susceptibility to fungal diseases, such as anthracnose, caused by *Colletotrichum sublineolum*. The identification of anthracnose resistance loci from different sorghum accessions is imperative to develop new varieties with broader resistant response and to increase its durability. We evaluated the anthracnose resistant response of 335 accessions from a sorghum association panel (SAP) (Cuevas *et al.* 2018, Plant Genome), 300 exotic sorghum accessions from NPGS Ethiopian germplasm collection, and two sets of recombinant inbred lines derived from anthracnose-resistant sources SC112-14 (Ethiopia) and QL3 (India). In parallel, two novel QTL associated with the resistant response in sorghum line Bk7 (chromosomes 7 and 9; Felderhoff *et al*. 2016,G3) are now being analyzed in more detail to identify the underlying resistance genes.

The evaluation of SAP identified 75 accessions resistant to anthracnose (Cuevas et al., 2018, Plant Genome). A phylogenetic analysis of these accessions showed high genetic diversity and multiple resistance sources. Genome-wide association scans (GWAS) using 268,289 single-nucleotide polymorphisms (SNPs; Morris et al. 2013) and logistic regressions for binary measures of resistance responses identified three loci within a region on chromosome 5 that have been previously associated with three sources of anthracnose resistance. Candidate genes within these loci were related to R-gene families, signaling cascades and transcriptional reprogramming, suggesting that the resistance response is controlled by multiple defense mechanisms. The strategic integration of exotic resistant germplasm into the SAP is needed to identify additional rare resistance alleles via GWAS.

The evaluation of 297 exotic sorghum accessions from NPGS Ethiopian germplasm collection identified 143 resistant accessions. Genetic characterization of this germplasm (Cuevas et al. 2017; BMC Genomics) and its anthracnose resistant response were merged with phenotypic and genetic characterization of the SAP (Morris et al. 2013) for a large GWAS comprising 592 accessions and 219,037 SNPs. Logistic regressions for binary measures of resistance responses identified the previously associated locus on chromosome 5 and an additional locus on chromosome 3, while mixed linear model using quantitative resistant response identified a locus

on chromosome 9. Candidate genes within loci on chromosome 5 and 3 include a resistance gene belonging to a family of genes encoding *F-box* proteins, while a candidate resistance gene on chromosome 9 is a gene with leucine-rich repeat and NACHT domain (i.e. R-gene family). Resistant alleles for loci on chromosomes 3 and 9 are present in the SAP at low frequency, thus, the integration of NPGS Ethiopian germplasm increased its frequency and power of detection.

Two sets of recombinant inbred lines (RILs) derived from the anthracnose-resistant sources SC112-14 (Ethiopia) and QL3 (India) were evaluated for anthracnose resistant response in Puerto Rico, Florida, Georgia and Texas for two consecutive years to identify broad resistance against multiple isolates of the pathogen. In parallel, a subset of RILs were evaluated in the greenhouse against eight anthracnose pathotypes to identify particular resistance loci. Composite interval mapping using two high-density genetic maps constructed using genotyping by sequencing revealed that the resistance in QL3 is controlled by multiple loci, while the resistance in SC112-14 is controlled by a single locus on chromosome 5. Segregation analysis of 1,500 individuals derived from SC112-14 delimited this locus to a 22 kb region and confirmed one of the GWAS loci on chromosome 5. Greenhouse evaluation validated this locus for the eight pathotypes. In contrast, greenhouse evaluation for QL3 identified two loci for four pathotypes that were not detected based on field evaluations.

Visual observation of the resistant line Bk7 and the susceptible line Early Hegari Sart in response to infection with *C. sublineolum* shows that rates of germination and appressorium formation were reduced in Bk7, and that this line also produces more dexoyanthocyanidins, visible as orange pigments. Mycelial growth is being monitored with a transgenic strain of *C. sublineolum* expressing the green fluorescent protein. We also determined that the resistant locus on chromosome 9 present in Bk7 included 12 candidate genes (Felderhoff et al., 2017) that were validated to be expressed in leaves. Each of the12 candidate genes from Bk7 is now being targeted for down-regulation using virus-induced gene silencing (VIGS) with a modified brome mosaic virus. Down-regulation of the gene(s) responsible for the anthracnose resistance is expected to result in successful infection by the pathogen. The locus on chromosome 7 contains too many candidate genes for VIGS, and expression profiling is being pursued to identify differentially expressed genes as candidates.

References

Cuevas et al. (2018). Genome-wide association mapping of anthracnose (*Colletotrichum sublineolum*) resistance in the U.S. sorghum association panel. *Plant Genome* (accepted)

Cuevas et al. (2017). Genomic characterization of a core set of the USDA-NPGS Ethiopian sorghum germplasm collection: implications for germplasm conservation, evaluation, and utilization in crop improvement. *BMC Genomics* **18**, 108.

Felderhoff et al. (2016). Using genotyping by sequencing to map two novel anthracnose resistance loci in *Sorghum bicolor*. *G3-Genes Genomes Genetics* **6**, 1935-1946.

Felderhoff et al. (2017) A cost-benefit analysis to select the most effective method for positional cloning: genotyping by sequencing versus allele-specific PCR. Euphytica 213: 286

Morris et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 453-458.

Funding statement.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0014171.

Genomics and Phenomics to Identify Yield and Drought Tolerance Alleles for Improvement of Camelina as a Biofuel Crop

Hussein Abdel-Haleem[1], Matthew Conley[1], Yufeng Ge[2], Daniel Schachtman[2], Cody Creech[2], Noah Fahlgren[3], Toni Kutchan[3], Sheeja George[4], and Russ Gesch[5], and **John M. Dyer**[1,*] (john.dyer@ars.usda.gov)

[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 will use 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 panel of 250 camelina accessions grown under well-watered and water-limited conditions. In year 1 of the study, the entire panel was analyzed using the greenhouse-based LemnaTec system available at the Donald Danforth Plant Science Center. The greenhouse-based system was used to monitor the growth of two replicates of each accession under well-watered and water-limited conditions from 7 to 35 days after planting. Phenotype data for each plant at each time point was derived from the approximately 200,000 images produced during the experiment using the PlantCV software package. Seeds were collected from each plant for further analysis.

In preparation for the field-based phenotyping experiments to be conducted in Maricopa, AZ and Sidney, NE, the entire population was planted in Maricopa during year 1 for multiplication of seed. Pilot studies were also conducted at each location with 10 selected varieties that showed genotypic and phenotypic diversity. In Nebraska, the ten selected varieties were planted in 5 replicates. The control was irrigated with approximately 1" per week and the treatment was not irrigated after stand establishment. The decreases in yield due to the treatment ranged from 26% to 1%, demonstrating good treatment effects and genotypic differences in the camelina lines

tested. A new field phenotyping cart for camelina plots with multiple sensing modules was constructed. Phenotyping was conducted at two time points. Initial analysis of the canopy temperature data showed distinct differences in the canopy temperatures between treatments towards the end of the growing season. In-depth analysis of 2017 phenomics/yield data for the ten selected varieties is ongoing. These preliminary trials helped us to establish the methods that will be used in 2018 and 2019 field trials to screen the 250 accessions of camelina for tolerance to drought based on yield reductions and phenomics data.

In Maricopa, the ten camelina varieties were planted in 4-row plots, three replicates each, under well-irrigated and reduced irrigation conditions. A high-clearance tractor was equipped with proximal sensors and imagers to collect field-based HTP data including crop height, canopy temperature and canopy multi-spectral reflectance. The HTP data were collected on a weekly basis throughout the growing season. The reflectance data were used to construct vegetation indexes. In addition to field-based HTP measurements, traditional morphological and physiological traits were collected and analyzed. Raw data were processed to remove outliers then the cleaned data sets were analyzed using analysis of variances and correlation analyses. The results showed that there were temporal associations between vegetation indexes and camelina growth stages. The indexes reached the highest values around flowering time then decreased in value at later growth stages when chlorophyll was degraded due to leaf senescence and desiccation. There were significant genotypic variations among tested varieties for measured indexes. Significant associations between HTP-measured traits and traditional morphological and physiological traits were observed and indicated the feasibility of using field-based HTP related traits to predict and measure traits controlling camelina plant structure, growth, and response to environmental conditions.

Collectively, these studies will serve as a foundation for future experiments that aim to characterize the phenotypes of the entire camelina panel planted under well-watered and water-limited conditions in the field. Genome-wide association studies will then be employed to identify genes and genetic markers associated with agronomically important traits (Goal 2). Comparison of results between field- and greenhouse-grown plants will also help determine how well the greenhouse-based studies translate to findings in the field. The results will also help identify a subset of plant lines that will be analyzed in an advanced yield trials at four different locations, including Maricopa, AZ, High Plains, NE, Quincy, FL, and Morris, MN (Goal 3).

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.

Title: Resistance to Stalk Pathogens for Bioenergy Sorghum

Deanna Funnell-Harris,[1,2] (Deanna.Funnell-Harris@ars.usda.gov), Patrick O'Neill,[1,2] Scott Sattler,[1,2] Maya Khasin[1,2] and Erin Scully[3,4]

[1]United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Wheat, Sorghum and Forage Research Unit, Lincoln, NE; [2]University of Nebraska, Lincoln; [3]USDA-ARS Center for Grain and Animal Health Research, Stored Product Insect and Engineering Research Unit; and [4]Kansas State University, Manhattan

Project Goals: This research is focused on discovering molecular and metabolic networks that drive sorghum resistance or tolerance to stalk rot pathogens. We will use advanced molecular techniques to identify key factors associated with resistance and tolerance to stalk pathogens in sorghum lines optimized for key bioenergy traits (modified phenylpropanoid metabolism), with enhanced drought tolerance (nonsenescence) or with increased tolerance to stalk pathogens. Some lignin altered lines and drought tolerant lines already have demonstrated increased tolerance to these pathogens but the mechanisms that mitigate pathogenic growth have yet to be identified. Our goal is to develop sorghum lines that withstand increased pathogen loads under reduced water conditions based on knowledge gained through this research.

Sorghum is a promising bioenergy crop with high yield potentials and significant tolerance to both drought and heat. However, sorghum is prone to stalk rots, which can significantly limit sorghum biomass yield through growth reductions and lodging. Stalk rot-causing fungi normally grow endophytically within sorghum plants. When sorghum plants experience water stress, host changes often trigger a developmental switch causing the fungi to become pathogenic. The underlying plant molecular circuits that either limit or exacerbate this fungal transition from endophytic to pathogenic growth are not known and are the focus of this proposal. Several publicly available lines have previously demonstrated resistance or tolerance to sorghum stalk pathogens, including lines with post-flowering drought tolerance (nonsenscence), which appears to suppress pathogenic growth, or a variety of lines that have exhibited increased resistance under field conditions. We have developed several near-isogenic sorghum *brown midrib* (*bmr*) *6* and *12* lines with altered lignin content and composition, which were previously demonstrated to have increased resistance or tolerance to sorghum stalk pathogens (1,3,4,5). Lignin, a component of plant cell walls, has been a focus for development of bioenergy sorghums because its presence increases recalcitrance of biomass to cellusolic ethanol conversion, but its presence also increases total energy content of biomass, which is important for thermal conversion technologies. To increase energy content, we have engineered sorghum plants overexpressing a Myb transcription factor that induces synthesis of monolignols, the lignin subunits, and a gene encoding caffeoyl-CoA *O*-methyltransferase, a monolignol pathway enzyme. Both the transgenic and *bmr* plants accumulate phenolic intermediates from monolignol biosynthesis that inhibit stalk pathogens *in vitro* (4). We have identified a procedure to determine pathogen survival in lesions and asymptomatic tissues of sorghum peduncles (top of the stalk, below the head; 3) and have recently developed a controlled-environment, water-stress assay, which reliably induces the developmental switch from endophytic to pathogenic growth of sorghum stalk rot fungi.

Our recent research may have identified sources of resistance in *bmr6* and *bmr12* lines, relative to the wild-type, to two stalk rot pathogens, *Fusarium thapsinum* and *Macrophomina phaseolina*. We have previously shown that following inoculation of peduncles with each of these fungi a visible lesion is first apparent at 3 days post inoculation (dpi) and lesion expansion is first apparent at 13 dpi (2). In the current research, there were significant differences in mean lesion lengths resulting on *bmr6* and *bmr12* plants at 13 dpi with each fungus under adequate water or water deficit conditions as compared to wild-type plants with these treatments. In particular, *bmr6* plants under the adequate water treatment and both *bmr6* and *bmr12* plants under water deficiency had significantly smaller mean lesion lengths than wild-type plants after inoculations with *M. phaseolina*. No significant differences were apparent after inoculations with *F. thapsinum* under adequate water, but both *bmr* lines had significantly smaller mean lesion lengths than wild-type under water deficit conditions. Interestingly, *bmr12* plants had significantly smaller mean lesion lengths under water deficit than under adequate water when inoculated with either pathogen, counter to expected response to stalk pathogens under water stress. Across both water conditions, *bmr12* plants had reduced *F. thapsinum* survival within lesions. At 3-cm beyond the lesion border, there was reduced pathogen survival in both *bmr6* (*F. thapsinum* and *M. phaseolina*) and *bmr12* (*M. phaseolina*) plants as compared with wild-type. These results suggest that reduced survival of the pathogens within *bmr6* and *bmr12* stalks is likely due to changes induced by impaired monolignol biosynthesis.

These results strongly indicate that *bmr6* and *bmr12* lines, and near-isogenic wild-type, are promising for identification of genomic and metabolic markers for increased resistance to stalk pathogens under water deficit conditions.

References

1. Funnell-Harris, D. L., O'Neill, P. M., Sattler, S. E., Gries, T., Berhow, M. A. and Pedersen, J. F. (2017). Response of sorghum stalk pathogens to *brown midrib* plants and soluble phenolic extracts from near isogenic lines. *European Journal of Plant Pathology, 148,* 941–953, doi:10.1007/s10658-017-1148-2
2. Funnell-Harris, D. L., O'Neill, P. M., Sattler, S. E., & Yerka, M. K. (2016). Response of sweet sorghum lines to stalk pathogens *Fusarium thapinum* and *Macrophomina phaseolina*. *Plant Disease, 100*, 896-903, doi:10.1094/PDIS-09-15-1050-RE.
3. Funnell-Harris, D. L., Pedersen, J. F., & Sattler, S. E. (2010). Alteration in lignin biosynthesis restricts growth of *Fusarium* spp. in brown midrib sorghum. *Phytopathology, 100*, 671-681, doi:10.1094/PHYTO-100-7-0671.
4. Funnell-Harris, D. L., Sattler, S. E., & Pedersen, J. F. (2014). Response of *Fusarium thapsinum* to sorghum *brown midrib* lines and to phenolic metabolites. *Plant Disease, 98*, 1300-1308, doi:10.1094/PDIS-09-13-0980.
5. Funnell, D. L., & Pedersen, J. F. (2006). Reaction of sorghum lines genetically modified for reduced lignin content to infection by *Fusarium* and *Alternaria* spp. *Plant Disease, 90*, 331-338, doi:10.1094/PD-90-0331.

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 2016-67009-25420.

Phylogenomic Discovery and Engineering of Nitrogen Fixation into the Bioenergy Woody Crop Poplar

Jean-Michel Ané,[1] Pamela Soltis,[2,3] Douglas Soltis,[2,3] Robert Guralnick,[2,3] Ryan Folk,[2,3] Sushmita Roy,[1] and **Matias Kirst**[2]* (mkirst@ufl.edu)

[1]University of Wisconsin, Madison; [2]University of Florida, Gainesville; [3]Florida Museum of Natural History, Gainesville

URL: http://nit_fix.org; Twitter: @nit_fix

Nitrogen (N) availability is critical for high biomass productivity of bioenergy crops. Despite the abundance of N_2 in the atmosphere, plants cannot access it. Instead, plants must absorb available N in the soil, provided through intensive and costly fertilization. Some species acquired the capability to obtain N through a mutualistic relationship with bacteria and archaea, but this capability is absent from most bioenergy crops. Our overall goal is to discover the underlying genome novelties that evolved this mutualistic relationship using a comparative phylogenetic framework to contrast related species that possess and lack this ability. Genomic novelties will be evaluated for their effect on root nodule development in *Medicago* (nodulating) and poplar *hairy root* organ cultures (non-nodulating). Next will engineer nodule development in poplar plants and test the impact of these structures on N-fixation, whole-plant development and biomass productivity and composition.

This project is funded by the Biosystems Design Program from the Biological and Environmental Research (BER) Office of Science at the U.S. Department of Energy (grant #DE-SC0018247).

Systems Biology to Improve Camelina Seed and Oil Quality Traits

John Browse[1*] (jab@wsu.edu), Jesse Bengtsson[1], Wilson Craine[1], Scot Hulbert[1], Kevin King[2], Huang Li[2], GunNam Na[2], Mehmet Ozseyhan[2], and **Chaofu Lu**[2]

[1]Washington State University, Pullman; [2]Montana State University, Bozeman

Project Goals: Camelina has great potential to become a sustainable high energy-yielding source of biofuel in the US. This project aims to address two critical needs for realizing this potential: to increase seed size and oil content for improved seedling establishment and oil yield, and to optimize oil quality for satisfactory fuel properties. Specifically, quantitative trait loci (QTLs) and molecular markers associated with these important traits will be identified using high-density genome maps and repeated field trials in Montana and Washington states. Modern genomics and biotechnological approaches will be employed to uncover novel molecular mechanisms (including genes and gene networks regulated by microRNAs and transcription factors) regulating fatty acid modification, oil accumulation and seed size in Camelina.

Camelina (*Camelina sativa*) is a low-input, non-food oilseed plant that has great potential to become a sustainable high energy-yielding source of biofuel in the US. This project aims to address two critical needs for realizing this potential: 1) to increase seed size and oil content for improved seedling establishment and oil yield, and 2) to optimize oil quality for satisfactory fuel properties. Progresses have been made in the following areas:

1. To identify quantitative trait loci (QTL) and molecular markers associated with seed size, oil content and other important agronomic traits. We are using two complimentary populations to identify QTL controlling these traits in camelina. 1) A panel consisting 230 accessions of *Camelina sativa* and 21 camelina varieties is used for genome wide association studies. These lines were planted in the fields in Bozeman, MT and Pullman, WA in April-May 2017 and showed wide variations in several important traits including seed size, oil content and flowering time. 2) A biparental population derived from a cross between two contrasting varieties is used for linkage mapping. Using the single-seed descent (SSD) method, we developed a population comprising 361 recombinant inbred lines (RILs) from the cross between a large-seed variety Pretyzh (2.21mm^2, 1.8 mg/seed, oil content 31.6%) originated from Ukraine, and Suneson (1.51mm^2, 1.3 mg/seed, oil content 36.9%), a variety resulted from the selection in a German accession (PI 633192). The SSD population is now at the F6 generation. Field evaluation in Bozeman, MT showed great variation in several phenotypic traits (e.g., seed size, pod size, oil content, plant height, and flowering time). Sequencing of the parental lines and the population is underway to discover SNP markers for construction of a genetic linkage map.

2. To discover novel molecular mechanisms (including gene networks regulated by microRNAs and transcription factors) regulating fatty acid modification and seed size in camelina. In collaboration with the Joint Genome Institute (JGI), RNA samples have been collected from camelina lines that differ in seed size and fatty acid composition (e.g., high linolenic acid) for

RNAseq and microRNA profiling. Over 40 miRNAs have been selected and specifically overexpressed in camelina seeds. Preliminary data in transgenic plants suggested that several microRNAs may play important roles in developing seed in camelina. For example, overexpression of miR167a showed pleiotropic effects such as increased seed size and delayed seed maturation. Especially, the miR167a overexpressed seeds had increased 18:2 and decreased 18:3 compared to the seeds from wild-type, while other fatty acids remained unchanged. Comparative transcriptome analysis and chromatin immunoprecipitation experiments suggest that miR167a regulates fatty acid composition in camelina seeds by suppressing the expression of *FAD3* through a regulatory cascade involving AUXIN RESPONSE FACTORs (ARF6/8), the putative targets of miR167, and several transcription factors involved in lipid biosynthesis such as ABSCISIC ACID-INSENSITIVE 3 (ABI3), BASIC LEUCINE-ZIPPER 2 (bZIP2).

3. Modification of fatty acid composition in camelina seeds. 1) An effective tool using the CRISPR/Cas9 technology has been successfully developed in camelina (Ozseyhan et al., 2018). Homozygous knockout mutants were successfully created in a single generation by simultaneously targeting three *FAE1* genes using an egg cell-specific Cas9/gRNA expression. Very-long-chain fatty acids in the mutants were reduced to less than 2% of total fatty acids compared to over 22% in the wild type, and the C18 unsaturated fatty acids were concomitantly increased. 2) Artificial microRNA was used to down-regulate the expression of *FATB* in camelina seed. Over 40% reduction of saturated fatty acids (16:0+18:0) was observed in transgenic seeds compared to the non-transgenic wild type. Transgenic seeds also contained increased linolenic acid (18:3). 3) A fatty acid desaturase (DES9*), derived from the cyanobacterial 16:0/18:0 acyl lipid desaturase by directed evolution, was used to reduce levels of 16 and 18-carbon saturated fatty acids. Saturated fatty acid levels were reduced by more than 60%, compared with control plants. Monounsaturated fatty acid products, 16:1 and 18:1, were also greatly increased in seeds expressing the desaturase. The seed from transgenic lines with the lowest saturate levels had reduced oil, but there were only mild phenotypic changes in other lines.

Publication

Ozseyhan, M.E., Kang J., Mu X. and Lu C., 2018. Mutagenesis of the *FAE1* genes significantly changes fatty acid composition in seeds of *Camelina sativa*. Plant Physiol Biochem **123**, 1-7.

This research project is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0016400. The authors thank the Joint Genome Institute to assist in camelina genome and RNA sequencing experiments.

Trace Metal Storage in Metal-Specific, Lysosome-Related Vacuolar Compartments

Stefan Schmollinger[1*](schmolli@chem.ucla.edu), Martina Ralle,[2] Si Chen,[3] Mugi Tsednee,[1] Anne Glaesener,[1] Daniela Strenkert,[1] Colleen Hui,[1] Kristen Holbrook,[1] Madeli Castruita,[1] Peter K. Weber,[4] Jennifer Pett-Ridge,[4] and **Sabeeha S. Merchant[1]**

[1]University of California, Los Angeles, CA; [2]Oregon Health and Sciences University, Portland, OR; [3]Argonne National Laboratory, Argonne, IL; [4]Lawrence Livermore National Laboratory, Livermore, CA

URL: http://www.chem.ucla.edu/dept/Faculty/merchant/#research

Project goals: Transition metals are of crucial importance for primary productivity; their scarcity limits crop growth and carbon sequestration on a global scale. The most important elements are copper (Cu), iron (Fe) and manganese (Mn), which all serve as cofactors to enable redox biochemistry, especially for oxygenic photosynthesis, the process that transformed earth's primitive anoxic environment and is largely responsible for primary production today. The single-celled, eukaryotic green alga *Chlamydomonas reinhardtii* is an excellent model system to study trace metal biology in phototrophic organisms, with all the advantages of a microbial system, well-characterized photosynthetic and trace metal metabolic machinery. The goal of this project is to identify, differentiate and characterize the different trace metal storage compartments in the Chlamydomonas cell, and uncover the dynamics of trace metal storage and mobilization in situations of excess and limitation in single cells.

Abstract: *Chlamydomonas reinhardtii* is a unicellular green alga that has been widely used as a plant reference system for six decades, it has a quick generation time (~ 6h), can be synchronized and grown to high densities and its three genomes are sequenced and well-annotated [1]. We have utilized Chlamydomonas as a reference organism for decades to understand the principles underlying trace metal utilization and economy in a photosynthetic cell, and have identified a repertoire of assimilatory and distributive transporters, discovered mechanisms for reducing the metal quota and recycling metal cofactors from non-essential to essential proteins in situations of sustained elemental deficiency [2].

Chlamydomonas requires a broad spectrum of metal cofactors to sustain its photosynthetic, respiratory and metabolic capabilities, and iron (Fe), copper (Cu) and manganese (Mn) are the major transition metals involved in these processes. The metal catalysts in the photosynthetic electron transfer chain span a potential of about 2V, with a strong Mn-containing oxidant in photosystem II, enabling the oxidation of water and a strong, Fe-containing reductant in PS I enabling the reduction of $NADP^+$ for the synthesis of reduced carbon-containing compounds from CO_2. While indispensable and often growth-limiting when absent, redox active metals are

toxic for cells, and are therefore tightly bound or sequestered upon uptake. They are either handled by a set of intracellular ligands (protein chaperones, metallothioneins, phytochelatins or glutathione, for example) or sequestered into specific compartments, like vacuoles.

We identified and characterized the acidocalcisome as a major storage site for Cu (in Zn deficiency and Cd toxicity), Fe (in Zn deficiency and in Fe excess conditions) and Mn (in Mn excess conditions). The acidocalcisome is an acidic vacuole in the cytosol, defined by the presence of pyrophosphate and polyphosphate complexed with calcium [3,4]. It can be identified as an electron-dense granule by transmission electron microscopy, or more precisely with multimodal X-ray fluorescence microscopy (XFM) analysis according to its characteristic high calcium and phosphorus content. XFM allows to absolutely quantify cellular trace metal contents, since no sectioning is required for one Chlamydomonas cells and metabolic states can be conserved rapidly using either vitrification or chemical fixation. We used XFM on the bionanoprobe (beamline 9-ID-C) at the Advanced Photon Source at the Argonne National Laboratory to determine the spatial distribution of trace metals within algae cells, and quantified the contents of the acidocalcisomes in situations of various trace metal hyper-accumulation (Fe, Cu, Mn). We utilized a set of different mutants, including the vacuolar transporter chaperone (*vtc1*) mutant strains, that are defective in polyphosphate synthesis and where acidocalcisomes are highly diminished [5], and the copper transporter (*ctr2*) mutant strain to distinguish two distinct vacuolar sub-types, depending or their elemental composition. Additionally, quantification of the trace metal content of individual cells via XFM and comparisons to data for cell cultures acquired with inductively-coupled plasma mass spectrometry (ICP-MS/MS) allowed us to distinguish the nutritional state for Cu and Fe in single cells.

References:

1. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. (2007) The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. Science 318: 245-250.
2. Merchant SS (2010) The elements of plant micronutrients. Plant Physiol 154: 512-515.
3. Blaby-Haas CE, Merchant SS (2014) Lysosome-related organelles as mediators of metal homeostasis. Journal of Biological Chemistry 289: 28129-28136.
4. Hong-Hermesdorf A, Miethke M, Gallaher SD, Kropat J, Dodani SC, et al. (2014) Subcellular metal imaging identifies dynamic sites of Cu accumulation in Chlamydomonas. Nat Chem Biol 10: 1034-1042.
5. Aksoy M, Pootakham W, Grossman AR (2014) Critical function of a Chlamydomonas reinhardtii putative polyphosphate polymerase subunit during nutrient deprivation. Plant Cell 26: 4214-4229.

This work was performed at the Department of Energy's UCLA-DOE Institute and is supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421 and under the auspices of the U.S. Department of Energy at LLNL 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. Use of the Advanced Photon Source at Argonne National Laboratory supported by the U.S. DOE, Office of Science, Office of Basic Energy Sciences, under contract number: DE-AC02-06CH11357.

Genomic-Enabled Breeding of *Miscanthus*

Lindsay V. Clark[1], Kossonou Anzoua[2], Larisa Bagmet[3], Pavel Chebukin[3], Bruce E. Coulman[4], Elena Dzyubenko[3], Nikolay Dzyubenko[3], Xiaoli Jin[5], Douglas A. Johnson[6], Uffe Jørgensen[7], Jens Bonderup Kjeldsen[7], Stephen P. Long[1], Hironori Nagano[2], Karen Koefoed Petersen[7], Andrey Sabitov[3], Ji Hye Yoo[8], Chang Yeon Yu[8], Toshihiko Yamada[2], and **Erik J. Sacks**[1*] (esacks@illinois.edu)

[1]University of Illinois, Urbana, IL; [2]Hokkaido University, Sapporo, Japan; [3]Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia; [4]University of Saskatchewan, Saskatoon, Canada; [5]Zhejiang University, Hangzhou, China; [6]USDA-ARS Forage and Range Research Lab, Logan, UT; [7]Aarhus University, Tjele, Denmark; and [8]Kangwon National University, Chuncheon, South Korea

Project Goals: The primary goal of this project is to facilitate the development of *Miscanthus* as a bioenergy crop by obtaining fundamental knowledge about its genetic diversity, population structure, and environmental adaptation. We also seek to identify molecular markers associated with traits of interest to greatly increase the efficiency of breeding improved cultivars. Such knowledge will also facilitate efforts to introgress useful genes for abiotic and biotic stress tolerance from *Miscanthus* into *Saccharum*, thereby bolstering the development of more winter-hardy and disease-free sugarcanes and energycanes than are currently available.

Miscanthus is a genus of C_4 perennial grasses native to East Asia and Oceania, and it is a promising biomass crop for bioenergy and bioproducts. However, for *Miscanthus* biomass production, only one sterile triploid genotype of *M. ×giganteus* has been grown commercially in North America and Europe; this clone was introduced from Japan to Denmark in the 1930s. Thus, there is a pressing need to expand the diversity of *M. ×giganteus* genotypes available to growers. *M. ×giganteus* is derived from *M. sinensis* and *M. sacchariflorus*. Thus, a prerequisite to breeding improved *M. ×giganteus* is to obtain, characterize, and breed diverse germplasm pools of both *M. sinensis* and *M. sacchariflorus*.

Through two parallel DOE Feedstock Genomics projects, we have revealed the population structure of *M. sinensis* and *M. sacchariflorus* throughout most of these species' native ranges in East Asia. From 617 *M. sinensis* genotypes and ~21,000 RAD-seq SNP markers, we identified seven genetic groups (all diploid) via Structure analysis and discriminant analysis of principal components that corresponded to geographically distinct regions. Similarly, from 764 *M. sacchariflorus* genotypes and ~35,000 SNPs, we identified six geographically distinct genetic groups, including two tetraploid groups from Japan, an independently derived tetraploid group from Korea and N China, and three diploid groups. *M. sinensis* shows evidence of a large population bottleneck during the last glacial maximum, but the more cold-adapted *M. sacchariflorus* was less restricted in range and has higher genetic diversity overall.

These population structure studies, combined with data from field trials at multiple locations in Asia and North America, have facilitated the identification of SNPs associated with key traits, such as biomass yield, overwintering ability, and flowering time. The field trials for the *M. sinensis* study have been completed but those for *M. sacchariflorus* are still underway. Notably, dry biomass yield of four *M. sinensis* accessions exceeded 80 Mg ha^{-1} in Zhuji, China, approaching the highest observed for any land plant. Additionally, six *M. sinensis* in Sapporo, Japan and one in Leamington, Canada also yielded more than the triploid *M. ×giganteus* 'Illinois' control, with values exceeding 20 Mg ha^{-1}. For genome-wide association (GWA) analyses, the *M. sinensis* germplasm panel was genotyped with 46,177 SNPs, enabling identification of 27 significant SNPs for yield. Genomic prediction accuracy for biomass yield of *M. sinensis* was 0.47 over five northern sites and 0.65 for our southern location in Zhuji, China. These results have established a baseline of data for initiating genomics-assisted breeding to improve biomass yield of *M. sinensis* and *M. ×giganteus* in a diverse set of relevant geographies. Preliminary results from the *M. sacchariflorus* study indicate that we are making similar gains with that species. Furthermore, the parallel nature of these *M. sinensis* and *M. sacchariflorus* studies, enable genomic comparisons between the two species; to illustrate this point, we present comparative GWA analyses for flowering time, a highly repeatable trait.

Future work, with support from Center for Advanced Bioenergy and Bioproducts Innovation (a DOE Bioenergy Research Center), will begin to explore gains that can be realized for yield and winter-hardiness by implementing genomic selection within the *M. sinensis* and *M. sacchariflorus* genetic groups that we identified previously. This work will benefit greatly from the *M. sinensis* reference genome released this year by JGI. For example, using the pre-release version of the genome, we have been able to identify hundreds of thousands of SNPs from the *M. sacchariflorus* panel, which is an order of magnitude increase.

Publications

1. Jin, X., X. Chen, C. Shi, M. Li, Y. Guan, C.Y. Yu, T. Yamada, E.J. Sacks, and J. Peng. 2017. Determination of hemicellulose, cellulose and lignin content using visible and near infrared spectroscopy in *Miscanthus sinensis*. Bioresource Technology 241:603-609 2017/10.
2. Clark, L.V., E. Dzyubenko, N. Dzyubenko, L. Bagmet, A. Sabitov, P. Chebukin, D.A. Johnson, J.B. Kjeldsen, K.K. Petersen, U. Jørgensen, J.H. Yoo, K. Heo, C.Y. Yu, H. Zhao, X. Jin, J. Peng, T. Yamada, and E.J. Sacks. 2016. Ecological characteristics and *in situ* genetic associations for yield-component traits of wild *Miscanthus* from eastern Russia. Annals of Botany doi: 10.1093/aob/mcw137.
3. Clark, L.V., J.R. Stewart, A. Nishiwaki, Y. Toma, J.B. Kjeldsen, U. Jørgensen, H. Zhao, J. Peng, J.H. Yoo, K. Heo, C.Y. Yu, T. Yamada, and E.J. Sacks. 2015. Genetic structure of *Miscanthus sinensis* and *M. sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. J. Exp. Bot. doi:10.1093/jxb/eru51.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012379, DE-SC0016264 and USDA-ARS NPGS Plant Exploration Program, Cooperative Agreement #59-1275-1-338.

Introgression of Novel Disease Resistance Genes from *Miscanthus* into Energycane

Alexander E. Lipka[1], Jack C. Comstock[2], Sushma Sood[3], Vanessa S. Gordon[3], Keiji Motomura[4], Toshihiko Yamada[5], and **Erik J. Sacks**[1*] (esacks@illinois.edu)

[1]University of Illinois, Urbana, IL; [2]Department of Plant Pathology, University of Florida, Gainsville, FL ; [3]USDA-ARS Sugarcane Field Station, Canal Point, FL; [4]University of the Ryukyus, Okinawa, Japan; [5]Hokkaido University, Sapporo, Japan

Project Goals: Our long-term goal is to improve energycane productivity and sustainability by providing resistance to key diseases with novel genes from *Miscanthus*.

1) **In miscane BC_1 populations (sugarcane x (sugarcane x *Miscanthus*)), identify molecular markers associated with novel genes from *Miscanthus* that confer resistance to at least two out of four of the following economically important diseases of sugarcane: ratoon stunt, yellow leaf, orange rust, and smut.**
2) **Compare effectiveness of different molecular marker analysis methods for selecting disease resistance alleles in miscane backcross populations. In particular, compare the pseudo-testcross QTL mapping strategy with genomic selection.**
3) **Screen germplasm collections of *M. sinensis* and *M. sacchariflorus* for resistance to ratoon stunt, yellow leaf, orange rust, and smut to confirm that *Miscanthus* is uniformly immune to these diseases as prior data suggest or to quantify genetic variation for resistance if not all accessions are resistant.**

Sugarcane is among the world's leading bioenergy crops. Modern sugarcane cultivars are derived from a relatively small set of founder genotypes, which has contributed to cultivar susceptibility to diseases. Modern sugarcane cultivars originated in the late 1800's when cultivars of *Saccharum officinarum* with high sugar yield potential but disease-susceptibility were crossed with the undomesticated, non-sugar producing but disease-resistant *S. spontaneum*, then backcrossed to *S. officinarum* to recover sugar yield. Recent efforts to improve sugarcane for disease resistance, pest resistance, and abiotic stress tolerance have continued to rely primarily on introgressions from *S. spontaneum*. Similarly, energycanes (sugarcanes bred specifically for energy) are also bred from modern sugarcane cultivars and *S. spontaneum*.

Yet, diseases are the primary constraints to cane productivity in commercial fields. Sugarcane is a perennial crop but a planting's commercial productivity in the U.S. is typically limited to 3-5 years because disease pressure reduces yield and decreases stand via plant death each year. Moreover, the utility of a successful cane cultivar for commercial production is typically ended by the emergence of a virulent strain of a common pathogen, resulting in reduced yields. Though many diseases affect sugarcane, the following four are of especially great concern (Rice, 2007): 1) ratoon stunt (bacterium *Leifsonia xyli* subsp. *xyli*.); 2) sugarcane yellow leaf (virus, *Sugar Cane Yellow Leaf Virus* (*SCYLV*); 3) orange rust (fungus *Puccinia kuehnii*); and, 4) smut (fungus *Sporisorium scitamineum*).

Arguably, *Miscanthus* would be a better source of genes for improving sugarcane than *S. spontaneum* because the former is highly resistant to diseases and pests, is more broadly adapted to diverse environments, and is more genetically distant from *S. officinarum* (thus providing more novel alleles). However, the use of *Miscanthus* for improvement of sugarcane has been limited, perhaps in large part due to prior lack of access to *Miscanthus* germplasm. PIs Sacks and Yamada have developed large and diverse *Miscanthus* germplasm collections. Recently, we have obtained more than a dozen F_1 hybrids between sugarcane and *Miscanthus* (including hybrids between cane × *M. sinensis* and cane × *M. sacchariflorus*). Backcrosses to sugarcane have been made and will soon be germinated.

Disease screenings were conducted in 2017 on miscane F_1 progeny, and core collections of *M. sinensis* and *M. sacchariflorus*. For the most part, our hypotheses of ubiquitous resistance in *Misccanthus* were confirmed. For smut, all of the 66 *Misccanthus* genotypes tested were fully resistant, and for orange rust all but one of the *Misccanthus* genotypes tested were fully resistant. For ratoon stunt disease, only two *Misccanthus* genotypes were susceptible and these were much less susceptible that the sugarcane positive control. Response of *Misccanthus* to sugarcane yellow leaf virus was more variable than for the other diseases tested, with 14/31 *M. sacchariflorus* and 24/35 *M. sinensis* fully resistant, but the remainder were partially to fully susceptible. Of six miscane genotypes tested to date, one was fully resistant to all four diseases, six were fully resistant to orange rust and smut, four were resistant to ratoon stunt, and three were resistant to sugarcane yellow leaf virus.

Subsequently, backcross progeny will be evaluated for disease-resistance and we will compare the standard QTL analysis method for sugarcane with genomic selection. To facilitate this comparison, we are developing optimized statistical models for relating genotype to phenotype in backcross (introgression) populations of polyploid plant species. In this work, we have simulated traits with contrasting genetic architectures in biparental crosses from multiple polyploid plant species, including sugarcane, and then compared the ability of markers that tag QTL to predict trait values to that using genome-wide marker sets. We expect these results to provide guidance on when it will be most appropriate to focus resources towards obtaining whole-genome marker sets versus sequencing only regions in the vicinity of QTL peaks.

Because disease susceptibility is a major limitation for cane production, cultivar durability, and sustainability, the introgression of effective and durable resistances to major cane diseases into energycane from *Miscanthus* is expected to increase the economic and environmental benefits of energycane while reducing costs and supply-side risks. Such benefits should promote further investment by industry in energycane production and processing capacity.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0016264.

Identifying productive energy sorghum germplasm for water-limited and nitrogen-limited environments using classical and advanced phenotyping methods - one piece of the puzzle for creating sustainable biomass production systems

Daniel Schachtman*[1] (Daniel.schachtman@unl.edu), Zachary Brenton[2], Asaph Cousins[3], Cody Creech[1], Ismail Dweikat[1], Stephanie Futrell[1], Emily Goren[4], Stephen Kresovich[2], Allyn Pella[1], Jessica Prenni[5], Jiating Li[1], Peng Liu[4], Ellen Marsh[1], Yeyin Shi[1], Amy Sheflin[5], Bala Sonawane[3]

[1] University of Nebraska, Lincoln, NE; [2]Clemson University, SC, [3]Washington State University, Pullman, Washington, [4]Iowa State University, Ames, Iowa, [5]Colorado State University, Fort Collins, CO

https://sorghumsysbio.org

Project Goals:

1. **Conduct deep census surveys of root microbiomes concurrent with phenotypic characterizations of a diverse panel of sorghum genotypes to define the microbes associated with the most productive lines under drought and low nitrogen.**
2. **Associate systems-level genotypic, microbial, and environmental effects with improved sorghum performance using robust statistical approaches.**
3. **Develop culture collections of sorghum root/leaf associated microbes**
4. **Perform controlled environment experiments for in-depth characterization and hypothesis testing of $G_{sorghum}$ x $G_{microbe}$ x $E_{interactions}$.**
5. **Validate physiological mechanisms, map genetic loci for stress tolerance, and determine the persistence of optimal microbial strains.**

This is a multifaceted project involves both plant genetics and studies of the soil microbial communities associated with sorghum. The overall objective is to create a sustainable biofuel feedstock system for production on marginal soils that lack sufficient moisture and nutrients. Part of the group is focused on understanding the relationship between soil microbes and sorghum genotype (See poster by Chiniquy et al.), culturing sorghum associated bacteria, characterizing the changes in the soil microbiome due to abiotic stresses and gaining a fundamental understanding of the genotypic changes in root metabolites and how those interact factors interact with soil microbe community composition (See poster by Sheflin et al.).

This poster will provide an update on the plant germplasm screening that was conducted in 2017. Two locations were used to characterize the abiotic stress tolerance of 24 sorghum lines. One field was located in western Nebraska and used to study response to drought conditions and the other field in eastern Nebraska was used to characterize response to low nitrogen. End of season relative biomass and height data will be presented. In order to further characterize the physiological and biochemical responses to stress more advanced methods are being developed. Plants and plots were characterized by both hyperspectral reflectance and UAV flights which were calibrated using data collected on individual leaves for chlorophyll, nitrogen content, specific leaf area and sucrose content of leaves and stems. Those results

will be presented. Using these more advanced phenotyping methods we plan to characterize segregating populations of energy sorghum for nitrogen and water use efficiency in 2018 with the aim of mapping these traits. In addition, leaves were collected for carbon and nitrogen isotope analyses which we aim to use for development of more advance physiological tools for further in-depth characterization of water and nitrogen use efficiency of this germplasm (see poster by Cousins et al.).

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 and Nitrogen Isotope Composition in Diverse Sorghum Lines Under Differential Water and Nitrogen Treatments

Asaph B. Cousins[1]* (acousins@wsu.edu), Balasaheb V. Sonawane[1], Kira Veley[2], Jeffrey Berry[2], Rebecca Bart[2], 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 objectives*** of the research presented here are to:
1. Conduct phenotypic characterizations of a diverse panel of sorghum genotypes to define photosynthetic and isotope response under drought and low nitrogen conditions.
2. Test sorghum genotype by environment interactions in both controlled environment and field growth conditions.
3. Determine if measurements of leaf carbon and nitrogen isotope composition can screen for differences in water and nitrogen use efficiency in diverse sorghum genotypes.

Towards achieving our project goals we have screened 30 diverse sorghum lines under controlled environment growth conditions in the Bellweather Phenotyping System at the Danforth Center. This population included 18 energy, 2 grain and 10 sweet sorghum lines. Under the phenotyping system two separate experiments were conducted using a random block design to study the growth, photosynthetic and stable isotope response of this diverse sorghum panel to changes in nitrogen and water availability. Additionally, whole plant nitrogen and water use efficiency were estimated from the phenotyping data. In the first experiment each line was provided with two levels of nitrogen (n=3 to 4 plants per line per treatment) over a three week time period. In the second experiment, each line received a constant nutrient supply under three different watering treatments over a three-week growth period (n=3 to 4 per line per treatment). Individual plants were imaged daily at 4 different angular rotations to calculate plant growth and approximate biomass accumulation. Additionally, weighing and watering of plants was automated to maintain the required soil volumetric water content and the predetermined nutrient supply. Towards the end of both experiments the upper most fully expanded leaf from individual plants was used for gas exchange measurements with a LI-6400XT open gas exchange system (Li-COR Biosciences, Inc. Lincoln, NE). A portion of the same leaf was sampled for nitrogen and carbon isotope composition ($\delta^{15}N$ and $\delta^{13}C$, respectively). Leaf samples for $\delta^{15}N$ and $\delta^{13}C$ analysis were also collected from two field locations in Nebraska

in both 2016 and 2017. The western Nebraska field site was used to study response to drought conditions and the eastern location was used to characterize response to low nitrogen.

Initial characterization of these data sets suggests significant variation between the sorghum genotypes in their responses to nitrogen and water availability. For the controlled environment phenotyping experiment data will be presented on differences in nitrogen and water use efficiency, rates of CO_2 assimilation, stomatal conductance, intrinsic transpiration efficiency (TE_i) and photosynthetic nitrogen use efficiency (PNUE). Measurements of total leaf C/N content and leaf $\delta^{13}C$ and $\delta^{15}N$ will be presented from both the controlled environment and field experiments.

Future directions

Leaf level and whole plant traits will be assessed across genotypes in response to both changes in nitrogen and water availability. This information will be analyzed in comparison to field grown material to help identify and select for genomic traits and potentially elite lines for enhanced nitrogen and water use efficiency in sorghum.

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.

Engineering mitochondrial metabolic networks to increase yield and water use efficiency in *Camelina sativa*

Bibin Paulose,[1] Thomas Sharkey,[1] and **Danny J. Schnell**[1]* (schnelld@msu.edu)

[1]Michigan State University, East Lansing, MI 48823

https://plantbiology.natsci.msu.edu/directory/danny-j-schnell/

Project Goals:

Our research plan aims to establish the non-food oilseed crop plant, *Camelina sativa,* as a commercially viable, dedicated biofuel and bioproducts feedstock. We will focus 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). The major limitation in widespread adoption of Camelina as an industrial oilseed crop is its modest oil yield. Our research 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. Our overall objective is to achieve up to a 300% increase per hectare in oil production, thereby meeting the yield and cost targets of a competitive biofuels and bioproducts crop while retaining its advantages for growth in marginal environments.

Abstract:

Environmental conditions and metabolic regulatory mechanisms exert major constraints on plant growth and the productivity by limiting the levels of carbon capture in photosynthetic source tissues (leaves) and the fixed carbon available for transport and allocation to sink tissues (e.g. seeds). A major constraint on carbon capture is the competition between the productive carboxylation reaction and non-productive oxygenation catalyzed by Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) in the Calvin-Benson cycle of photosynthesis. Our research addresses this constraint by investigating how metabolic networks interact to achieve high rates of carbon fixation while removing the toxic products of oxygenation via photorespiratory metabolism. These studies utilize the model oilseed crop, *Camelina sativa* engineered with a mitochondrial metabolite transporter, Organellar Carrier Protein 1 (OCP1). Plants expressing OCP1 ($Camelina^{OCP}$) assimilate carbon dioxide 20-30% times faster, and show increased growth rates and 38-50% higher seed yields. As such, $Camelina^{OCP}$ plants provide an excellent system to investigate the metabolic networks of central carbon metabolism in source tissues and to identify the control points that limit carbon capture and constrain photoassimilate export for seed production in a crop plant.

On the basis of preliminary studies, OCP1 is hypothesized to enhance carbon dioxide assimilation as part of a metabolic release valve to reduce the levels of metabolites that inhibit

photosynthesis by 1) accelerating their recycling through the photorespiratory pathway, 2) aiding in the transfer of excess reducing equivalents from chloroplasts to mitochondria for dissipation via respiration, or 3) participating in the dissipation of excess reducing equivalents by augmenting the uncoupling capacity of mitochondria. We will present the details on the photosynthetic parameters impacted by OCP1 activity using advanced gas exchange and chlorophyll fluorescence measurements. These studies are complemented membrane transport studies and ^{13}C metabolic flux data and to define the metabolic network dynamics responsible for increased carbon capture in Camelina^{OCP}.

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/Pages/Research.aspx

Project Goals: This project aims to genetically improve the agronomic traits of Field Pennycress (*Thlaspi arvense* L.; pennycress) for its use as a profitable oilseed winter cover crop grown throughout the U.S. Midwest. We have identified a large number of EMS-induced pennycress mutant lines exhibiting a variety of improved agronomic traits. We have also developed and demonstrated the utility of pennycress *Agrobacterium*-mediated plant transformation and CRISPR-Cas9 genome editing, generating pennycress lines with undetectable levels of erucic acid in seed oil, reduced seed coat fiber, reduced pod shatter, and reduced seed dormancy. We are working to identify, characterize, and introgress into breeding lines these and additional trait-improving mutations so as to generate elite pennycress 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 so as to improve the seed meal nutritional value 4) Shortened time to maturity to consistently allow pennycress harvest in time to plant full-season soybeans.

Pennycress (*Thlaspi arvense*; Field pennycress) is an oilseed plant of the Brassicaceae family closely related to Arabidopsis, camelina, and rapeseed canola. Pennycress is native to Eurasia and naturalized to North America, growing widespread throughout temperate regions of the world. Pennycress can be grown as a winter annual oilseed-producing cover crop, for example, planted in the fall in standing corn and harvested in the spring in time to plant full-season soybeans throughout the 80 million-acre U.S. Midwest Corn Belt. Once commercialized, elite pennycress varieties will provide additional income to farmers and agribusinesses thereby strengthening rural communities. Pennycress will also provide ecosystem services as a cover crop, reducing soil erosion and nutrients runoff and providing habitat and pollinator support on otherwise vacant farmland.

Field trials with current isolates have demonstrated that pennycress can be seeded in upper Midwest cornfields in the late summer and fall, at which time the plants begin to grow then overwinter, producing mature seed in the spring that can be harvested without disrupting soybean planting or yields. As an energy crop adopted throughout the U.S. Midwest, pennycress varieties could annually produce 1.3 billion gallons of liquid transportation fuels and 15 million tons of high-protein seed meal, once modest breeding improvements are made (facilitated by resources and germplasm from this project). Many other products could also be produced from this oilseed feedstock. Longer-term agronomic and genetic improvements have the potential to more than double this impact.

While pennycress holds much agronomic promise, economically-viable varieties remain to be developed. Current varieties are hampered by suboptimal seed germination and stand

establishment, un-optimized maturity for a given growth zone, high seed glucosinolate and fiber content, and significant harvest loss due to pod shatter. We will provide highlights of our progress employing modern forward and reverse genetics and genomics strategies to rapidly generate and identify lines of pennycress that harbor mutations/natural gene variants conferring superior agronomic traits. These trait-improving alleles are being incorporated into breeding programs located in the Midwest. Our goal is to help develop elite pennycress varieties for commercialization thereby launching pennycress as a profitable oilseed-producing winter cover crop in the 2020's.

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.

"SyPro Poplar"
Improving Poplar Biomass Production under Abiotic Stress Conditions: An Integrated Omics, Bioinformatics, Synthetic Biology and Genetic Engineering Approach

Amir H. Ahkami [1]* (amir.ahkami@pnnl.gov), C. Neal Stewart, Jr.[2], Yongil Yang [2], Stephen DiFazio [3] Ryan Kelly [1], Yiran Liang [1], William Chrisler [1], and **Eduardo Blumwald [4]**

[1] Environmental Molecular Science Laboratory (EMSL), 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: In SyPro Poplar we intend to (i) study the functions of selected stress-responsive genes; (ii) discover novel motifs and construct stress-responsive synthetic promoters; and (iii) use these promoters to drive the expression of genes shown to confer abiotic stress tolerance in a variety of crops and develop abiotic stress-tolerant poplar seedlings in a coordinated fashion. We will use a combinatorial gene stacking approach with key transgenes driven by stress-responsive synthetic promoters to confer stress resistance. Our plan is to develop a series of abiotic stress-responsive synthetic promoters comprised of the stretch of DNA containing multiple copies of abiotic responsive cis-motifs upstream of a core-promoter in which abiotic stress specific transcription factors (TFs) bind to their cognate sequences to drive transcription under multiple abiotic stresses. The aim is the development of transgenic trees with sustained photosynthetic activity and increased biomass production under individual and the simultaneous occurrence of water deficit, increased soil salinity and elevated temperatures.

Abiotic stresses such as drought, high temperature and salinity are major causes for decreased crop yield, including decreased biomass from bioenergy feedstocks. Producing crops with robust phenotypes under variable growth conditions has proven to be challenging, with abiotic stress tolerance proving to be among the most challenging traits. Given the low input requirements needed for competitive bioenergy feedstock cultivation, irrigation and other mitigation is not economically feasible. Therefore resistance to stresses will likely be required to be genetically encoded within crops. The development of multi-stress tolerant plants is critical for both food and energy security with predictable and environmentally-independent biomass production being the most important trait[1]. In this regard, advanced -omics and plant synthetic biology methods are promising ways to accumulate favorable alleles associated with stress tolerance in a plant genome. Although various genes relevant to plant responses to different types of abiotic stresses have been characterized and used to improve plant productivity under stress, only limited success has been achieved in producing stress-tolerant bioenergy crops [2]. To date, constitutive overexpression (using CaMV 35S or ubiquitin promoters) of single genes has been the most common strategy for improving abiotic stress tolerance in plants. However, this strategy has had a relatively limited success due to involvement of multiple pathways in plant adaptation to stress and possible negative pleiotropic effects on plant growth [3]. For that reason, tissue-specific and stress responsive synthetic promoters can be designed to respond specifically to different stress conditions, effectively inducing transgene expression in plants while maintaining a low activity under normal growth conditions. Compared to natural promoters, synthetic promoters provide higher (or lower) expression level with high specificity, lower basal expression level, shorter length, and less sequence homology to any host genomic sequences [4]. However, the use of integrated strategies, including -omics, bioinformatics and biotechnological tools, for the development and testing of synthetic promoters has been restricted. Thus, construction of synthetic promoters with appropriate tissue specificity and inducibility is a nascent science. Moreover, to date, synthetic promoters have been mainly used to drive expression of reporter genes under controlled conditions. In SyPro Poplar, we aim to construct novel monopartite and tripartite synthetic tissue-specific promoters (by multimerization of novel abiotic responsive cis-elements) capable of driving transcription of novel stress-responsive genes under multiple abiotic stresses and under field conditions.

The proposed research program will focus on poplar leveraging its vast suite of genomic resources, including several high quality reference genomes for strategically-important species, and re-sequenced genomes, transcriptomes, proteomes, and metabolomes for large population samples. Moreover, it is a prevalent woody feedstock for improved next-generation biofuels, representing the most productive native trees in the northern hemisphere. Poplar has distinct advantages over other biofuels crops because the timing of harvest is very flexible, necessitating minimal long-term storage at processing facilities and thereby solving some major logistical challenges for biofuel production [5].

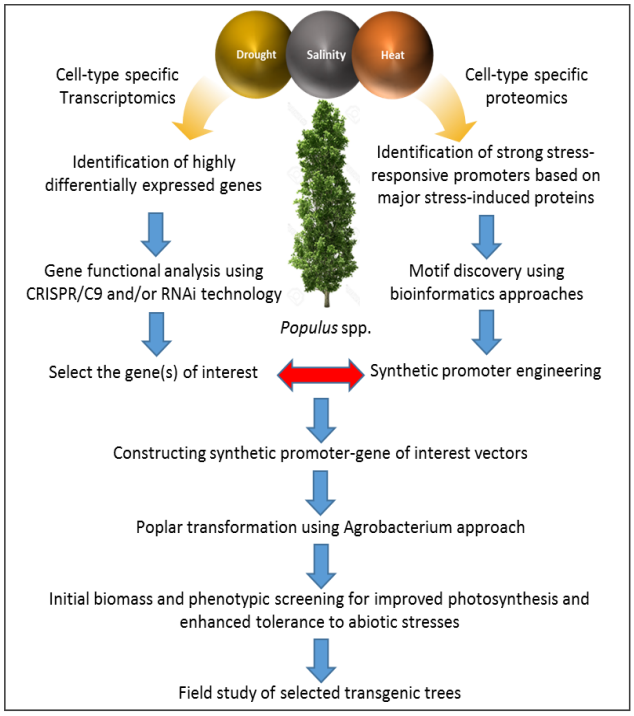


Figure 1. SyPro Poplar workflow.

In SyPro Poplar we intend to integrate omics, bioinformatics, synthetic biology and genetic engineering approaches to develop transgenic poplar trees with sustained photosynthetic activity and increased biomass production under individual and the simultaneous occurrence of water deficit, increased soil salinity and elevated temperatures.

Populus tremula x alba clone INRA 717 1-B4, as a model genotype with extremely high transformation efficiency and favorable silvicultural characteristics, has been propagated and used by SyPro Poplar team for initial test experiments and method optimizations including laser capture microdissection (LCM) and tissue specific proteomics based on newly developed PNNL capabilities, nanowell-based preparation in one-pot for trace samples (nanoPOTS), small-volume sample preparation, and ultrasensitive liquid chromatography-mass spectrometry (LC-MS). Preliminary genomics analysis and motif discovery of available poplar data is in progress using bioinformatics tools such as BioMart and MEME. To build a prototype in the first year of the project, vector construction and transformation of clone INRA 717 1-B4 with some selected stress-responsive genes under the control of both constitutive and stress-responsive inducible promoters is also in progress.

This research is supported by the U. S. Department of Energy (DOE) office of Biological and Environmental Research (BER), Genomic Science Program.

References:

1. Oliver RJ, Finch JW, Taylor G, 2009. Second generation bioenergy crops and climate change: a review of the effects of elevated atmospheric CO2 and drought on water use and the implications for yield. *Global Change Biology Bioenergy* **1**, 97-114.
2. Reguera M, Peleg Z, Blumwald E, 2012a. Targeting metabolic pathways for genetic engineering abiotic stress-tolerance in crops. *Biochim Biophys Acta* **1819**, 186-94.
3. Rivero RM, Kojima M, Gepstein A, *et al.*, 2007. Delayed leaf senescence induces extreme drought tolerance in a flowering plant. *Proc Natl Acad Sci U S A* **104**, 19631-6.
4. Liu W, Mazarei M, Peng Y, *et al.*, 2014. Computational discovery of soybean promoter cis-regulatory elements for the construction of soybean cyst nematode-inducible synthetic promoters. *Plant Biotechnol J* **12**, 1015-26.
5. Langholtz MH, Stoke BJ, Eaton LM, 2016. Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy. In. *Economic Availability of Feedstocks.* Oak Ridge National Laboratory. (1.)

Developing Non-food Grade *Brassica* Biofuel Feedstock Cultivars with High Yield, Oil Content, and Oil Quality that are Suitable for Low Input Production Dryland Systems.

Jack Brown,[*1], Jim B. Davis,[1] Aaron Esser,[2] Kurt Schroeder[1], Fangming Xiao,[1] and Zhang Zhiwu[2]. [1] University of Idaho, Moscow. [2] Washington State University, Pullman. [*] Presenting PI (jbrown@uidaho.edu)

Project Goals: (1) Develop superior non-food grade oilseed cultivars suitable for a range of PNW and other US environments with blackleg and abiotic stress resistance suitable for high-quality biofuel feedstocks; (2) Dissect genetic architecture of industrial *Brassica* oilseed germplasm for yield, quality, and resistance to blackleg fungus through a genome-wide association study with high-density genetic markers; (3) Develop and apply marker-assisted genomic selection techniques for blackleg resistance that will accelerate development of disease resistant oilseed cultivars; and (4) Use bioinformatics to identify putative PRR resistance genes that confer durable resistance to blackleg and use transgenic approaches to introgress PRR genes into adapted cultivar background.

Abstract: This project utilizes rapeseed (*Brassica napus*) and Indian mustard (*B. juncea*) oilseed germplasm and the long-term industrial oilseed breeding program at the University of Idaho, and the agronomy and molecular biology expertise from the University of Idaho and Washington State University. We have recently completed the first full growing season of the project.

Two hundred and forty spring *B. napus* genotypes were planted under field condition in spring of 2017. All these accessions had previously been genotyped, and were planted in replicated trials into ground with a proven history of blackleg. Plots were evaluated throughout the growing season for blackleg disease lesions on leaves. Few diseases symptoms were observed due to drier than usual weather post-planting. In addition, 20 spring *B. napus* lines were planted at two locations in replicated field trials where blackleg was controlled (via two foliar fungicide applications) and without fungicide. Blackleg lesions were recorded throughout the growing season. At maturity, seed was harvested and stem canker recorded, although few incidences were noted.

In fall of 2017, 280 winter *B. napus* genotypes were planted under field condition at two locations in northern Idaho (Grangeville, 4 replicates with smaller plots, and Genesee, two replicates with larger plots to allow combine harvest at maturity). All these accessions had previously been genotyped, and were planted in replicated trials into ground with a proven history of blackleg disease. To date, we have recorded blackleg leaf lesions in fall of 2017 and we will repeat lesion counts on two occasions in spring of 2018. Blackleg has always appeared a greater problem in winter *B. napus* than in spring plantings and significantly greater lesions counts were observed and markedly different resistance levels were found within the genotypes screened. At harvest, we will record yield from the Genesee trial which has larger plot size compared to Grangeville. All of these genotypes were included in an earlier project were we recorded seed yield. However, in that study we had severe winterkill in two of the four years and so yield data on these winter types was limiting. The 2017-2018 plots look good at the time of reporting, and the additional yield data will be invaluable to add to our data set and allow greater genotype by association studies.

In addition, 20 winter *B. napus* lines were planted at Grangeville ID, in replicated field trials where blackleg was controlled (via two foliar fungicide applications) and without fungicide. Blackleg lesions were recorded in fall of 2018 and we will repeat lesion counts on two occasions in spring of 2018. At maturity, seed will be harvested and stem canker recorded, although few

incidences were noted. This trial will help us understand the potential yield loss caused by blackleg disease in Pacific Northwest winter *B. napus* cultivars without fungicide control or genetic plant resistance to blackleg.

A genetically diverse genotypic collection of *B. juncea* had been collected at the University of Idaho. This diversity panel included Plant Introduction accessions (268 lines) from USDA, along with germplasm from within the University of Idaho *B. juncea* breeding program (232 lines). To ensure, continuity between genotyping and phenotyping, seed from each of the 500 genotypes was planted in the glasshouse in fall of 2016. A single plant was harvested from each accession and that seed thereafter used to plant the first field evaluation trials in spring of 2017. While the selected plants were at the seedling stage in the glasshouse a leaf sample was taken from each genotype and freeze dried for genotyping. In spring of 2017 the glasshouse seed increase produced sufficient seed to allow planting of a replicated yield trial at one location (Genesee) and a seed increase plot at another location (Moscow). A wide range of plant morphological data was recorded on plots from the replicated yield trial including, crop establishment, flower date, plant height, lodging, pod length, seeds pod^{-1}, and maturity. Any blackleg lesions observed were counted, and other disease and insect damage recorded. At crop maturity, seed was harvested and yield recorded. A sample of seed was taken from each plot and is currently being evaluated for oil content, oil quality, and seed mean quality.

Pattern Recognition Receptor (PRR) mediated resistance is dependent on perception of pathogen-associated molecular pattern (PAMP) by pattern recognition receptor (PRR) in plants. In addition, PRR-mediated resistance is potentially durable because PAMPs, as structural elements of molecules, are highly conserved and usually essential for liability or lifestyle of pathogens; and because pathogens are less likely to evade host resistance through mutation or deletion of PAMPs, compared with virulence effectors. We have been successful in identifying several putative PRR genes and these are being further investigated, before we begin transferring these PRR genes into the regionally adapted *B. napus* and *B. juncea* oilseed cultivars being selected in the breeding program. In addition we plan to pyramid multiple PRR genes into a single winter *B. napus* line by multiple crossing among transgenic plants expressing individual PRR genes.

Progress results from all research to date will be discussed in the context of the work that is to follow. The genetically superior *Brassica* non-food grade oilseed cultivars developed will drastically increase domestic industrial oilseed production. In addition, we will identify new and novel plant resistance genes for blackleg resistance and molecular marker assisted selection tools to accelerate plant breeding procedures. Increased domestic production of biofuel feedstock oil will reduce dependence and importation of fossil fuels, increase fuel security, and ensure US agricultural competitiveness with foreign countries.

Funding for this project is provided by: DOE-USDA, Plant Feedstock Genomics for Bioenergy: DE-FOA-0001444.

Genetics and Genomics of Pathogen Resistance in Switchgrass

Serge Edmé[1*] (serge.edme@ars.usda.gov), Gautam Sarath[1], **Nathan Palmer**[1], Rob Mitchell[1], Satyanarayana Tatineni[1], Gary Yuen[2], Anthony Muhle[2], and R.V. Chowda Reddy[1]
[1]USDA-ARS Wheat, Sorghum and Forage Research Unit, Lincoln, NE; [2]Plant Pathology Department, University of Nebraska, Lincoln, NE

Project Goals: This project aims at applying genomic selection (GS) in switchgrass, a key bioenergy crop in the US, on the genic regions causally linked to high biomass, ethanol yield, and durable resistance to fungal and viral diseases. The specific objectives are:
(1) Understand the genetic and genomic bases of pathogen response in regionally-adapted upland and lowland switchgrass breeding populations with contrasting disease symptoms under field conditions,
(2) Dissect the molecular underpinnings of the broad resistance or tolerance to pathogens in 'Kanlow' vs 'Summer', and
(3) Discover the molecular differences that permit systemic viral infections in some switchgrass plants, but not in other genetically-related plants.

Diseases, such as viral mosaic (caused by *Panicum mosaic virus*, PMV) and rust (caused by *Puccinia emaculata* and *Uromyces graminicola*), can cause significant losses in yield and quality traits in switchgrass (*Panicum virgatum*) bred for forage and bioenergy. Breeding by phenotypic recurrent selection can be effective at increasing genetic resistance in switchgrass populations. However, GS approaches are expected to increase the accuracy of selection, to concurrently improve yield and disease traits in switchgrass, and to ultimately generate greater genetic gains. Dense markers, such as single nucleotide polymorphisms (SNP) that span the whole genome, are required by genomic selection in order to predict breeding values of selection candidates by combining genotypes, pedigree, and SNP effects with phenotypic data.

Two interconnected reciprocal switchgrass populations derived by crossing 'Kanlow' (lowland) with 'Summer' (upland) are being followed across three generations. A reference germplasm set of switchgrass cultivars (from several US geographic regions) and selections was assembled to constitute a good allelic representation of the genome. Parental and progeny populations were phenotyped for biomass yield, Klason lignin content, ethanol yield, and disease ratings and genotyped with DArTseq to develop SNP markers. Two RNA-seq experiments (3 replicates of 10 pooled individual plants/timepoint/treatment) were sequenced and are being analyzed. In the first experiment, the 4th leaf from greenhouse-grown Kanlow and Summer plants were collected at seven timepoints over the course of two months in order to examine basal gene expression differences between the two populations. In the second experiment, the response of Kanlow and Summer plants to rust infection was examined by collecting the 4th leaf of infected and uninfected plants at 2, 7, 11, and 18 days after inoculation.

Genetic variation was assessed for viral mosaic and yield traits (dry matter yield, Klason lignin, and predicted ethanol yield) in an inter-ecotypic Summer x Kanlow population using linear and generalized linear mixed model approaches with restricted maximum likelihood. The two models were compared in their assessment of the genetic parameters, estimation of breeding values, and prediction of genetic gain. The analyses were also performed in the context of a multivariate animal model, which traces the pedigree back through three generations. A pedigree (Fig. 1) was

built to follow the flow of genes from parents to progeny across generations and to increase the efficiency of prediction models by including the pedigree and molecular relationship matrices. The resulting genetic parameters were used to update the selection index with disease resistance and determine the relative importance and contributions of each trait to maximizing genetic gains. Heritability estimates were 0.53 for dry matter yield, 0.36 for KL, 0.39 for ETOH, and 0.49 for mosaic ratings. Dry matter yield was genetically but negatively correlated with mosaic ratings (-0.44) and KL (-0.12ns), indicating that higher yielding genotypes were more resistant/tolerant to the virus (Edmé et al., 2017). 'Liberty', a cultivar released from the Summer x Kanlow population, was intermediate in resistance to rust and viral pathogens, which imply introgression of resistance factors from the paternal parent (Kanlow) into the maternal genome (Muhle et al., 2017).

The synergistic interaction between PMV and sPMV (*satellite PMV*) was investigated, using infectious cDNA clones of NE and TX isolates of PMV and clones of KS and TX isolates of sPMV (Chowda-Reddy et al. 2018). Both PMV-NE and TX elicited mild mosaic symptoms on proso millet (*Panicum miliaceum*) whereas co-infection by PMV-NE+sPMV-KS elicited severe mosaic, yellowing, and stunting symptoms, compared with moderate symptoms by PMV-TX+sPMV-TX. The severe symptoms caused by PMV-NE or PMV-TX with sPMV-KS indicated that sPMV-KS was the main contributor to an efficient synergistic interaction. The genome sequences of sPMV-KS and sPMV-TX differ by 11 nucleotides with four non-synonymous and three synonymous changes in the coat protein ORF. These genomic differences between sPMV isolates provide the basis for the differential synergistic interaction with PMV.

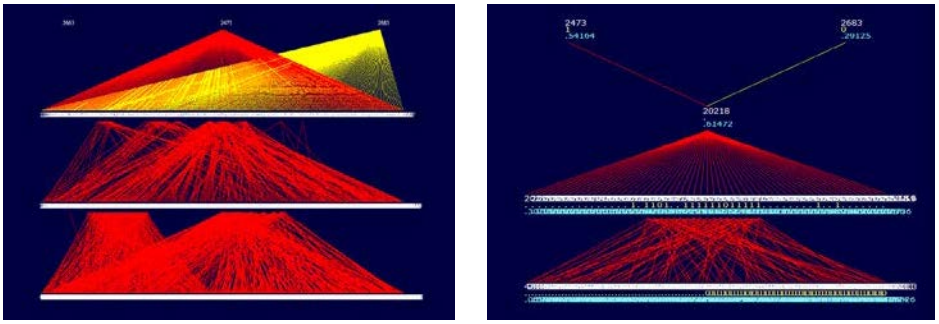


Fig. 1. Three cycles in the Pedigree of the Summer x Kanlow population. Left panel: female parents in red and male parents in yellow. Right panel: Original fullsib population with disease ratings and breeding values and successive generations bred by open-pollination.

References

Chowda-Reddy et al. 2018. Mapping synergistic interaction determinants between *Panicum mosaic virus* and *Satellite Panicum mosaic virus*. Intl Congress Plant Pathology. July 29-Aug 3, Boston, Mass.

Edmé et al. 2017. Realized and Expected Genetic Gains after Three Cycles of Breeding Switchgrass (*Panicum virgatum*) for Bioenergy. Manuscript under review.

Edmé et al. 2017. Genetic variation for resistance to viral mosaic in a Summer x Kanlow switchgrass breeding population based on the generalized linear mixed animal model. Switchgrass IV. Aug. 7-10, Lincoln, NE.

Muhle A. et al. 2017. Resistance of 'Liberty' switchgrass to rust and viral diseases. Switchgrass IV. Aug. 7-10, Lincoln, NE.

Funding statement: This research project is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. *DE-SC0016108.*

"Building on success in systems design of high yielding low-input energycanes for marginal lands" OR "Renewable Oil Generated with Ultra-productive Energycanes (ROGUE)"

Steve Long[1*] (slong@illinois.edu), Fredy Altpeter[2], Brian Baldwin[3], Li-Qing Chen[1], Don Ort[1], John Shanklin[4] and Vijay Singh[1]

[1]University of Illinois at Urbana-Champaign, [2]University of Florida, [3]Mississippi State University, [4]Brookhaven National Laboratory.

Project goals

Combining functional genomics and metabolic modeling, through DOE ARPA-E support, we identified and up-regulated key genes for TAG synthesis and transport and down regulated those involved in TAG catabolism to force accumulation of TAG in sugarcane within a modified oleosin coating in a Florida cultivar of sugarcane. Because oil contains far more combustion energy per unit mass than both structural and non-structural carbohydrates, we similarly identified and engineered increased photosynthetic capacity to provide the additional energy required to support lipid synthesis without loss of total biomass. In sugarcane we achieved 10% increases in total productivity and up to 8% tri-acyl glycerides (TAGs) in the leaves. This project now aims to further advance these technologies in Energycane and *Miscanthus x giganteus,* which is also something termed an energycane. Why have we chosen these two crops? They are: 1) Technologically farm- and factory-ready. That is, technologies for planting, growing, harvesting, transporting and crushing the juice, including extraction of oil, from their stems are well established (we have patented an efficient method for extracting this lipid from grass stems). 2) The most productive crops known, for the sub-tropical and temperate zones of the USA, respectively, in terms of their ability to convert sunlight energy into the chemical energy stored in the plant. 3) Energycane is highly productive on marginal soils in the Gulf States; Miscanthus can be grown on such soils from the north of the Gulf States into eastern Canada, so avoiding competition with food crop production and opening up many thousands of acres to under-utilized land to profitable use. 4) Both energycane and Miscanthus are closely related and share much of their genetic code with sorghum, allowing reference to this now detailed structural and functional genomic database. Having this template greatly speeds the pace at which genetic changes will be engineered. 5) Neither crop requires good soil or much fertilizer and they use water more efficiently than grain crops. 6) Both are easily contained under field conditions because they do not produce viable seeds under continental US field conditions, making them biosafe for transgenic modification. 7) As non-food crops, biolistically transformed, deregulation should not be required, facilitating quick commercial uptake. 8) Both crops are clonally propagated and since we are transforming elite US cultivars, the transformed plants will provide commercially viable material without the requirement of any further breeding.

At their exceptional levels of productivity, engineering energycanes and Miscanthus to accumulate vegetable oils (tri-acyl glycerides) to 20% of dry weight will provide ≥17x the amount of oil per acre than can currently be obtained from soybean, and 50x that of Camelina. With ~2,400,000 acres of marginal, under-utilized or abandoned cropland in the SE and E of the USA (Figure 1) this could provide sufficient diesel and jet fuel to satisfy >66% of the nation's current use.

These crops will be biolistically engineered with multi-gene constructs for both oil accumulation in their vegetative tissues and increased photosynthetic capacity to provide the energy for oil synthesis. This will be achieved by: 1) Continued development of our mathematical models allowing *in silico* optimization and engineering of photosynthesis and of oil synthesis/accumulation to guide *in vivo* genetic transformation. 2) Further development of our energycane and Miscanthus biolistic transformation systems to deliver higher efficiencies and speeds, including DNA editing. 3) Expansion of our Golden Gate libraries for the development of multi-gene constructs, which will be used in combination with re-transformation strategies. 4) Increased activity of tri-acyl glyceride (TAG) synthesis through up-regulation and addition of genes involved in synthesis, including up-regulation of transcription factors and editing, RNAi knock-downs of TAG catabolic pathways, and accumulation of modified protein coats to protect the oil globules. 5) Redirection of sucrose within the parenchyma storage cells, via modification of transporters (up- and down-regulation) to fuel TAG synthesis in addition to targeting of TAG accumulation to storage parenchyma. 6) Computationally guided engineering improvement of photosynthetic efficiency, including genetic modification of the crop canopy, leaf structure and metabolism to increase efficiency of light, water and nitrogen use. 7) Further development of our transient expression systems for testing and selecting best constructs for stable transformation. 8) Field testing in FL, MS and IL of resulting germplasm. Overall goal to increase photosynthetic efficiency by 50% and accumulate oil to 20% of shoot biomass, and provide farm-ready germplasm.

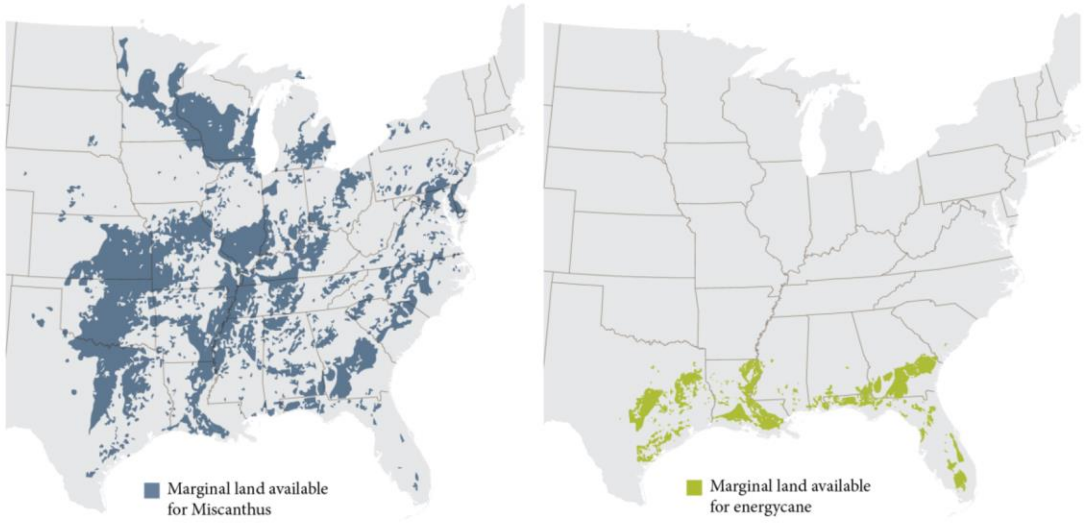


Figure 1. Marginal land availability for energycane and Miscanthus

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 established Agrobacterium-mediated transformation methods, we will develop a comprehensive toolset for genetic manipulation of *Lemnaceae*. We will establish CRISPR/Cas9 genome editing to complement our previously demonstrated 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. **Enabled by multiple *Lemnaceae* genome sequences and annotations and an extensive living collection of global accessions, we will use comparative genomics and systems network analysis to catalog transcription factors and promoters in 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. They have a much reduced morphology comprising leaf-like growing fronds, starch-filled resting fronds, and simple roots. While they are sometimes regarded as invasive due to their ability to rapidly cover the surface of freshwater ponds in the presence of nitrogen-rich agricultural runoff, duckweeds are native to all continents but Antarctica, and their extreme growth rate is ideal for vast biomass production. *Lemnaceae* in optimal conditions have an exponential growth rate that 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. Existing commercial strains of *Lemnaceae* have been optimized for protein production, environmental sensing, or wastewater remediation; we propose to redesign these strains for biofuel production. 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. Clonal propagation, limited seed set, and variable chromosome number are shared with sugarcane and *Miscanthus*, and many of the design principles and technologies we develop will have applications in other energy crops.

Under prior support from DOE, the Long Island Biofuels Alliance has achieved significant milestones in harnessing the potential of duckweed as a bioenergy crop. The Shanklin Lab completed a survey of fatty acid and TAG composition across 30 *Lemnaceae* species, while the Schwender lab has constructed a constraint-based model of carbon flux. A reliable and rapid protocol for stable transformation of *Lemna minor* was published by the Martienssen and Shanklin labs, along with gene-knockdown by artificial

miRNA. The Lam and Martienssen labs have contributed to the sequence and gene content of three Lemna genomes complete with chromosome structures, methylomes, small RNA transcriptomes, and structural variant analysis across accessions. Current genome assemblies have yielded validated orthologs in all the major lipid biosynthesis pathways. 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. By the end of the 4-year funding period, we will have dramatically increased oil content in fast growing aquatic plants, providing a novel, robust and highly productive source of biodiesel.

Publications

1. Yan, Yiheng, Jason Candreva, Hai Shi, Evan Ernst, Robert Martienssen, Jorg Schwender, and John Shanklin. 2013. "Survey of the Total Fatty Acid and Triacylglycerol Composition and Content of 30 Duckweed Species and Cloning of a Δ6-Desaturase Responsible for the Production of γ-Linolenic and Stearidonic Acids in Lemna Gibba." *BMC Plant Biology* 13 (1). BioMed Central:201.
2. Cantó-Pastor, A., A. Mollá-Morales, E. Ernst, W. Dahl, J. Zhai, Y. Yan, B. C. Meyers, J. Shanklin, and R. Martienssen. 2015. "Efficient Transformation and Artificial miRNA Gene Silencing in Lemna Minor." *Plant Biology* 17 Suppl 1 (January). Wiley Online Library:59–65.
3. Michael, Todd P., Douglas Bryant, Ryan Gutierrez, Nikolai Borisjuk, Philomena Chu, Hanzhong Zhang, Jing Xia, et al. 2017. "Comprehensive Definition of Genome Features in Spirodela Polyrhiza by High-Depth Physical Mapping and Short-Read DNA Sequencing Strategies." *The Plant Journal: For Cell and Molecular Biology* 89 (3). Wiley Online Library:617–35.

Funding for this project is provided by the DOE Office of Biological & Environmental Research (DE-SC0018244).

Characterizing the Defense Hierarchy of *Populus trichocarpa* and its Hybrids.

George Newcombe[1*] (georgen@uidaho.edu), Posy Busby[2], Wellington Muchero[3], Dale Pelletier[3], Mary Ridout[1], Devin Leopold[2], and Patricia Blair[3].

[1]University of Idaho, Moscow; [2]Oregon State University, Corvallis; [3]DOE National Laboratory, Oak Ridge.

https://www.researchgate.net/project/The-eriophyid-mite-Schizoempodium-mesophyllincola-in-Populus-trichocarpa

Project Goals: Plants host a diverse array of endophytes that can affect plant disease severity either by interacting directly with pathogens or by modulating the plant defense response. Plants also possess genes for pathogen resistance. Thirdly, competition among pathogens can reduce overall damage. Our ongoing project seeks to characterize the contingency rules of the defense hierarchy that includes all three mechanisms: genes for resistance to leaf rust and eriophyid mite pathogens, the mite/rust competition, and fungal and bacterial endophytes for which we first developed methodologies (Brown et al, 2018). Our model system involves the leaves of *Populus trichocarpa*. We are also seeking to determine whether the contingency rules of the foliar defense hierarchy apply to other tissues (i.e., fruits and seeds) of the host. We still need to complete seven manuscripts in addition to the two listed that are in press.

Abstract.

Genes for resistance to the leaf-bronzing, eriophyid mite, *Schizoempodium mesophyllincola*, were identified in both the GWAS common garden of *Populus trichocarpa* genotypes, and in *P. trichocarpa* x *P. deltoides* (TxD) hybrid pedigrees (Newcombe et al, 2018). In the former case, a cutin/suberin gene has been identified; in the latter an exapted gene inherited from *Populus deltoides* was associated with a 'sapsucker' or aphid resistance QTL. In this latter case resistance is exapted because the mite has a restricted, endemic distribution in the maritime Pacific Northwest, a region in which *Populus deltoides* is not native. Exapted resistance has been the norm in past studies of resistance to various pathogens of *Populus trichocarpa* and its hybrids, so this new finding fits the overall pattern. In terms of larger significance, exapted resistance poses a challenge to our understanding of the evolutionary basis of genes for resistance. The discovery of the cutin/suberin gene provides a possible answer to this evolutionary question. The cutin/suberin gene confers exapted resistance to *Sphaerulina musiva*, a non-native pathogen of the Pacific Northwest, but its evolutionary persistence could be due to adapted resistance to the mite that is native to the region. Genetic resistance is, however, only part of the resistance hierarchy that our research is revealing. We have also found that rust and mite pathogens can directly compete for leaf tissue and that endophytes can antagonize the rust fungus. These three mechanisms form a hierarchy: first genetic resistance, then competition between mites and rust, and lastly endophyte-mediated resistance. What this means is that endophytes mediate resistance only when genetic resistance and competition are absent. In spite of these advances in understanding of the defense hierarchy in

poplar leaves, we also have uncovered new complexity. For example, we have found a fungus that may attack and regulate the mite. Secondly, we have found strong geographic structuring of foliar endophytes that has not yet been integrated into our understanding of the hierarchy. Finally, we have discovered that hierarchies may be tissue-specific. Our exploratory research with fruits (capsules) and seeds of *Populus trichocarpa* hints at a hierarchy unlike that of poplar leaves. This reproductive hierarchy involves a new disease (*Marssonina* capsule blight), a previously unknown seedling pathogen of poplar (a pathovar of *Pseudomonas syringae*), and another bacterium that antagonizes *Pseudomonas syringae*. The main distinguishing feature is that pathogen competition has not been observed between *Marssonina* and *Pseudomonas*. In its place is facilitation of one pathogen by another.

References

Brown, PS, Leopold, D, and Busby PE. 2018. Protocols for investigating the foliar mycobiome using high throughput DNA sequencing. Methods in Molecular Biology: Plant Pathogenic Fungi and Oomycetes. Springer Nature. In press.

Newcombe, George, Muchero, Wellington, and Posy Busby. 2018. Exapted resistance to an eriophyid mite in a hybrid pedigree of *Populus*. PLOS ONE: accepted pending suitable revision.

Funding statement. We thank DOE, Office of Biological & Environmental Research. Federal Award Identification Number: DE-SC0014547.

Comparative Genomics, Transcriptomic Analysis, and Ecophysiology of Individual Environmental Methane-oxidizing Syntrophic Consortia

Connor T. Skennerton,[1] Grayson Chadwick,[1] Hang Yu[1], Rex Malmstrom,[2] Tanya Woyke,[2] Roland Hatzenpichler, [1] Robert Hettich,[3] Christof Meile,[4] Mark Ellisman,[5] Chris Kempes,[6] Gene W. Tyson,[7] 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]Oak Ridge National Lab, US Department of Energy, Tennessee; [4]University of Georgia, Athens, Georgia; [5]University of California, San Diego, California; [6]Santa Fe Institute, Santa Fe, New Mexico; [7]University of Queensland, St. Lucia, Queensland, Australia

http://orphanlab.caltech.edu

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

Our goal is to utilize multiple analytical strategies to substantially expand our understanding of the key microorganisms, metabolic strategies, and interspecies relationships involved in the formation and oxidation of methane in the environment. Our research incorporates novel meta'omics strategies, state-of-the-art microscopy, and stable isotope geochemistry spanning spatial scales of nanometers to millimeters. Applied in tandem, these complementary methodologies have the potential to provide a new and more holistic 'eco-systems level' understanding of the factors which regulate methane cycling in anoxic sedimentary ecosystems. These analyses will be combined with characterizations of field samples and geochemically-characterized laboratory microcosms in order to further our understanding of methane cycling in anoxic habitats.

The anaerobic oxidation of methane (AOM) is a major contributor to the global methane cycle. At sites of methane seepage, AOM is catalyzed by diverse syntrophic partnerships between anaerobic methanotrophic (ANME) archaea and sulfate-reducing deltaproteobacteria (SRB), frequently found in well-structured multi-celled consortia. These organisms are not yet available in pure culture, however environmental studies have identified multiple genera of ANME and deltaproteobacterial partners often co-existing in the same sediment environment. The physiological differences and potential for niche differentiation between distinct ANME groups is poorly understood. Our earlier work using conventional environmental metagenomics resulted in a subset of reassembled genomes from these groups, enhancing our understanding of ANME and SRB metabolism. Much of the information on fine scale population structure and information regarding specific syntrophic pairings is lost however when reconstructing genomes from bulk sediment metagenomes. To address these more targeted ecologically relevant questions, we developed assays that combine fluorescent activated cell (consortia) sorting

(FACS) with bioorthogonal noncanonical amino acid tagging (BONCAT) of anabolically active AOM consortia in sediment and sequenced the genomes of multiple individual consortia of ANME-SRB recovered from the same sample. Using this unique sequence dataset from the BONCAT sorted aggregates, we have examined fundamental questions regarding the clonal nature of within consortia archaea and bacteria and the broader microheterogeneity that exists between the genomes of closely related ANMEs recovered from different aggregates. All ANME lineages across the dataset were observed to harbor large multiheme cytochromes that appear to be a hallmark of these organisms, and distinct from closely related methanogen relatives. Examination of potential interspecies interactions including metabolic complementation between the syntrophic partners is ongoing. These single aggregate genomes have also been used in concert with environmental metatranscriptomics of sediment microcosms maintained under conditions supporting syntrophic sulfate-coupled AOM as well as active archaeal methane oxidation with AQDS in the absence of an active bacterial partner. This approach revealed distinct differences between the transcriptional responses of individual aggregates containing distinct ANME types in the same sediment under different conditions. These observations are now being combined with microscopy and nanoSIMS-based analyses of different spatial arrangements between ANME and SRB subgroups within aggregates in order to better understand variations in ecophysiology between co-occurring ANME groups and the interplay between syntrophic metabolism and microbial community structure.

References

1. Skennerton, C. T. *et al.* Methane-Fueled Syntrophy through Extracellular Electron Transfer: Uncovering the Genomic Traits Conserved within Diverse Bacterial Partners of Anaerobic Methanotrophic Archaea. *MBio* **8,** (2017).
2. Scheller, S., *et al*. Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. *Science* **351,** 703–707 (2016).
3. Skennerton, C. T. *et al*. A genomic atlas of anaerobic methanotrophic archaea. *In preparation.*

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-FOA-0001458]

Defining the Functional Genome Associated with Enhanced Water and Nitrogen Use Efficiency in Bioenergy Sorghum

Maxwell Braud,[1] Andrea L. Eveland,[1*] (aeveland@danforthcenter.org), Rajiv Parvathaneni,[1] Indrajit Kumar,[1] Edoardo Bertolini,[1] Brendon Morrison,[1] Xiaoping Li,[1] Thomas P. Brutnell,[1] and Daniel Schachtman[2]

[1] Donald Danforth Plant Science Center, St. Louis, MO; [2] University of Nebraska, Lincoln, NB

https://sorghumsysbio.org

Project Goals:

- **Establish controlled environment conditions for repeatable testing of low water and low nitrogen regimes on diverse sorghum accessions.**
- **Generate RNA-seq-based transcript profiles and genome-wide accessible chromatin maps for developing leaf and root tissues across genotypes in response to water and nitrogen stresses.**
- **Integrate transcriptional and epigenomic signatures with existing genetics and genomics resources to prioritize genetic loci for functional validation.**

The overall goal of this work is to establish a baseline functional genomics resource that can be leveraged for investigating effects of low water and nitrogen inputs on growth and development of three diverse sorghum accessions; BTx623 (reference line, NAM parent, EMS population), Grassl (bioenergy, NAM parent, EMS population) and PI_510757 (high NUE, WUE and large biomass, NAM parent). These lines are representatives of three diverse races of sorghum (Kafir, Caudatum and Durra, respectively) and thus will serve as founders for deeper explorations of the genetic diversity present in sorghum. A series of chamber-based, controlled environment experiments are being carried out to interrogate gene regulation in these three genotypes at various stages of development; from seedling establishment to the floral transition. Transcriptome profiles analyzed across emerging leaf and root samples were aligned with respective chromatin accessibility maps for these tissues generated using ATAC-seq. These datasets will be integrated to identify transcriptional signatures that are associated with enhanced growth under low water and N conditions. An integrated resource that utilizes the same founder lines (i.e. three diverse sorghum lines), tissues and developmental time series will **i)** improve resolution of our gene regulatory networks, **ii)** provide a foundation for a sorghum pan-genome ENCODE-like project to identify regulatory regions across sorghum races and **iii)** identify targets for engineering enhanced biomass, NUE and drought tolerance.

Co-Optimization of Multicomponent Enzyme Mixtures and Saccharification Conditions

Carolina A. Barcelos[1]* (cbarcelos@lbl.gov), Taya G. Feldman,[1] Joel M. Guenther,[1] Blake A. Simmons,[1,2] and **Kenneth L. Sale**[1,3]

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA; and [3]Biomass Science and Conversion Technologies Department, Sandia National Laboratories, Livermore, CA

http://www.jbei.org

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

For the efficient production of sugars for platform chemicals and biofuels, lignocellulosic biomass is pretreated and subsequently hydrolyzed with a combination of enzymes to produce glucose and xylose. Although ionic liquids pretreat biomass very efficiently, the enzymatic hydrolysis of ionic liquid pretreated biomass has not been optimized. Moreover, enzymes for hydrolysis of pretreated biomass remain a major cost in the production of chemicals and fuels from lignocellulosic biomass. The goals of this research are to use the JBEI's jSALSA robotics platform to 1) determine the optimal operating temperature and pH of glycosyl hydrolases from thermophiles, 2) optimize synthetic multi-component enzyme mixtures for hydrolysis of several ionic liquid/substrate combinations.

Funding statement.

This work was part of the DOE Joint BioEnergy Institute (http:// www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

A Comparison of Lignocellulose Solubilization by Cellulolytic Monocultures and a Mixed Enrichment Yields Surprising Results

Evert K. Holwerda[1,3*] (evert.k.holwerda.th@dartmouth.edu), Robert S. Worthen[1,3], Sean J-L. Murphy[1,3], Xiaoyu Liang[1,3], Jason M. Whitham[3], James G. Elkins[3], Steven D. Brown[2], Lee R. Lynd[1,3], **Gerald A. Tuskan[3]**

[1]Dartmouth College, Hanover, New Hampshire, [2]LanzaTech, Skokie, Illinois, [3]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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

The question "What biocatalysts are most effective at mediating biomass deconstruction?" is of considerable fundamental interest and is foundational for developing processes for biological conversion of cellulosic biomass. Yet controlled comparative studies are scarce. Recently, substantial differences in solubilization capability have been demonstrated with respect to the ability of various biocatalysts to solubilize unpretreated switchgrass,[1] 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.[2]

Mixed lignocellulose-fermenting enrichments (aka microbiomes, microbiota) have repeatedly been shown to contain a diversity of cellulolytic microbes and cellulase genes. It is generally thought that lignocellulose solubilization by mixed enrichments is enabled by synergistic interactions and more effective than solubilization by pure cultures. However, few, if any, studies have quantitatively compared the relative rates and extents of solubilization by pure cultures and mixed enrichments. We sought to provide such a comparison in this study.

In an initial experiment carried out in batch culture, we compared solubilization of 30 g/L mid-season harvested switchgrass by *Clostridium thermocellum* and an inoculum from a mixed enrichment (thermophilic microbiome) maintained for over 2 years on this feedstock. Both cultures were incubated at 55°C under anaerobic conditions. Based on

16S rRNA and metagenomic analysis, the mixed enrichment inoculum appeared to have substantial diversity with respect to both the microorganisms and glycosyl hydrolase genes present. The change in diversity during solubilization was followed using relative abundance for 16S rRNA genes. Total carbohydrate solubilization (TCS) was greater for *C. thermocellum* than the mixed enrichment for early time points, and essentially equal for later time points.

In a second experiment we cultivated *C. thermocellum* and mixed enrichments on the same switchgrass feedstock in intermittently-fed continuous cultures maintained at various residence times. The extrapolated concentration of inaccessible carbohydrate was the same for *C. thermocellum* and the mixed enrichment, which also agreed well with results from batch culture. For both the pure and mixed culture, the rate of TCS was first order in accessible substrate. To our surprise, the first order rate constant was over 2-fold higher for the pure culture compared to the mixed culture.

Our results contradict the understanding that mixed enrichment cultures achieve more complete and more rapid deconstruction of lignocellulosic biomass, but will need further confirmation.

1 J. M. D. Paye, A. Guseva, S. K. Hammer, E. Gjersing, M. F. Davis, B. H. Davison, J. Olstad, B. S. Donohoe, T. Y. Nguyen, C. E. Wyman, S. Pattathil, M. G. Hahn and L. R. Lynd, *Biotechnol. Biofuels*, 2016, **9**, 8.

2 L. R. Lynd, A. M. Guss, M. E. Himmel, D. Beri, C. Herring, E. K. Holwerda, S. J.-L. Murphy, D. G. Olson, J. Paye, T. Rydzak, X. Shao, L. Tian and R. Worthen, in *Industrial Biotechnology*, John Wiley & Sons, Ltd, 2016, pp. 365–394.

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.

CELF/CBP for Enhanced Utilization of Polysaccharides and Lignin

Priyanka Singh,[1,2,3] Charles Cai,[2,3*](ccai@engr.ucr.edu), Rajeev Kumar,[2,3] Charles E. Wyman[1,2,3], and **Gerald Tuskan**[3]

[1]Department of Chemical and Environmental Engineering, University of California, Riverside, California; [2]Center for Environmental Research and Technology (CE-CERT), University of California, Riverside, California; [3]The 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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Cosolvent Enhanced Lignocellosic Fractionation (CELF) pretreatment of cellulosic biomass produces hydrolysate that is rich in hemicellulose sugars and lignin and leaves solids containing mostly glucan. Consolidated bioprocessing (CBP) can then very effectively deconstruct the glucan and potentially residual hemicellulose in these solids without adding enzymes. An integrated CELF-CBP process therefore offers the potential to achieve low cost bioconversion of lignocellulosic biomass through effective utilization of the three major biomass fractions. Furthermore, efficient lignin fractionation by CELF pretreatment reduces lignin toxicity to *C. thermocellum* and provides a clean lignin stream for further upgrading in addition to production of fermentable sugars with high yields. Accordingly, our goal in this project is to maximize total sugar release from hemicellulose and cellulose via the CELF-CBP combination and CELF lignin recovery from CBP residuals from both switchgrass and poplar. The project includes developing and applying lignin fractionation and characterization techniques to define promising options for conversion of CELF lignin into value-added products. Understanding sugar release by the CELF-CBP combination and lignin features that are amenable to conversion to valuable products over a wide range of pretreatment and post-fermentation conditions can provide valuable insights into pathways that can enhance value generation from biomass conversion. This poster will include preliminary results from co-optimization of CELF-CBP coupled with lignin fractionation and characterization.

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.

Towards Understanding Plant Cell Wall Structure and Properties During Microbial Deconstruction - From Chemical Bonds to Wall Architecture

Qi Xu[1], Roman Brunecky[1], Daehwan Chung[1], Bryon S. Donohoe[1], Yining Zeng[1], Breeana R. Urbanowicz[2], Maria Peña[2], Jan Westpheling[3], Debra Mohnen[2], and Michael E. Himmel[1], Yannick J. Bomble[1*] (yannick.bomble@nrel.gov), **Gerald A. Tuskan**[4]

[1]National Renewable Energy Laboratory, Golden, CO; [2]Complex Carbohydrate Research Center, Athens, GA, [3]University of Georgia, Athens, GA, [4]The 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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Many aspects of plant cell wall deconstruction by cellulolytic microbes remain unknown. Specifically, cellulolytic bacteria that directly bind to substrates show modalities of deconstruction that are important in the: 1) understanding of cell wall deconstruction and 2) discovery other efficient microbes to enable the complete conversion of biomass to sugars. Closing this knowledge gap will inform consolidated bioprocessing (CBP) applications wherein the microbe both deconstructs and converts biomass to advanced biofuels and biochemicals. For microorganisms which bind to plant biomass, it is known that the proximity of the microbe to the biomass surface and the titer of the enzymes secreted contribute to greater cell wall solubilization than the enzymes alone. Interestingly, the molecular mechanisms governing these complex systems are not yet understood. Our goals are to: 1) obtain in-depth understanding of the barriers that confront biomass deconstruction by biocatalysts, 2) gain the fundamental knowledge needed to inform deconstruction processes by consolidated bioprocessing microorganisms and 3) develop a multi-length scale understanding of plant cell wall deconstruction by combining integrated experimental and computational techniques. This research will increase understanding of the emergent mechanisms by which biocatalysts interact with and utilize biomass; as well as the roles that biomass polymers, cell wall architectural units/surfaces, and tissues play in conversion processes. Ultimately, we aim to gain a holistic, predictive view of biomass deconstruction by studying the physical, chemical, and biochemical phenomena occurring during natural deconstruction.

Understanding biomass synthesis and structure, as well as their relationships to deconstruction, is also vital for reaching these goals. We are working to provide an in-depth understanding of the biochemical and biophysical mechanisms that promote (or inhibit) the deconstruction of plant cell walls during CBP. Here we present our most recent work on understanding cellulolytic synergism in secretomes from thermophilic microbes during deconstruction of biomass and the importance of studying the relationship between microbial binding strength and extent of solubilization focusing on *Caldicellulosiruptor bescii and Clostridium thermocellum*. Understanding the functional diversity of enzymes in the *C. bescii* exoproteome and how inter-molecular synergy between them confers *C. bescii* with its high cellulolytic activity is an important endeavor to potentially confer cellulolytic capability to non-cellulolytic microbes considered for high titer production of biochemicals and biofuels. We found that the combination of three or four of the most highly expressed enzymes in the *C. bescii* exoproteome exhibits such synergistic activity. For example, some discrete combinations of these enzymes mimic and even improve upon the activity of the exoproteome, even though some of the enzymes lack significant activity on their own. Regarding our work with *C. thermocellum*, we have found that modifying the cellulolytic machinery of this microbe reduces its cellulolytic activity but also the strength of binding to biomass significantly, leading to a change in its deconstruction mechanism.

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

Post-harvest Induction of Hyperthermophilic Cellulases Reduces Recalcitrance in Poplar

Charleson Poovaiah (crpoovai@syr.edu), Yao Xiao, Lilibeth Northern, and **Heather D. Coleman***

Syracuse University, Syracuse, NY

http://colemanlab.syr.edu/index.html

Project Goals: The overall goals of the project are to verify in poplar In Plant Activation (INPACT) technology[1], which enables inducible expression of genes and accumulation of proteins at very high levels *in planta*, and to evaluate the ability of cellulases produced through this technology to hydrolyze cellulose to glucose for fermentation. In addition, we aim to identify transgenic lines that have improved processability due to the accumulation of cellulases, either through traditional expression or through the use of the INPACT system.

Plant cell walls provide a vast, untapped source of sugars in the form of cellulose available for fermentation into ethanol and other biofuels. There remain several challenges in the attempt to developing economically viable cellulosic ethanol based on biochemical conversion, including the cost of cellulolytic enzymes and thermochemical pretreatments to reduce recalcitrance of the cell wall. One potential approach to reduce enzyme cost is the production of the enzymes within the feedstock itself. This would require a system that prevents the degradation of the cell wall until the desired time.

Here we verify in poplar a transgenic technology (In Plant Activation Technology – INPACT) that allows for the controlled production and accumulation of enzymes within the plant. This system is also being used to drive the expression of cellulases in poplar, with the transgenic lines currently being grown in the greenhouse.

When paired with a cellulose synthesis promoter, driving gene expression in secondary cell wall of vascular tissue, INPACT drives inducible expression of the GUS reporter gene. However, the level of expression does not reach 'extreme' levels at this time. It is likely that the improvement of the system, as carried out previously[2] will be necessary to drive high-level accumulation of enzymes.

While carrying out analysis of these INPACT lines, we have also assessed traditional overexpression lines with the Cauliflower Mosaic Virus 35S promoter driving the expression of cellulase genes. In particular, we have assessed a line expressing a hyperthermophilic *Thermotoga neapolitana* endoglucanase (TnCelB). Poplar-derived TnCelB retains high activity to substrates at 100°C, and in all transgenic lines, biomass was significantly increased. While there were some alterations in the plant cell wall composition, the line with the highest TnCelB activity showed enhanced glucan saccharification efficiency over WT with and without a post-harvest heat treatment. The same transgenic line showed gains in saccharification over WT without a traditional pre-treatment, solely from a heat treatment of 100°C immediately following harvest. This set of plants provided a feedstock substrate that yields increased glucose without a chemical pretreatment, solely through the activation of a hyperthermophilic cellulase.

Overall, this project has yielded evidence that the INPACT system is viable in poplar, though it will require additional fine-tuning. In addition, the control lines from the project, which are traditional overexpression lines, have yielded significant results in terms of improved glucose release without a chemical pretreatment.

References

1. Dugdale, B. *et al*. (2013) In plant activation: an inducible, hyperexpression platform for recombinant protein production in plants. The Plant Cell 25: 2429-2443.
2. Kinkema, M. *et al*. (2014) An improved chemically inducible gene switch that functions in the monocotyledonous plant sugar cane. Plant Molecular Biology 84:443-454.

Funding statement: This material is based upon work supported by the Department of Energy under Award Number DE-SC0010411.

Comprehensive Functional Characterization of the Glycoside Hydrolase Family 3 Enzymes from *Cellvibrio japonicus* Reveals Unique Metabolic Roles in Biomass Saccharification

Cassandra E. Nelson,[1] Mohamed A. Attia, [2] Artur Rogowski,[3] Carl Morland,[3] Harry Brumer,[2] Estela C. Monge, [1]* (emonge1@umbc.edu), and **Jeffrey G. Gardner**[1]

[1]University of Maryland - Baltimore County, Baltimore, MD; [2]Univeristy of British Columbia, Vancouver, Canada; and [3]Newcastle University, Newcastle upon Tyne, UK

Project Goals: This project will generate a predictive systems-level model of lignocellulose deconstruction for the saprophytic soil bacterium *Cellvibrio japonicus*. In addition, this project will characterize novel Carbohydrate Active Enzymes and probe their utility for biotechnology applications, such as renewable fuels and commodity chemicals.

Lignocellulose degradation is central to the carbon cycle and renewable biotechnologies. The xyloglucan (XyG), β(1→3)/β(1→4) mixed-linkage glucan (MLG), and β(1→3) glucan components of lignocellulose represent significant carbohydrate energy sources for saprophytic microorganisms. The bacterium *Cellvibrio japonicus* has a robust capacity for plant polysaccharide degradation, due to a genome encoding a large contingent of Carbohydrate-Active Enzymes (CAZymes), many of whose specific functions remain unknown. Using a comprehensive genetic and biochemical approach we have delineated the physiological roles of the four *C. japonicus* Glycoside Hydrolase Family 3 (GH3) members on diverse β-glucans. Despite high protein sequence similarity and partially overlapping activity profiles on disaccharides, these β-glucosidases are not functionally equivalent. Bgl3A has a major role in MLG and sophorose utilization, and supports β(1→3) glucan utilization, while Bgl3B underpins cellulose utilization and supports MLG utilization. Bgl3C drives β(1→3) glucan utilization. Finally, Bgl3D is the crucial β-glucosidase for XyG utilization. This study not only sheds the light on the metabolic machinery of *C. japonicus*, but also expands the repertoire of characterized CAZymes for future deployment in biotechnological applications. In particular, the precise functional analysis provided here serves as a reference for informed bioinformatics on the genomes of other *Cellvibrio* and related species.

Publications

1. Gardner JG**.** 2016. Polysaccharide degradation systems of the saprophytic bacterium *Cellvibrio japonicus*. *World Journal of Microbiology and Biotechnology*. 32:121-32.

2. Nelson CE, Beri NR, and Gardner JG**.** 2016. Custom fabrication of biomass containment devices using 3-D printing enables bacterial growth analyses with complex insoluble substrates. *Journal of Microbiological Methods*. 130:136-43.

3. 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.

Spatial Connectomics to Identify Agents Relevant to Lignocellulose Deconstruction in Fungi

Jiwei Zhang[1], Gerald N. Presley[1], Kevin Silverstein[1], Melania Figueroa[1], Katarina Sweeney[1], Kenneth E. Hammel[2], Christopher J. Hunt[2], Ellen A. Panisko[3], and **Jonathan S. Schilling**[1]

[1]University of Minnesota, Saint Paul, [2]U.S. Forest Products Laboratory (FPL); [3]Pacific Northwest National Laboratory

URL: http://schillinglab.cfans.umn.edu

Project Goals: Our goal is to discover which genes are differentially up-regulated across the mycelia of brown rot wood-degrading fungi *in planta*, particularly at the leading edge of wood decomposition. These unique fungi accomplish what we have difficulty achieving – energy from plant biomass. To do this, brown rot fungi apparently couple an oxidative pretreatment step with enzymatic saccharification in discrete space, via partitioning of reactions. We have previously shown evidence that these are governed by differential expression, but with genes putative, transport out of hyphae unclear, and the secretome poorly studied, historically, a comprehensive approach is needed that can also limit ('winnow') data sets from powerful global analytical tools to focus on the genes and pathways that matter. To address this, we are comparing global expression profiles among mycelial regions and matching what we see to what they do, in terms of changing wood physiochemistry. To map a 'connectome' in a wood-fungal interaction, specifically, we must also match gene expression patterns with the extracellular secretome and with physiochemical wood modifications.

Objective 1 – **Zone localization:** Use the wood wafer design to resolve a depolymerization zone in *P. placenta* near the mycelial leading edge, and optimize RNA extraction for thin-sectioning.

Objective 2 - Fungal connectomics: Co-localize gene expression with the secretome and with relevant physiochemical modifications made to the wood, e.g. hemicellulose loss, porosity changes.

Objective 3 - Clade comparisons: Compare key zones among brown rot clades, in context of white rot same-clade ancestors, to target universal 'brown rot' genes and candidates for bioprocessing.

Abstract:
Some fungi are uniquely able to deconstruct lignocellulose, and their mechanisms have potential biofuels applications. A key hindrance to harnessing these 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 analyses of these fungi from artificial media or from colonized wood ground en masse fail to distinguish expression of key gene products 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*. A similar approach, 'connectomics,' has been used to map the human nervous system, and its application here is timely. First, wood-degrading fungal genomes are an emerging resource, particularly brown rot functional types. Second, 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.

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. To do this, we need to compare global expression profiles among mycelial regions. To map a 'connectome' in a wood-fungal interaction, specifically, we must also match gene expression patterns with the extracellular secretome and with physiochemical wood modifications. Given this potential for substrate-fungus feedback, we will cross-check 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 the USDA Forest Products Laboratory and the Pacific Northwest National Laboratory.

References:

Zhang J., Schilling, J.S. (2017) Role of carbon source in the shift from oxidative to hydrolytic wood decomposition by Postia placenta. Fungal Genetics & Biology 106: 1-8.

Presley, G.N., Zhang, J., Schilling, J.S.* (online first) A genomics-informed study of oxalate and cellulase regulation by brown rot wood-degrading fungi. Fungal Genetics & Biology doi.org/10.1016/j.fgb.2016.08.004.

Zhang, J., Presley, G.N., Hammel, K.E., Menke, J.R., Hu, D., Ryu, J.S., Orr, G., Schilling, J.S. *(2016) Localizing gene regulation reveals a staggered wood decay mechanism for the brown rot fungus *Postia placenta*. Proceedings of the National Academy of Sciences 113: 10968-10973.

Kaffenberger, J.T. and Schilling, J.S. (2015) Comparing lignocellulose physiochemistry after decomposition by brown rot fungi with distinct evolutionary origins. Environmental Microbiology 17(12):4885-97.

Schilling, J.S., Kaffenberger, J.T., Liew, F.J., Song, Z. (2015) Signature wood modifications that reveal decomposer community history. PLoS ONE 10(3): e0120679. doi:10.1371/journal.pone.0120679.

Schilling, J.S., Duncan, S.M., Presley, G.N., Filley, T.R., Jurgens, J.A, and Blanchette, R.A. (2013) Colocalizing incipient reactions in wood degraded by the brown rot fungus *Postia placenta*. International Biodeterioration and Biodegradation 83: 56-62.

Schilling, J.S., Ai, J., Blanchette, R.A., Duncan, S.M., Filley, T.R., and Tschirner, U.W. (2012) Lignocellulose modifications by brown rot fungi and their effects, as pretreatments, on cellulolysis. Bioresource Technology 116: 147-154.

Funding statement

This work was supported by the U.S. Department of Energy Office of Science (Early Career Grant DE-SC0004012 from the Office of Biological and Ecological Research (BER) to J.S.S.; and BER Grant DE-SC0012742). Confocal microscopy is funded by user Facility grant #48607 at the Environmental Molecular Sciences Laboratory of Pacific Northwest National Laboratory (to J.S.S.).

Insights into Hemicellulose-Cellulose Interactions from Thermochemical Pretreatment of Model Composite Materials

Riddhi Shah,[1] Sai Venkatesh Pingali,[2] Barbara R. Evans,[2] Daisuke Sawada,[2] Hugh O'Neill,[1,2] and **Brian H. Davison**[2*] (davisonbh@ornl.gov)

[1]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville; [2]Oak Ridge National Laboratory, Oak Ridge, Tennessee

http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

Thermochemical pretreatment of biomass results in structural reorganization of the various cell wall polymers during cell wall deconstruction. However, the underlying molecular level interactions that occur during these processes are poorly understood.[1] Model systems of polymers that mimic cell wall structures can potentially help in understanding polymer-polymer interactions that occur during pretreatment.[2] In this study, model hemicellulose-cellulose composites were prepared by synthesizing bacterial cellulose in presence of xyloglucan (XG) or glucomannan (GM) dissolved in the growth media. Small-angle neutron scattering (SANS) was used to study the nanoscale structural changes in the composites as a result of dilute acid pretreatment (DAP). By growing deuterium-labeled cellulose[3] in the presence of hydrogenated hemicellulose it was possible to deconvolute the scattering signatures of the two components. Changes in crystallite size, crystallinity, and glucan chain packing for the native and DAP-treated cellulose and composites were also studied using X-ray diffraction (XRD).

At the nanoscale, DAP-treated cellulose showed a collapse in structure, as indicated by a decrease in the radius of gyration (R_g) from 250 Å to 130 Å that was interpreted as expulsion of water from the space between microfibrils resulting in formation of a tightly packed macrofibril. The change of power law exponent (α) from 2.71 ± 0.26 to 2.23 ± 0.048 showed that the pretreated cellulose microfibril network was less entangled than that of native cellulose. In the case of the xyloglucan-cellulose (XGC) composite, the structure of the cellulose network was relatively unchanged after

pretreatment. Interestingly, in glucomannan-cellulose (GMC) composites, the changes that occurred due to DAP were similar to native cellulose except that α was increased (α = 2.82 ± 0.06) after pretreatment, indicating that the microfibrillar network became more entangled.

XRD analysis showed no significant change in the crystallinity (69.4 ± 3.9%) and peak positions (14.47°, 16.86°, and 22.72°) after pretreatment of native cellulose, but an increase in the crystallite size along the (010) plane was evident. The crystallinity of XGC was lower than native cellulose (36.59 ± 2.30 %) but increased to 53.33 ± 1.72% after DAP. There was a change in the peak positions in XGC compared to native cellulose that was interpreted as an increase in the Iβ content in the cellulose. Unlike native cellulose, the crystallite size of the XGC was unchanged after pretreatment. Similar to the SANS data, the structural changes observed in GMC as a result of pretreatment were very similar to native cellulose.

Our results show significant differences between XG and GM interactions with the growing cellulose network. We can propose that XG interacts directly with the cellulose microfibrils as they are formed. This is supported by increased crystalline Iβ content in the cellulose microfibrils and the size of the microfibril remaining unchanged after DAP. On the other hand, GM most likely interacts at the surface of the macrofibrils because it does not change the crystalline form of the cellulose and the macrofibrils collapse as a result of DAP, similar to native cellulose. This study provides insight into cellulose-hemicellulose interactions that may help in studies related to understanding accessibility to enzymes for biofuels, developing cellulosic bioproducts and engineering plants with increased digestibility.

References

1. Cheng, G., Zhang, X., Simmons, B., and Singh, S., "Theory, Practice and Prospects of X-ray and Neutron Scattering for Lignocellulosic Biomass Characterization: Towards Understanding Biomass Pretreatment," *Energy & Environmental Science* **8**(2), 436-455 (2015). doi:10.1039/C4EE03147D
2. Penttilä, P. A., Imai, T., and Sugiyama, J., "Fibrillar Assembly of Bacterial Cellulose in the Presence of Wood-Based Hemicelluloses," *International Journal of Biological Macromolecules* **102**(Supplement C), 111-118 (2017). doi:https://doi.org/10.1016/j.ijbiomac.2017.04.010
3. O'Neill, H., Shah, R., Evans, B. R., He, J. H., Pingali, S. V., Chundawat, S. P. S., et al., "Chapter 6 - Production of Bacterial Cellulose with Controlled Deuterium-Hydrogen Substitution for Neutron Scattering Studies," **565**, pp. 123-146 in *Isotope Labeling of Biomolecules - Labeling Methods* (2015). doi:10.1016/bs.mie.2015.08.031

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.

Molecular Interactions Enhanced by Co-Solvent Use for Effective Biomass Pretreatment

Sai Venkatesh Pingali,[1] Loukas Petridis,[1], Riddhi Shah,[1] Hugh M. O'Neill (oneillhm@ornl.gov),[1*] Micholas Dean Smith,[1] Yunqiao Pu,[1] Charles Cai,[2] Charles E. Wyman,[2] Volker Urban,[1] Arthur J. Ragauskas,[1] Jeremy C. Smith,[1] and **Brian H. Davison**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee; [2]Bioenergy Science Center, University of California-Riverside, Riverside, California

http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

The deconstruction of cellulose is crucial for the chemical conversion of lignocellulosic biomass into fuel and precursors for bio-products. Biomass is composed of cellulose fibers embedded in a matrix of hemicellulose and lignin, a hydrophobic polymer. The mixture of hydrophilic and hydrophobic properties for these plant polymers have triggered a concerted effort to use a co-solvent such as organic/water co-solvents. For example, the Co-solvent Enhanced Lignocellulosic Fractionation (CELF) treatment method augments traditional aqueous-based pretreatment by employing tetrahydrofuran (THF) as a co-solvent to water.[1] As developed by our collaborators in the Wyman group at the University of California-Riverside, current understanding of CELF is derived from experiments examining bulk chemical composition of biomass that show CELF significantly improves biomass delignification and deconstruction of sugar polymers compared to pure water. But, the fundamental understanding of how the

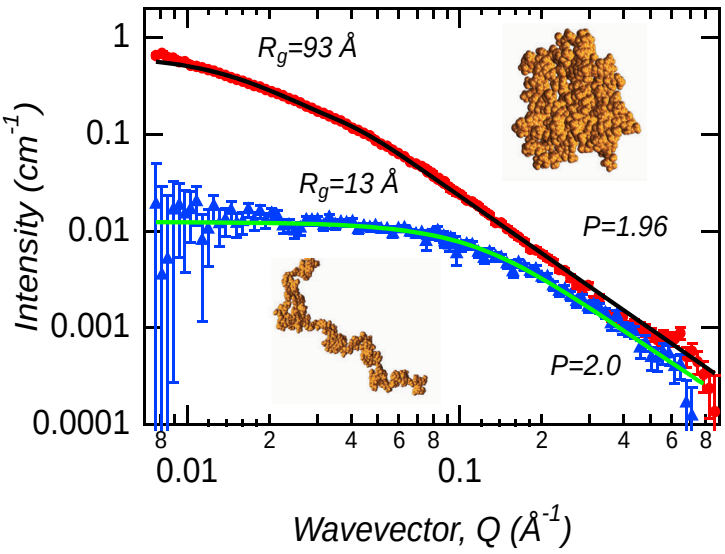


Figure 1. SANS data of lignin in THF only (red dots and black solid line) and water-THF co-solvent (blue filled up-triangle and green solid line). The solid lines are SANS data fits to determine the particle size (as R_g) and polymer bulk morphology (as P). The conformational models obtained by MD simulations for water only (top) and water-THF co-solvent (bottom) systems.

co-solvents can lead to such a dramatic increase in pretreatment efficiency is lacking. Here, we report molecular dynamic simulations (MD) and small-angle neutron scattering (SANS) results that lend insights to how a co-solvent provides superior environment. Recent molecular dynamics (MD) simulations have revealed two potential molecular-scale mechanisms explaining CELF process's high efficiency.[2-3] Lignin in pure water adopts collapsed conformations and binds to cellulose. In contrast, lignin forms coil conformations and detached to cellulose in THF-water co-solvent. Furthermore, water and THF could phase separate on the surface of cellulose promoting enhanced cellulose deconstruction. Preliminary SANS studies show lignin forms smaller particles in co-solvent compared to pure solvent conditions, consistent with MD results. In fact, this is the first observation at nanoscale of small particles of lignin molecules in any solvent. SANS experiments on co-solvent phase separation are complete and data analysis is underway. In addition, the water-THF studies can be extended into other co-solvents like organosolv-type pretreatment studies.

References

1. Cai, C. M., Zhang, T., Kumar, R., and Wyman, C. E., "THF Co-Solvent Enhances Hydrocarbon Fuel Precursor Yields from Lignocellulosic Biomass," *Green Chem.* **15**(11), 3140-3145 (2013).
2. Mostofian, B., Cai, C. M., Smith, M. D., Petridis, L., Cheng, X., Wyman, C. E., and Smith, J. C., "Local Phase Separation of Co-Solvents Enhances Pretreatment of Biomass for Bioenergy Applications," *J. Am. Chem. Soc.* **138**(34), 10869-10878 (2016).
3. Smith, M. D., Mostofian, B., Cheng, X., Petridis, L., Cai, C. M., Wyman, C. E., and Smith, J. C., "Cosolvent Pretreatment in Cellulosic Biofuel Production: Effect of Tetrahydrofuran-Water on Lignin Structure and Dynamics," *Green Chem.* **18**(5), 1268-1277 (2016).

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.

Fractionation of Lignocellulosic Biomass using biomass derived γ-valerolactone as solvent

Ali Hussain Motagamwala[1,2,*] (motagamwala@wisc.edu), **James A. Dumesic**[1,2]

[1]University of Wisconsin-Madison, Madison, Wisconsin; [2]Great Lakes Bioenergy Research Center, Madison, Wisconsin.

https://www.glbrc.org/research/deconstruction

Project Goals: Achieve carbon efficient fractionation of lignocellulosic biomass through solvent systems tailored for production of separate aqueous carbohydrate and lignin streams for downstream biological conversion.

Lignocellulosic biomass is a bountiful resource of renewable carbon and has the potential for the production of fuels and chemicals derived currently from fossil resources[1-2]. Lignocellulosic biomass can be grown on marginal land with limited resources, thus competing minimally with the food chain. However, compared to traditional carbon sources, like coal and petroleum, biomass is an expensive resource, and the atom-efficient utilization of this resource is pivotal for the economic success of a biomass-based industry. Biomass-derived sugars have been proposed to be versatile platform intermediates for production of fuels (through fermentation) and chemicals (through biological upgrading). Fractionation of biomass using γ-valerolactone (GVL) as solvent is unique in that GVL can be produced from the by-products of lignocellulosic biomass depolymerization, thereby leading to a self-sustainable biorefinery.

GVL-based fractionation process is potentially feedstock agnostic, allows virtually complete solubilization of the biomass, and leads to polysaccharide recovery at 70-85% yield.[3] Moreover, two separate carbohydrate streams comprising mainly C_5 or C_6 sugars can be obtained, thus the GVL-based process provides significant flexibility by allowing either co-utilization of these streams or unique processing strategies for each component (Figure 1). Biological upgrading of carbohydrates to biofuels or bio-based chemicals requires that the carbohydrates be separated from GVL. We have developed a solvent system consisting of GVL, water and an organic co-solvent which is mono-phasic at the fractionation temperatures (e.g., 160^{o}C) and is bi-phasic at lower temperatures (e.g., 25^{o}C).[4] This phase transition leads to a spontaneous partitioning of carbohydrates from the organic solvent components into an aqueous hydrolysate stream, thereby avoiding the need for expensive and potentially hazardous separation processes.

Finally, we have demonstrated that the solvent system comprised of GVL, water and co-solvent is also advantageous with respect to the recovery of lignin, as it renders a substantial portion of lignin (45 weight. %) insoluble in the solvent system upon cooling. Precipitating a portion of lignin present in the initial feed simplifies post-reaction processing.[4] Moreover, the lignin obtained using GVL as a solvent has a native-like structure with little evidence of lignin degradation other than some β-ether cleavage and concomitant molecular weight reduction.[3]

In summary, the GVL-based deconstruction method can enable conversion of all three biomass fractions (hemicellulose, cellulose, and lignin) in a carbon-efficient, renewable manner. When integrated into a biorefinery process, the GVL/co-solvent method is predicted to render significant cost savings as it bypasses the need for cellulolytic enzymes, thus making this a highly attractive alternative to traditional deconstruction methods.

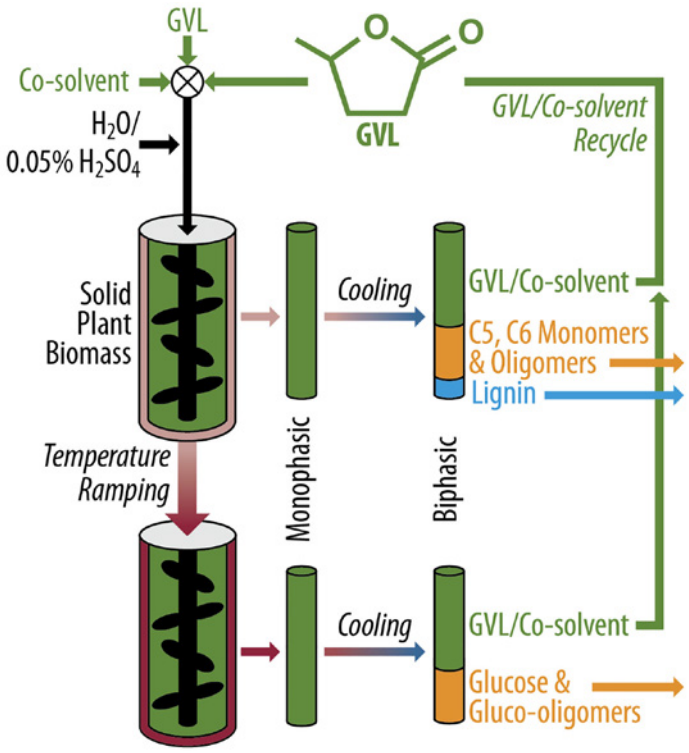


Figure 1. GVL-mediated biomass deconstruction. Conceptual diagram showing the fractionation of lignocellulosic biomass (~20 wt %) using GVL, water and a co-solvent and subsequent phase separation and product recovery.

References

1. 1. J. C. Serrano-Ruiz, R. M. West and J. A. Dumesic, Annu. Rev. Chem. Biomol. Eng., 2010, 1, 79–100.
2. G. W. Huber, I. Sara and A. Corma, Chem Rev., 2006, 106, 4044–4098.
3. J. S. Luterbacher, J. M. Rand, D. M. Alonso, J. Han, J. T. Youngquist, C. T. Maravelias, B. F. Pfleger and J. A. Dumesic, Science, 2014, 343, 277–280
4. A. H. Motagamwala, W. Won, C. T. Maravelias, J. A. Dumesic, Green Chem., 2016, 18, 5756.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE---FC02---07ER64494).

Discovery of Phenylacetate Decarboxylase, a New Glycyl Radical Enzyme Enabling First-Time Biochemical Production of Toluene

Andria Rodrigues[1]* (avrodrigues@lbl.gov), Kamrun Zargar[1], Yu-Wei Wu[2], Avneesh Saini[1], Renee Saville[1], Jose Henrique Pereira[1], Paul Adams[1], Susannah Tringe[3], Christopher Petzold[1], Jay Keasling[1], **Harry Beller**[1]

[1]Joint BioEnergy Institute (JBEI), Emeryville, CA; [2]Taipei Medical University, Taipei, Taiwan, [3]Joint Genome Institute (JGI), Walnut Creek, CA

Project Goals: The Joint BioEnergy Institute (JBEI) aims to produce a chemically diverse suite of biofuels from lignocellulosic biomass. Our objective in this project was enzyme discovery to enable first-time biochemical production of toluene, an important octane booster and petrochemical with a global annual market of 29 million tons, from cellulosic sugars.

Toluene is an important fuel additive and chemical feedstock with a global market of 29 million tons per year. It would be desirable to offset the enormous volume of petroleum-derived toluene with biochemically produced toluene made from a renewable resource (such as cellulosic biomass) using engineered microbes. Microbial toluene biosynthesis was reported in anoxic lake sediments more than 3 decades ago, however the enzyme(s) catalyzing this reaction have never been elucidated. Here we report the discovery of a toluene synthase (phenylacetate decarboxylase) from an anaerobic, sewage-derived enrichment culture that stoichiometrically produced toluene from phenylacetate. The discovery process (Zargar et al. 2016) included metagenome sequencing of the culture (which included more than 340,000 protein-coding genes), anaerobic FPLC (Fast Protein Liquid Chromatography) of cell-free extracts of the culture, and differential metaproteomic analyses to identify proteins present in active (toluene-producing) FPLC fractions but absent in adjacent inactive FPLC fractions (i.e., toluene synthase candidates). Toluene synthase candidates included a novel glycyl radical enzyme (GRE) of bacterial origin and its cognate activating enzyme [AE; a radical SAM (*S*-adenosyl-L-methionine) enzyme]. For functional confirmation, recombinant, N-terminally tagged, codon-optimized versions of the GRE and AE genes were expressed in *E. coli* and purified under anaerobic conditions. After *in vitro* reconstitution of the AE to restore its [4Fe-4S] cluster, its activity was confirmed *in vitro* by measuring conversion of SAM to methionine. *In vitro* assays with purified GRE, AE, and SAM were shown to successfully convert ^{13}C-labeled phenylacetate to ^{13}C-labeled toluene, whereas no toluene was produced in control assays lacking SAM or in assays containing mutant versions of GRE in which the site of the glycine radical was converted to alanine. Notably, metagenomic analysis of anoxic, toluene-producing lake sediment enrichment cultures contained nearly identical versions of the GRE and AE (>85% protein sequence identity) even though the lake sediment and sewage cultures had very different community composition. We have begun to further characterize the GRE (e.g., with respect to its activity on substituted phenylacetic acid analogs). Overall, this first-time identification of a

phenylacetate decarboxylase will ultimately enable bio-based toluene production *via* engineered microbial hosts.

References:

1. Zargar, K., R. Saville, R. Phelan, S. G. Tringe, C. J. Petzold, J. D. Keasling, and H. R. Beller. 2016. *In vitro* characterization of phenylacetate decarboxylase, a novel enzyme catalyzing toluene biosynthesis in an anaerobic microbial community. *Scientific Reports* (*Nature*) **6**, 31362 doi: 10.1038/srep31362.

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.

From Field to Fuel: How Do Feedstock Variability and Plant-Experienced Stresses Affect Microbial Conversion?

Rebecca G. Ong[1,2]* (rgong1@mtu.edu), Yaoping Zhang,[3,4] Lawrence G. Oates,[3,4] Trey K. Sato,[3,4] Gregg R. Sanford,[3,4] Richard Amasino,[3,4] Bruno Basso,[5,6] David Lowry,[5,6] Kurt Thelen,[5,6] Audrey Gasch,[3,4] and **Timothy Donohue**[3,4]

[1]Michigan Technological University, Houghton, MI; [2]DOE Great Lakes Bioenergy Research Center, Houghton, MI; [3]University of Wisconsin-Madison, Madison, WI; [4]DOE Great Lakes Bioenergy Research Center, Madison, WI; [5]Michigan State University, East Lansing, MI; and [6]DOE Great Lakes Bioenergy Research Center, East Lansing, MI

https://www.glbrc.org/

Project Goals:

Different growth conditions and deconstruction methods yield surprisingly diverse hydrolysates in which various small-molecule inhibitors disrupt microbial metabolism, cellular integrity, and product formation. But how plant species, growth conditions, and processing methods contribute to microbial inhibition is poorly understood. In our previous work, we showed, for the first time, that drought can severely inhibit downstream microbial conversion processes.[1] In this study, yeast growth was completely inhibited in drought-year switchgrass hydrolysate. This was due to a stress response in switchgrass that led to the accumulation of soluble sugars, which were ultimately degraded during pretreatment to compounds (imidazoles and pyrazines) that were inhibitory to yeast growth. These results demonstrated that plant-experienced environmental stresses can strongly affect fermentation microorganisms and biofuel yields. Our project aims to determine how variation in plants and growth environments, particularly stressful conditions, coupled with the effects of different deconstruction methods, impact microbial conversion.

References

1. Ong, R.G.; Higbee, A.; Bottoms, S.; Dickinson, Q.; Xie, D.; Smith, S.A.; Serate, J.; Pohlmann, E.; Jones, A.D.; Coon, J.J.; Sato, T.K.; Sanford, G.R.; Eilert, D.; Oates, L.G.; Piotrowski, J.S.; Bates, D.M.; Cavalier, D.; Zhang, Y., Inhibition of microbial biofuel production in drought-stressed switchgrass hydrolysate. *Biotechnology for Biofuels* **2016,** 9(1), 237. [10.1186/s13068-016-0657-0]

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

Evolutionary Dynamics of a Secondary Metabolite Gene Cluster in Budding Yeasts

David J. Krause,[1,4*] (dkrause2@wisc.edu), Jacek Kominek[1,4], Dana A. Opulente[1,4], Xing-Xing Shen[2], Xiaofan Zhou[2], Jeremy DeVirgilio[3], Amanda B. Hulfachor[1], Cletus P. Kurtzman[3], Antonis Rokas[2], and **Chris Todd Hittinger[1,4]**

[1]Laboratory of Genetics, Genome Center of Wisconsin, Wisconsin Energy Institute, J.F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, WI 53706, USA; [2]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA; [3]Mycotoxin Prevention and Applied Microbiology Research Unit, National Center for Agricultural Utilization Research, Agricultural Research Service, US Department of Agriculture, Peoria, IL USA; and [4]DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI 53706, USA

Project Goals

This project aims to improve our understanding of how microorganisms apportion carbon flux into biofuel-relevant metabolic pathways, including how we can rationally engineer flux to improve the production of specialty biofuels.

Abstract

The focus of biomass conversion into biofuels has recently shifted away from the production of ethanol and into the production of specialty biofuels, such as isobutanol. For the yeast *Saccharomyces cerevisiae*, this presents several challenges, as native production of isobutanol is minimal compared to ethanol. One major challenge is that the native flux through the branched-chain amino acid (BCAA) pathway, from which isobutanol is produced, is quite low. We therefore sought to identify species of yeast with naturally higher flux through this pathway. To do this, we studied a small subset of yeast species that produce a pigment called pulcherrimin, which is identifiable by its characteristic red color when bound to iron. Pulcherrimin is a cyclic-dipeptide derivative produced from two leucine molecules, linking it directly to BCAA biosynthesis. We identified a four-gene cluster conserved among species that produced pulcherrimin, which we named the PULcherrimin gene cluster. Using targeted gene replacements in the pulcherrimin-producing yeast *Kluyveromyces lactis*, we found that all four genes play a role in either pulcherrimin biosynthesis or re-utilization of pulcherrimin-bound iron from culture medium. To our knowledge, this is the first demonstration of a functional secondary metabolite gene cluster in budding yeasts. In characterizing the species distribution of the *PUL* gene cluster, we found the presence of partial clusters in several more species, including *S. cerevisiae*. These partial clusters always consisted of the genes required for utilization of pulcherrimin, and not the biosynthesis genes. We performed targeted gene replacements in *S. cerevisiae* and confirmed the roles of these genes, neither of which had previously been assigned a known function. We found no evidence for acquisition of the *PUL* cluster through gene gains via horizontal gene transfer, and we therefore hypothesize that the *PUL* cluster was present in the ancestor of all budding yeasts, but was lost in most yeast lineages. We also predict a public goods dilemma that emerges from lineages that have lost the pulcherrimin biosynthesis genes, but maintain the genes involved in pulcherrimin utilization. Future work will focus on the molecular mechanisms by which pulcherrimin-producing lineages vary in their levels of production to better understand how they differently apportion their BCAA carbon flux.

Funding Statement

This work was supported by the National Science Foundation under Grant No. DEB-1442148, the USDA National Institute of Food and Agriculture, Hatch project 1003258, and by DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-08ER64494). CTH is a Pew Scholar in the Biomedical Sciences, supported by the Pew Charitable Trusts.

Microbial metabolic engineering to produce alcohols from cellulosic hydrolysates

Jaewon Lee[1,2], Stephen L. Lane[1,2], **Yong-Su Jin[1,2]* (ysjin@illinois.edu)**

[1]Center for Advanced Bioenergy and Bioproducts Innovation (CABBI); [2]Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, IL

http://cabbi.bio

Project Goals:

Alcohols are much less toxic than other advanced biofuels and yeast can produce and tolerate them up to 150 g/L. We have engineered yeast strains capable of producing high titers (100-150 g/L) of ethanol (mono-alcohol) and 2,3-butanediol (di-alcohol) not only from glucose but also from prevalent sugars (cellobiose, xylose, and galactose) in the hydrolysates of terrestrial and marine biomass. While alcohols can be used as a biofuel directly, they can be also catalytically upgraded into various chemicals. Ethanol and related alcohols, such as n-butanol can be dehydrated into corresponding olefins with high selectivity and conversion yield. 2,3-Butanediol can be also dehydrated into methyl ethyl ketone (MEK), a potential platform chemical towards synthesis of jet and diesel fuels and 1,3-butadiene, a monomer of synthetic rubber361. Given the feasibility of high titer production and catalytic upgrading, we will engineer east strains to produce various alcohols from inexpensive sugar sources with a high yield and titer. We will develop and use a versatile and fully automated biofoundry, the Illinois Biological Foundry for Advanced Biomanufacturing (iBioFAB) for strain improvement.

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-SCxxxxxx (J.L., S.L.T. and Y.-S. Jin through CABBI).

Expanding the Scope of Biofuels and Chemicals Produced by Microorganisms through Design of Artificial Enzymes Containing Unnatural Amino Acids and Non-native Cofactors

Chang Cui*,[1] (ccui5@illinois.edu), Lam Tung Vo[2], Huimin Zhao[2] and **Yi Lu**[1]

[1]Department of Chemistry, [2]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL

https://chemistry.illinois.edu/directory/profile/yi-lu

Project Goals: Microorganisms are increasingly used to produce biofuels and chemicals. However, developing robust microorganisms for the economical production of biofuels and bioproducts from low-cost, often-recalcitrant feedstocks at large scale with high titers, rates, and yields remains a significant challenge, especially when the products are not natural compounds. Compared to the vast number of chemicals and fuels produced by chemical processes, the number of biofuels and bioproducts that have been produced by microorganisms is rather small. While progress has been made in biotransformation of novel compounds using native enzymes and their variants, it becomes increasingly difficult to use native enzymes for many other biotransformations that produce novel molecules that are unprecedented in nature, or even in chemical or biological fields. We hypothesize that a major reason for such a limitation is that native enzymes using 20 natural amino acids and a limited number of biosynthesized cofactors, such as protoporphyrin IX, for structures and functions. To overcome the limitation in order the meet the challenge, we plan to incorporate unnatural amino acids and non-native cofactors into native enzymes to expand the functional groups and exert control over reactivity and selectivity.

To incorporate unnatural amino acids into a protein, we plan to use genetic codon expansion by engineering a pair of orthogonal amino acyl tRNA synthetase (aaRS) and tRNA.[1-5] We choose this method because it has been demonstrated to work well in many organisms ranging from bacteria like *E. coli* to the multicellular organism *Caenorhabditis elegans* and *Arabidopsis thaliana*, and most recently to mammalian cells. To incorporate non-native cofactors into proteins, we will first replace native cofactors, such as heme, with non-native cofactors of similar structure, such as another planer molecule, MnSalen or other metalloSalens (Salen =N,N'-bissalicylidene-1,2-ethanediamino anion)[6-7] and organometallic catalysts.[8] The resulting artificial enzymes will first be characterized by spectroscopic methods, such as UV-vis and EPR, followed by enzymatic activity measurements to assess the effectiveness of this approach and its products.

While the above method works well in test tubes, it becomes very difficult to apply the above method for living cells. To overcome this limitation, we use iBioFAB in the Carl R. Woese Institute for Genomic Biology at the University of Illinois at Urbana-Champaign to explore different combinations of cofactors, enzymes and reaction conditions. This use of iBioFAB for high throughput cloning and enzyme characterization will allow us to design artificial and select metalloenzymes with not only high efficiencies, but also high selectivity.

References

1. Yi Lu, Natasha Yeung, Nathan Sieracki and Nicholas M. Marshall, “Design of functional metalloproteins”, *Nature* 460, 855-862 (2009).
2. Yi Lu, "Design and Engineering of Metalloproteins Containing Unnatural Amino Acids or Non-native Metal-containing Cofactors," *Curr. Opin. Chem. Biol.* 9, 118-126 (2005).
3. Yang Yu, Cui Chang, Xiaohong Liu, Igor Petrik, Jiangyun Wang and Yi Lu, “A Designed Metalloenzyme Achieving the Catalytic Rate of a Native Enzyme," J. *Am. Chem. Soc.* 137, 11570–11573 (2015).
4. Yang Yu, Qing Zhou, Li Wang, Xiaohong Liu, Wei Zhang, Meirong Hu, Jianshu Dong, Jiasong Li, Xiaoxuan Lv, Hanlin Ouyang, Han Li, Feng Gao, Weimin Gong, Yi Lu and Jiangyun Wang, “Significant improvement of oxidase activity through the genetic incorporation of a redox-active unnatural amino acid,” *Chem. Sci.* 6, 3881-3885 (2015).
5. Yang Yu, Xiaoxuan Lv, Jiasong Li, Qing Zhou, Chang Cui, Parisa Hosseinzadeh, Arnab Mukherjee, Mark J. Nilges, Jiangyun Wang and Yi Lu, “Defining the Role of Tyrosine and Rational Tuning of Oxidase Activity by Genetic Incorporation of Unnatural Tyrosine Analog,” *J. Am. Chem. Soc.* 137, 4594-4597 (2015).
6. James R. Carey, Steven Ma, Thomas D. Pfister, Dewain K. Garner, Hyeon K. Kim, Joseph A. Abramite, Zhilin Wang, Zijian Guo, and Yi Lu, "A Site-Selective Dual Anchoring Strategy for Artificial Metalloprotein Design," *J. Am. Chem. Soc.* 126, 10812-10813 (2004).
7. Dewain K. Garner, Lei Liang, David A. Barrios, Jun-Long Zhang, and Yi Lu, “The Important Role of Covalent Anchor Positions in Tuning Catalytic Properties of a Rationally Designed MnSalen-Containing Metalloenzyme,” *ACS Catal.* 1 (A special thematic issue on Biocatalysis and Biomimetic Catalysis), 1083–1089 (2011).
8. Hee Jung Hwang, James R. Carey, Evan T. Brower, Alan Gengenbach, Joseph Abramite and Yi Lu, “Blue Ferrocenium Azurin: An Organometalloprotein with Tunable Redox Properties,” *J. Am. Chem. Soc.* 127, 15356-15357 (2005).

This material by Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research.

"Ideal lignin" facilitates full biomass utilization

Yanding Li[1,2]* (yli594@wisc.edu), Li Shuai,[1,3] Hoon Kim,[1,4] Ali Hussain Motagamwala,[1,5] Justin K. Mobley,[1] Fengxia Yue,[1,4] Daphna Havkin-Frenkel,[6,7] Jeremy S. Luterbacher,[8] James A. Dumesic,[1,5] and **John Ralph**[1,4]

[1]DOE Great Lakes Bioenergy Research Center (GLBRC), Wisconsin Energy Institute, University of Wisconsin-Madison (UWM), WI. [2]Dept. of Biological Systems Engineering, UWM, WI. [3]Depart. of Sustainable Biomaterials, Virginia Tech, Blacksburg, VA. [4]Dept. of Biochemistry, UWM, WI. [5]Dept. of Chemical and Biological Engineering, UWM, WI. [6]Dept. of Plant Biology and Pathology, Rutgers U., New Brunswick, NJ. [7]Bakto Flavors, LLC, North Brunswick, NJ. [8]Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

http://www.glbrc.org

Project Goals: The mission of the GLBRC is to perform the basic research that generates technology to convert cellulosic biomass to advanced biofuels. In combining aims for the "design" of superior plants with lignins that allow efficient depolymerization, we have the following goals: 1) Delineate the features of an "ideal lignin" for streamlined processing of lignocellulosics in a biorefinery; 2) Demonstrate high-yield lignin conversion to a small array of phenolic monomers after the usual processing to derive monomeric sugars (and/or liquid fuels) from the polysaccharides.

Lignin, one of the major components of lignocellulosic biomass, is crucial to plant growth and development, but poses a major impediment to biomass utilization in various processes. It is, however, increasingly recognized that valorizing lignin is essential for sustainably and profitably converting biomass to biofuels and coproducts.[1] Hydrogenolysis, i.e., the hydrogenation of un-processed solid biomass by a solid noble metal catalyst, remains one of the most promising methods for producing aromatic monomers. By cleaving the major linkage (β–ether) in native lignin, this process can produce simple monomeric products in high yields. However, hydrogen-olysis still suffers from some issues. In most wild-type biomass, the lignin polymer is composed of three phenylpropanoid subunits, *p*-hydroxyphenyl (**H**), guaiacyl (**G**), and syringyl (**S**), derived by combinatorial radical coupling from the three main monolignols, *p*-coumaryl, coniferyl, and sinapyl alcohols. Although **H**-units are typically low-level, this results in at least three different types of monomers (**H**, **G**, and **S**) each with ethyl, propyl, and propanol sidechains as the primary hydrogenolysis products, which makes monomer separation and utilization difficult. Lignin's principal alkyl-aryl-ether units with their β–O–4 inter-unit bonds (45-85%) can be selectively cleaved, but other linkages including β–5 (1-12%), β–β (5-12%), 5–5 (1-9%), 4–O–5 (~2%) and β–1 (1-2%) that are also present in lignins remain largely intact;[1] carbon-carbon (C–C) and diaryl ether (4–O–5) units typically result in dimeric or higher oligomeric products. The use of extracted lignins rather than whole biomass has the advantage that the material can be fully

dissolved in organic solvents and may facilitate catalyst recovery and continuous processing. However, industrial lignin fractionation, particularly by acidolytic methods, is known to cause some β-ether cleavage and result in the condensation between units via the electrophilic substitution of acid-generated carbocation intermediates on the electron-rich aromatic rings,[2] limiting hydrogenolytic depolymerization yields.

With the revelations regarding lignins' structural malleability from studies on lignin pathway mutants and transgenics as well as on various 'natural' plants discovered to possess unusual lignins, researchers have been able to contemplate actually designing lignins for improved utilization.[3] It is now a realistic juncture to posit the characteristics for an "ideal lignin" archetype for biomass processing. For the depolymerization of the lignin polymer to monomers, lignin should have the following three necessary and sufficient characteristics. First, it should contain only ether (C–O) inter-unit linkages in its backbone. Second, it should be stable under acidic conditions to prevent condensation and the generation of undesired new C–C bonds, as well as the sidechain truncation that often accompanies hydrolysis, during pretreatment. Last, it should be generated *in planta* from a single phenylpropanoid monomer to allow the production of the simplest array of compounds.

The recently discovered catechyl lignin (**C**-lignin),[4] which is essentially a homopolymer biosynthesized purely by β–O–4-coupling of caffeyl alcohol with the growing polymer chain to produce benzodioxanes as the overwhelmingly dominant unit in the polymer, might represent such an ideal lignin that overcomes biomass fractionation issues and expands the opportunities for lignin valorization. Here we discovered that **C**-lignin is highly acid resistant and is able to maintain its native structure even after harsh acidic treatments. Hydrogenolysis of **C**-lignin resulted in the cleavage of essentially all benzodioxane structures to cleanly produce catechylpropane and catechylpropanol, and a small amount of a chromane, in high yield. The unique properties make the valorization of **C**-lignin compatible with traditional inexpensive acid pretreatment methods and allow continuous upgrading of the extracted lignin. Such valuable properties suggest that there is unrecognized value in certain biomass residues, and highlight the merit in pursuing genetic means to modify biomass to produce **C**-lignin.

References

1. Rinaldi, R.; Jastrzebshi, R.; Clough, M. T.; Ralph, J.; Kennema, M.; Bruijnincx, P. C. A.; Weckhuysen, B. M., Paving the way for lignin valorisation: Recent advances in bioengineering, biorefining and catalysis. *Angewandte Chemie (International Edition)* **2016,** *55* (29), 8164-8215.
2. Shuai, L.; Amiri, M. T.; Questell-Santiago, Y. M.; Héroguel, F.; Li, Y.; Kim, H.; Meilan, R.; Chapple, C.; Ralph, J.; Luterbacher, J. S., Formaldehyde stabilization facilitates lignin monomer production during biomass depolymerization. *Science* **2016,** *354* (6310), 329-333.
3. Mottiar, Y.; Vanholme, R.; Boerjan, W.; Ralph, J.; Mansfield, S. D., Designer lignins: harnessing the plasticity of lignification. *Current Opinion in Biotechnology* **2016,** *37* (1), 190-200.
4. Chen, F.; Tobimatsu, Y.; Havkin-Frenkel, D.; Dixon, R. A.; Ralph, J., A polymer of caffeyl alcohol in plant seeds. *Proceedings of the National Academy of Science USA* **2012,** *109* (5), 1772-1777.

Funding was provided by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

Bioconversion of Sugars and Aromatics from Engineered Plants to Methyl Ketones

Jie Dong[1,2], Aymerick Eudes[1,3], James Sun[1,2], Veronica Texeira Benites[1,2], Edward Baidoo[1,2], Paul Adams[1,3], Henrik Scheller[1,4], Blake Simmons[1,2], **Steven W. Singer[1,2,*] (SWSinger@lbl.gov)**

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory; [3]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory; [4]Environmental Genomics and Systems Biology Division Lawrence Berkeley National Laboratory.

http://www.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 tuning plant and microbial metabolic engineering to match the substrates provided by the plant with the metabolic capabilities demonstrated by the microbial hosts.

Abstract

Current strategies for bioconversion of plant biomass often focus on glucose as the sole substrate. Expanding the substrate range of bioconversion to include both xylose and soluble aromatics from lignin will improve the efficiency of bioconversion. This improvement will enable economic production of highly reduced biofuels and bioproducts obtained from the fatty acid and isoprenoid pathways. Here, we describe the construction of *Pseudomonas putida* strains that produce high titers of medium chain methyl ketones from both sugars and monoaromatic lignin-related compounds. These *P. putida* strains also produce methyl ketones from plants that were engineered for reduced recalcitrance and increased production of soluble monoaromatics. These studies provide a blueprint for linking plant and microbial engineering to maximize bioconversion of carbon in biomass.

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.

Rapid Domestication of Non-Model Microbial Hosts for Biofuels Production

Lauren Riley[1,2], E. Anne Hatmaker[1], Beth Papanek[1], Kaela O'Dell[1], Irenee Payne[1], Carly Duffy[1], Bob Schmitz[3], Jan Westpheling[3], Adam Guss[1,2*] (gussam@ornl.gov) and **Gerald A. Tuskan**[1]

[1]Oak Ridge National Laboratory, Oak Ridge, TN; [2]University of Tennessee, Knoxville, [3]University of Georgia, Athens, 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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Many organisms naturally possess complex physiological phenotypes that are of interest for biotechnology research. Often, these traits are challenging to transfer into traditional host organisms such as *Escherichia coli* and *Saccharomyces cerevisiae*. Therefore, the ability to rapidly domesticate non-model organisms to harness these traits could usher in a new era of biotechnology where synthetic biology is routinely applied to these organisms. However, these organisms are typically unable to be bioengineered due to a lack of available genetic tools and an insufficient foundation of knowledge about the organism.

The development of genetic tools is limited largely by the inability to efficiently transform DNA into these organisms. A critical barrier to transformation is DNA restriction-modification systems, which act as a bacterial immune system to cut DNA that is methylated differently than in the host. Typically, these systems are comprised of methylation and restriction subunits. To prevent host death, the cognate DNA methyltransferases recognize the same target sequence as the restriction enzymes, and the methylated DNA is protected from restriction. Therefore, in order to prevent restriction of heterologous DNA, the DNA needs to be methylated in the same manner as the host organism prior to transformation. In order to determine the sites targeted for restriction in these strains, we performed methylome analysis for organisms of interest in collaboration with the Department of Energy's Joint Genome Institute. This information was used to choose methyltransferases for expression in *E. coli* to protect DNA for *Clostridium clariflavum* and *Megasphaera elsdenii* transformation. For *C. clariflavum*, nine distinct DNA sequences were found to be methylated in the native host. Methyltransferases targeting five of these sites have been expressed in *E. coli* and demonstrated to be functional using methylome analysis on the *E. coli* expression strain. While not methylated completely in the same positions

as the native host, this DNA was used to demonstrate successful transformation of *C. clariflavum*. Similarly, through a combination of methylome analysis and methyltransferase expression in *E. coli*, we demonstrated DNA transformation of two different *M. elsdenii* strains that each have different methylation patterns.

To develop *M. elsdenii* into a new bioengineering platform, we also began addressing the knowledge gap in *Megasphaera*. We sequenced the *M. elsdenii* genome and used this information to build a metabolic reconstruction in the DOE Systems Biology Knowledgebase, which will serve as the foundation for a metabolic model. Transcriptomic data is also helping to elucidate the metabolic pathways used by *M. elsdenii* for the bioconversion of lactic acid and sugars into butyric and hexanoic acids. The genetic tools are now being combined with the new physiological knowledge to engineer *M. elsdenii* to expand the substrate range and to produce new value-added products. Similar approaches are being taken with other organisms of interest to demonstrate that this is a broadly applicable approach to developing new host organisms for advanced bioengineering.

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

Exposure to oxygen induces a transient bottleneck in the methylerythritol 4-phosphate pathway in *Zymomonas mobilis*: the role of iron-sulfur cluster assembly proteins and flavodoxin reductase in recovering pathway activity

Julia Martien[1,2]* (martien@wisc.edu), David Stevenson[1,2] , Alex Hebert[3] , Joshua Coon[3] , **Daniel Amador-Noguez**[1,2]

[1]University of Wisconsin-Madison Department of Bacteriology; [2]Great Lakes Bioenergy Research Center; [3]University of Wisconsin-Madison Department of Biomolecular Chemistry

Project Goals: The goal of this research is to inform metabolic engineering efforts to improve yields of isoprenoid-based bioproducts via the methylerythritol 4-phosphate (MEP) pathway. By examining native responses to a severe metabolic bottleneck in the MEP pathway, we hope to gain a new perspective about the key regulatory enzymes controlling MEP pathway activity. www.glbrc.org/research/conversion

The methylerythritol 4-phosphate (MEP) pathway generates isoprenoid precursors in bacteria. Isoprenoids are a diverse class of molecules, encompassing several desirable bioproducts including pharmaceuticals, synthetic polymers, and high-grade fuels. In recent years, efforts have been made to metabolically engineer microbes for over-production of isoprenoid commodity molecules via the MEP pathway. However, directed engineering has been limited by an incomplete understanding of the regulatory modules controlling pathway activity.

Our work investigates the metabolic regulation of the MEP pathway in the emerging biofuel-producer *Zymomonas mobilis* by monitoring metabolic changes in the MEP pathway and connected pathways in response to oxygen exposure. Using metabolite quantification with UHPLC-MS, we have observed a dramatic and transient metabolic bottleneck in the MEP pathway induced by exposure to oxygen. Metabolite profiling indicates this bottleneck is caused by inactivation of the final two enzymes of the pathway, IspG and IspH, likely due to oxidative damage of their Fe_4S_4 iron-sulfur cluster cofactors. In *E. coli*, and likely in *Z. mobilis*, these Fe_4S_4 clusters are reduced in by flavodoxin, which is in turn reduced by an NADPH-dependent flavodoxin reductase.[1,2].

Proteomics and transcriptomics data suggest the bottleneck in the MEP pathway is relieved by increased expression of the SUF iron-sulfur assembly operon and flavodoxin reductase, recovering IspG and IspH function by maintaining in-tact and reduced iron-sulfur cluster cofactors in both enzymes. Our findings have thereby identified the SUF iron-sulfur cluster assembly operon and flavodoxin reductase as potential targets for overexpression to increase microbial production of isoprenoid commodity molecules.

References:

1. Puan, K.-J., Wang, H., Dairi, T., Kuzuyama, T. & Morita, C. T. *fldA is an essential gene required in the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid biosynthesis*. *FEBS Letters* **579,** (2005).
2. Rohdich, F. *et al.* The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 1586–91 (2003).

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02- 07ER64494).

Multi-omics analysis of *Rhodococcus opacus* strains evolved for optimized lignocellulose bioconversion

William R. Henson[1], Tayte P. Campbell[2]* (tayte.campbell@wustl.edu), Aki Yoneda[2,3], Yu Gao[1], Drew DeLorenzo[1], Bertram Berla[2], Soo Ji Kim[1], **Marcus Foston[1]**, **Tae Seok Moon[1]**, and **Gautam Dantas[2,3,4,5]**

[1] Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, St. Louis, MO, 63130, USA; [2] The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63110, 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

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. 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.

Lignin-derived aromatics inhibit microbial conversion of lignocellulose-based sugars to fuels and chemicals, and microbial lignin valorization strategies require consumption of complex aromatic mixtures. *R. opacus*, a promising biofuel production strain, can tolerate and consume individual lignin-derived aromatics, but degradation pathways and tolerance mechanisms for aromatic mixtures are not well characterized. To address this knowledge gap, we performed adaptive evolution using lignin model compound mixtures and characterized adapted strains using multiple omics approaches. Adapted strains showed up to 1900% growth improvement compared to the wild-type strain in model lignin compound mixtures. We identified multiple genes involved in redox reactions that were affected by mutations in four distinct adaptation experiments. Both the wild-type and adapted strain PVHG6 co-consumed multiple lignin model compounds, and transcriptomics showed distinct funneling degradation pathways upregulated in response to each aromatic compound. Together, our results provide new insights into the complex metabolism of aromatic compound mixtures by *R. opacus* and suggest that tuning redox metabolism can enable improved catabolism of aromatic compounds.

Publications

1. 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)

2. Yoneda A, Henson WR, Goldner NK, Park KJ, Forsberg KJ, Kim SJ, Pesesky MW, Foston M, Dantas G, Moon TS. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating *Rhodococcus opacus* PD630, Nucleic Acids Res. 18, 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-SC0012705.

Rerouted PKA Signaling Coordinates Sugar And Hypoxia Responses For Anaerobic Xylose Fermentation In Yeast

Kevin S. Myers[1*] (kmyers2@wisc.edu), Nicholas M. Riley[2], Matthew E. MacGilvray[3], Trey K. Sato[1], Mick McGee[1], Justin Heilberger[1], Joshua J. Coon[2,4,5,6], **Audrey P. Gasch**[1,3,4]

[1] Great Lakes Bioenergy Research Center, University of Wisconsin - Madison, Madison, WI; [2] Department of Chemistry, University of Wisconsin - Madison, Madison, WI; [3] Laboratory of Genetics, University of Wisconsin - Madison, Madison, WI; [4] Genome Center of Wisconsin, University of Wisconsin - Madison, Madison, WI; [5] Department of Biomolecular Chemistry, University of Wisconsin - Madison, Madison, WI; [6] Morgridge Institute for Research, Madison, WI

http://www.glbrc.org

Project Goals:

We aim to better understand the metabolism of strains engineered and evolved for anaerobic xylose fermentation using multi-omics and network analysis. Using these data, we also set out to improve anaerobic xylose fermentation and to decouple metabolism from growth in order to promote flux of carbon toward ethanol production rather than cell biomass. The knowledge gained from this study can potentially be applied to other strains for production of other desired chemicals using cellulosic biomass and yeast.

Microbes can be engineered for novel metabolism to produce biofuels and chemicals, but rerouting metabolic flux toward products remains a major hurdle. We used multi-omics and network analysis to explore cellular rewiring across a panel of yeast strains engineered for anaerobic xylose fermentation, important for sustainable biofuel production from plant cellulosic biomass. We show that rerouted Protein Kinase A (PKA) signaling regulates both hypoxia and sugar responses, in part via the transcription factor Azf1, catalyzing a cascade of effects correlated with anaerobic xylose growth. By deleting the PKA regulatory subunit we successfully decoupled growth and metabolism, enabling us to distinguish phosphorylation changes related to xylose-dependent growth *versus* metabolism. Using this information, we generated an industrially relevant strain with high rates of anaerobic xylose conversion.

Funding statement.

This work is funded by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494).

SteadyCom: Predicting Microbial Abundances while Ensuring Community Stability

Siu H. J. Chan[1] (sxc554@psu.edu), Margaret Senftle[1,&], Lin Wang[1], Satyakam Dash[1], and **Costas D. Maranas**[1]

[1]Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802
[&]Current affiliation: Department of Chemical and Biomolecular Engineering, Rice University

http://www.maranasgroup.com/

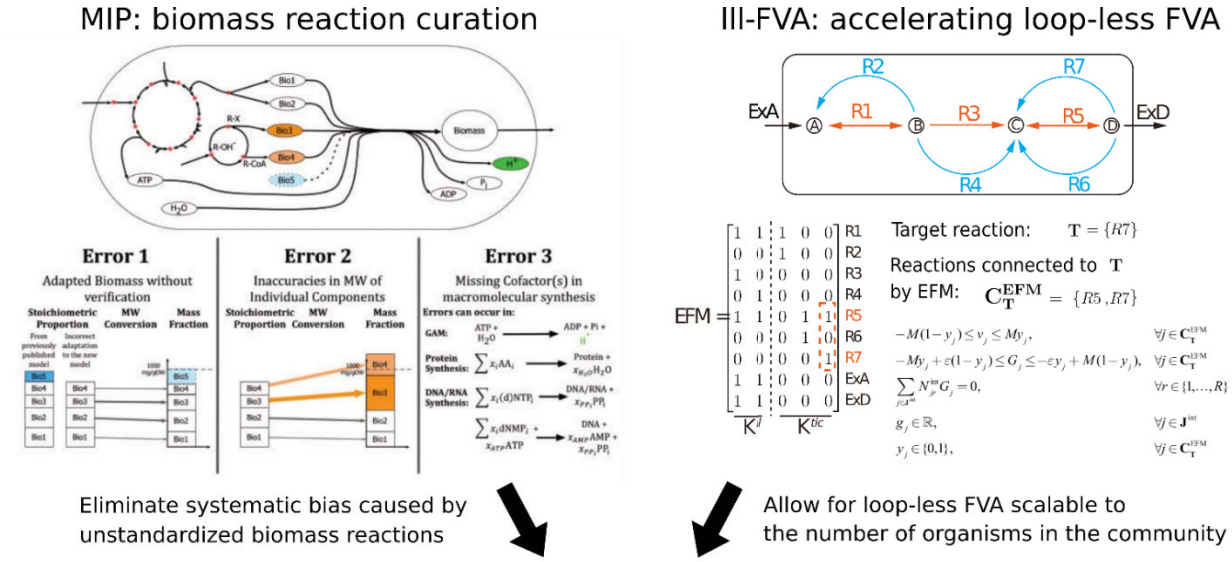
Project Goals: This project aims to apply and extend the genome-scale metabolic modeling approach to microbial communities. To simulate the metabolism of realistic microbial communities which can consist of up to hundreds species, scalable computational methods for simulation and analysis as well as a standardization procedure for unifying single-organism models are required. To this end, an optimization framework termed SteadyCom was developed to simulate the steady-state community metabolism of microbial communities. Together with a procedure introduced to standardize the biomass reactions in single-organism models and a scalable method devised for loop-less flux variability analysis, the results propose a workflow to analyze microbial communities at steady-state that is scalable to a large number of organisms.

Genome-scale metabolic modeling has become widespread for analyzing microbial metabolism. Extending this established paradigm to more complex microbial communities is emerging as a promising way to unravel the interactions and biochemical repertoire of these omnipresent systems. While several modeling techniques have been developed for microbial communities, little emphasis has been placed on the need to impose a time-averaged constant growth rate across all members for a community to ensure co-existence and stability. In the absence of this constraint, the faster growing organism will ultimately displace all other microbes in the community. We introduced the SteadyCom [1] optimization framework for predicting metabolic flux distributions consistent with the steady-state requirement. SteadyCom can be rapidly converged by iteratively solving linear programming (LP) problem and the number of iterations is independent of the number of organisms. A significant advantage of SteadyCom is compatibility with flux variability analysis (FVA). SteadyCom is first demonstrated for a community of four *E. coli* double auxotrophic mutants and is then applied to a gut microbiota model consisting of nine species, with representatives from the phyla Bacteroidetes, Firmicutes, Actinobacteria and Proteobacteria. In contrast to the direct use of flux balance analysis (FBA), SteadyCom is able to predict an abundance profile with a good agreement to experimental gut microbiota. SteadyCom provides an important step towards the cross-cutting task of predicting the composition of a microbial community in a given environment.

During simulating the community metabolism, an unstandardized biomass reaction of any of the organisms in the community that produces biomass with a molecular weight (MW) different from the defined standard 1 g $mmol^{-1}$ introduces a systematic error to the simulation. We developed the systematic procedure termed Minimum Inconsistency under Parsimony (MIP) [2] for checking the biomass weight and curating the biomass reaction so as to eliminate the systematic error in community simulations. We demonstrated that if the biomass reactions are not standardized, biomass MW discrepancies are accentuated in microbial community simulations as they can cause significant and systematic errors in the community composition. Microbes with underestimated biomass MWs are overpredicted in the community whereas microbes with overestimated biomass weights are underpredicted. The observed departures in community composition are disproportionately larger than the discrepancies in the biomass weight estimate. The procedure represents an important preprocessing step to ensure unbiased simulations of community metabolism.

To effectively perform FVA in the absence of thermodynamically infeasible cycles (TICs) in community models, we devised a method termed localized loop-less flux variability analysis (lll-FVA) [3] with significantly improved computational performance. We identified the fewest needed

constraints sufficient for optimality under the loop-less requirement in terms of elementary flux modes and put forth the concept of localized loop-less constraints to enforce this minimal required set of loop-less constraints. The computational time for loop-less FVA is reduced by a factor of 10-150 compared to the original loop-less constraints and by 4-20 times compared to the currently fastest method Fast-SNP with the percent improvement increasing with model size. Importantly, lll-FVA offers a scalable strategy for loopless flux calculations for multi-compartment/multi-organism models of very large sizes (e.g. $>10^4$ reactions) not feasible before. The Matlab implementations for SteadyCom, MIP and lll-FVA are available at https://github.com/maranasgroup.



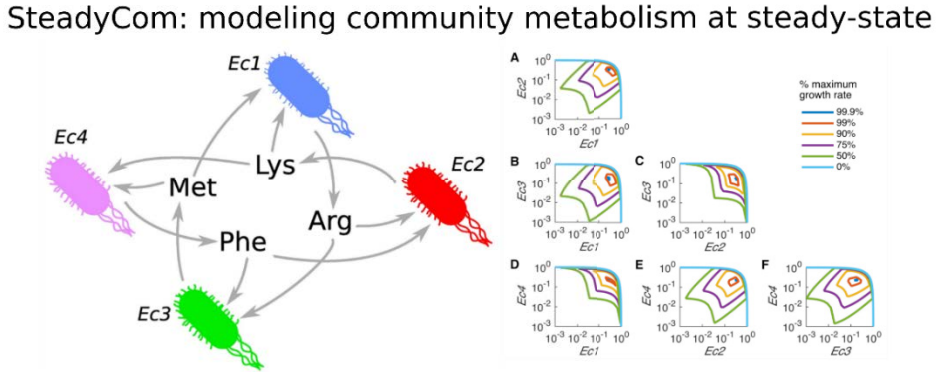


Figure 1. Modeling microbial communities using the SteadyCom framework. By integrating individual models standardized and verified through the MIP procedure [2], SteadyCom [1], which is compatible with flux variability analysis (FVA), can predict metabolic interactions in microbial communities at steady-state. The lll-FVA technique allows for loop-less flux analysis in microbial communities that is scalable to the increasing number of community members [3].

References

1. Chan S.H.J., Simons-Senftle M.N., and Maranas C.D. (2017) SteadyCom: Predicting microbial abundances while ensuring community stability. *PLoS Comput Biol* 13(5): e1005539.
2. Chan S.H.J., Cai J., Wang L., Simons-Senftle M.N., and Maranas C.D. (2017) Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics*, 33:3603-3609.
3. Chan S.H.J., Wang L., Dash S., and Maranas C.D. (2017) Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints. *arXiv preprint arXiv*:1711.04084

Supported by funding from the U.S. Department of Energy by grant DE-FG02-05ER25684 and *DE-SC0008091*.

A Framework for the Identification of Economically Promising Bio-Based Chemicals

Tony (Wenzhao) Wu[1]* (wwu67@wisc.edu), Matthew Long,[1] Xiaolin Zhang,[1] Jennifer Reed,[1] and **Christos Maravelias**[1]

[1]University of Wisconsin, Madison;

Project Goals

Recent progress in metabolic engineering enables the use of engineered microorganisms for the production of chemicals ("bio-based chemicals")[1,2,3,4]. However, it is still unclear which chemicals have the highest economic prospect. To this end, we develop a screening framework (see **Figure 1**) to identify economically promising ones. Specifically, we first develop a genome-scale constraint-based metabolic modeling approach (based on Flux Balance Analysis - FBA), which is used to identify a candidate pool of 209 chemicals (together with the estimated yield, productivity and residence time for each) from the intersection of the High-Production-Volume (HPV) chemicals[5] and the KEGG and MetaCyc databases. Second, we design 3 screening criteria based on a chemical's profit margin, market volume and market size. The total process cost, including the downstream separation cost, is systematically incorporated into the evaluation[6,7,8]. Third, given the 3 criteria assumed in this work, we identify 32 products which are economically promising if the maximum yields estimated based on the FBA (g product/g glucose) can be achieved, while 22 are promising if the maximum productivities (g product L^{-1} day^{-1}) can be achieved. Comparisons between producing the products extracellularly and intracellularly, as well as between using *E. coli* and *S. cerevisiae* are also discussed (see **Figure 2**). The proposed framework provides important guidance for future studies in the production of bio-based chemicals. It is also flexible in that the databases, yield estimations, and criteria can be modified to customize the screening.

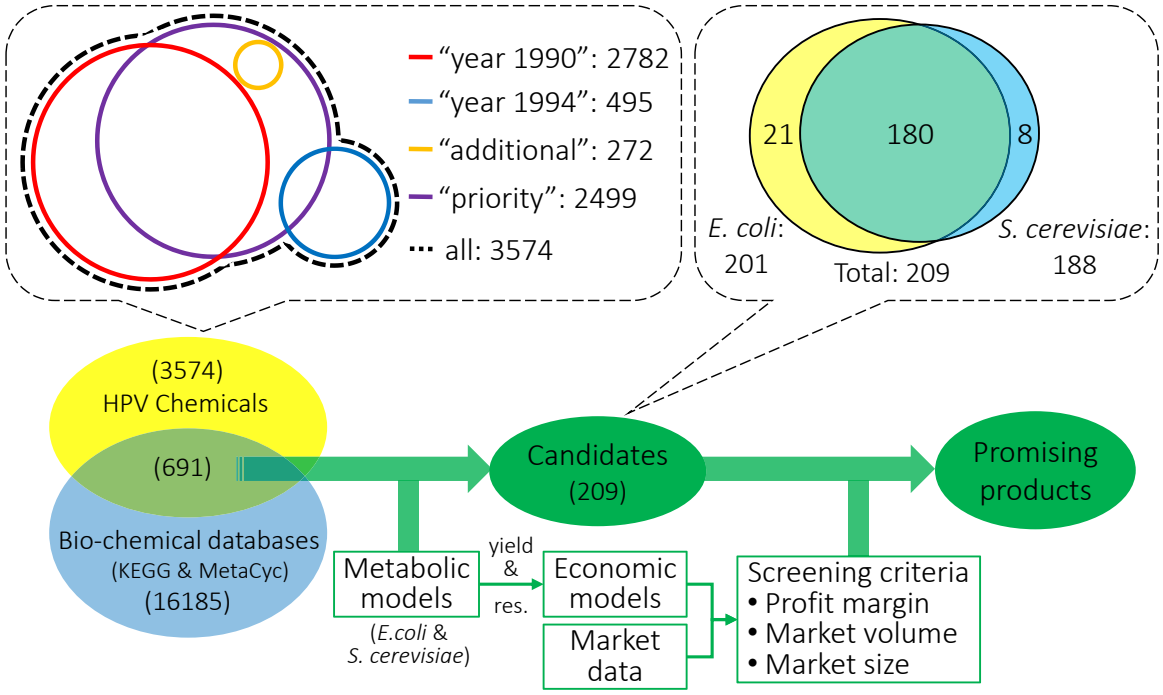


Figure 1. Promising products identification framework, including the compilation of HPV chemicals, identification of the candidate pool and development of the screening criteria. "Res." = residence time.

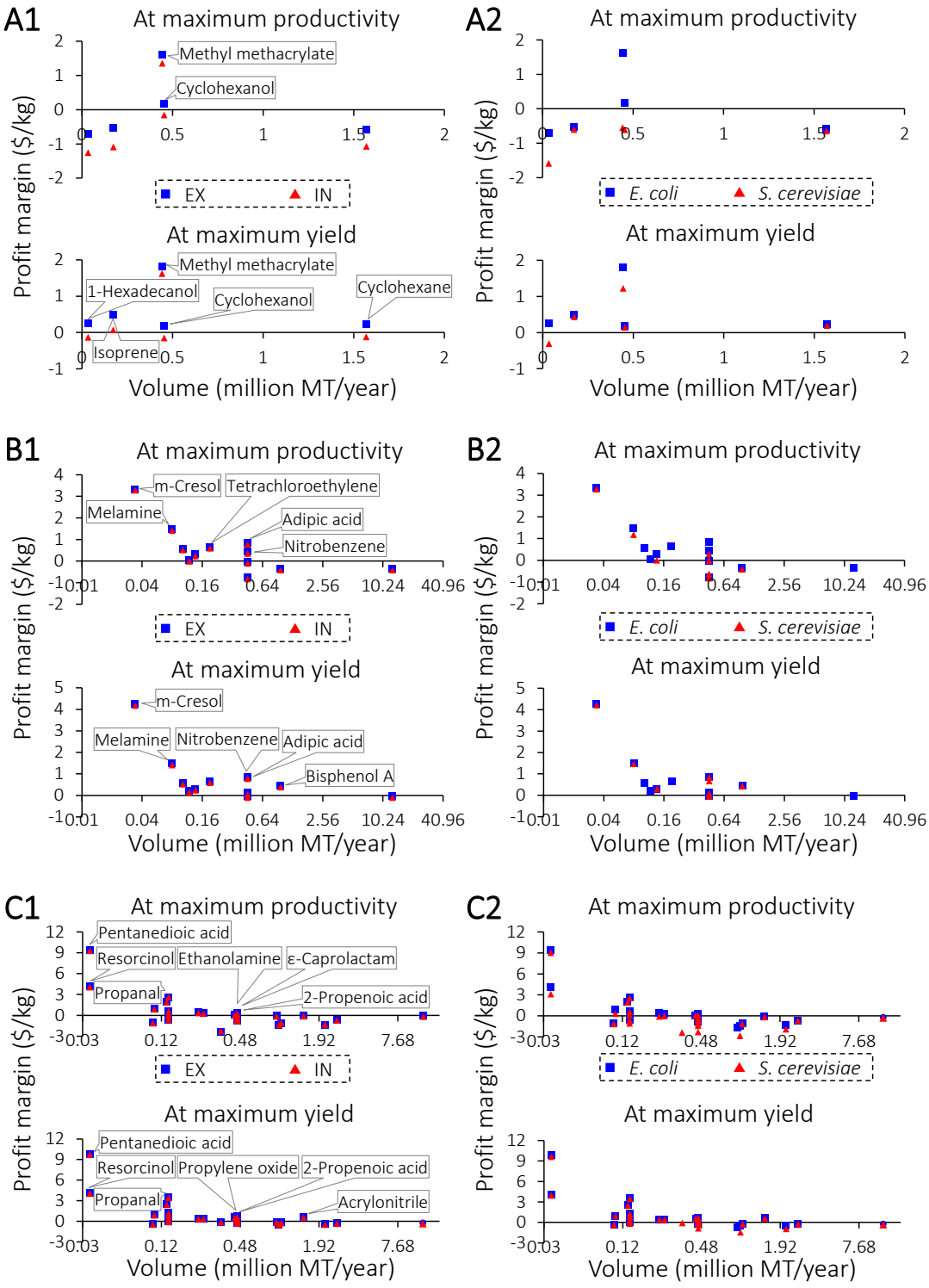


Figure 2. Graphical representation of the screening results at the maximum yield and maximum productivity. (A) Insoluble and light (in terms of density with respect to water) products; (B) insoluble and heavy products; (C) soluble products. A1, B1 and C1 compare EX (extracellular) and IN (intracellular) under the optimal microorganism condition; A2, B2 and C2 compare microorganisms under the optimal localization condition. B1-C2 are plotted on logarithmic scales. Promising products with top 3 profit margins and volumes are labeled in A1, B1 and C1.

References

[1] Gavrilescu, M. & Chisti, Y., 2005. Biotechnology- a sustainable alternative for chemical industry. *Biotechnology advances,* Volume 23, pp. 471-499.
[2] Bornscheuer, U. T. & Nielsen, A. T., 2015. Editorial overview: chemical biotechnology: interdisciplinary concepts for modern biotechnological production of biochemicals and biofuels. *Current Opinion in Biotechnology,* Volume 35, pp. 133-134.
[3] Wilson, S. A. & Roberts, S. C., 2014. Metabolic engineering approaches for production of biochemicals in food and medicinal plants. *Current Opinion in Biotechnology,* Volume 26, pp. 174-182.
[4] Zhang, X., Tervo, C. J. & Reed, J. L., 2016. Metabolic Assessment of E. coli as a Biofactory for Commercial Products. *Metabolic Engineering,* Volume 35, pp. 64-74.
[5] US EPA, 2004. High-production-volume (HPV) chemicals status and future directions of the HPV Challenge Program. *Office of Pollution Prevention and Toxics*, Washington DC.
[6] Wu, W., Yenkie, K.M. & Maravelias, C.T., 2017. A superstructure-based framework for bio-separation network synthesis. *Computers & Chemical Engineering*, Volume 96, pp. 1-17.
[7] Yenkie, K. M., Wu, W., Clark, R. L., Pfleger, B. F., Root, T. W., & Maravelias, C. T., 2017. A roadmap for the synthesis of separation networks for the recovery of bio-based chemicals: Matching biological and process feasibility. *Biotechnology Advances*, Volume 34(8), pp. 1362-1383.
[8] Yenkie, K. M., Wu, W., Maravelias, C.T., 2017. Synthesis and analysis of separation networks for the recovery of intracellular chemicals generated from microbial-based conversions. *Biotechnology for Biofuels*

Funding statement.

This work was funded by National Science Foundation through the Emerging Frontiers in Research and Innovation program (EFRI-1240268), and the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494).

Development and screening of one-pot IL conversion technologies

Gabriella Papa[1,2*] (gpapa@lbl.gov), Jian Sun[1,3], José A. Pérez-Pimienta[4], Victoria Rigual Hernandez[5], John Gladden[1,3]and Blake A. Simmons[1,2]

[1] Deconstruction Division, Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA, USA
[2] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, USA
[3] Biological and Engineering Science Center, Sandia National Laboratories, 7011 East Avenue, Livermore, CA, USA
[4] Laboratorio de Futuros en Bioenergía, Unidad Guadalajara de Ingeniería Avanzada, Centro de Investigación y Estudios Avanzados (CINVESTAV), Zapopan, México; Department of Chemical Engineering, Universidad Autónoma de Nayarit, Tepic, Mexico
[5] Department of Chemical Engineering, Faculty of Chemistry, Complutense University of Madrid, Avda. Complutense s/n, 28040 Madrid, Spain

Project Goal: To assess the capabilities and potential of the "one pot" ionic liquid based [Ch][Lys] configuration for application to different feedstocks.

Certain ionic liquid (ILs) are known to be effective biomass pretreatment solvents that produce high yields of fermentable sugars suitable for the production of biofuels and bioproducts. This poster will cover the discovery and synthesis of new ILs used in new process configurations that reduce the need for solid-liquid separations and water washing of pretreated solids. The development of a "one-pot" system configuration for the integration of pretreatment-saccharification-fermentation unit operations will be a central focus of the poster. This one-pot IL conversion technology is enabled by the use of biocompatible ILs that are not toxic to the enzymes and microbes used in saccharification and fermentation, respectively. The performance of biocompatible ILs, such as the new group of bio-derived cholinium-based ILs is explored in this study. The results provide evidence of the potential of the IL one-pot configuration as a compelling integrated scheme for the deconstruction and conversion of a wide range of lignocellulosic feedstocks. The fundamental understanding generated by this work will contribute to the design and realization of affordable and scalable IL-based biomass conversion technologies.

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.

Engineering the oleaginous yeast *Rhodosporidium toruloides* for the production of lipids and lipid-based chemicals

Jingjing Liu* (jil263@illinois.edu), Shuyan Zhang, Sujit Japtap, Hanna Walukiewicz, Matthew Plutz, and **Christopher V Rao**

Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana-Champaign

https://cabbi.bio/

Project Goal: The goal of this project is to engineer the oleaginous yeast *Rhodosporidium toruloides* for the production of lipids and lipid-based chemicals.

Natural lipids can be used to produce a wide variety of compounds, including fuels (biodiesel), lubricants, surfactants, solvents, waxes, and creams. These lipid-derived compounds, or so-called oleochemicals, potentially offer a renewable alternative to traditional petroleum-based manufacturing. Oleaginous yeast naturally produce lipids from simple sugars when some other essential nutrient such as nitrogen is limiting. In these regards, they potentially offer an economical and renewable route for natural lipid production from low-cost, plant-based sugars

We are investigating *Rhodosporidium toruloides*, a red basidiomycetous yeast, that naturally produces lipids at high titers. This oleaginous yeast produces lipids from a wide variety of common sugars, including glucose, xylose, cellobiose, arabinose, and sucrose. However, far less is known about this oleaginous yeast than the model *Yarrowia lipolytica*. In addition, the genetics are still quite primitive. Despite these limitations, we have made significant progress increasing lipid production in *R. toruloides*. In particular, we have been able to engineer a strain that doubles lipid production during batch growth and quintuples it during fed-batch growth using glucose as a feedstock [1-2]. However, these strains have significantly lower productivities when xylose is the feedstock [3]. To address these lower productivities on xylose, we are also investigating the mechanism of pentose utilization in *R. toruloides*.

References

[1] Zhang S, Ito M, Skerker JM, Arkin AP, Rao CV. Metabolic engineering of the oleaginous yeast *Rhodosporidium toruloides* IFO0880 for lipid overproduction during high-density fermentation. Applied Microbiology and Biotechnology. 2016 Nov;100(21):9393-9405. doi: 10.1007/s00253-016-7815-y.

[2] Zhang S, Skerker JM, Rutter CD, Maurer MJ, Arkin AP, Rao CV. Engineering *Rhodosporidium toruloides* for increased lipid production. Biotechnology and Bioengineering. 2016 May;113(5):1056-66. doi: 10.1002/bit.25864.

[3] Jagtap SS, Rao CV. Production of D-arabitol from D-xylose by the oleaginous yeast Rhodosporidium toruloides IFO0880. Appl Microbiol Biotechnol. 2018 Jan;102(1):143-151. doi: 10.1007/s00253-017-8581-1.

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.

Engineering Bacteria to Produce Branched-chain Fatty Acid Derived Advanced Biofuels

Wen Jiang,[1] Shaojie Wang[1]* (shaojiewang@email.wustl.edu), Wenqin Bai,[1]

and **Fuzhong Zhang**[1,2]

[1]Department of Energy, Environmental and Chemical Engineering,

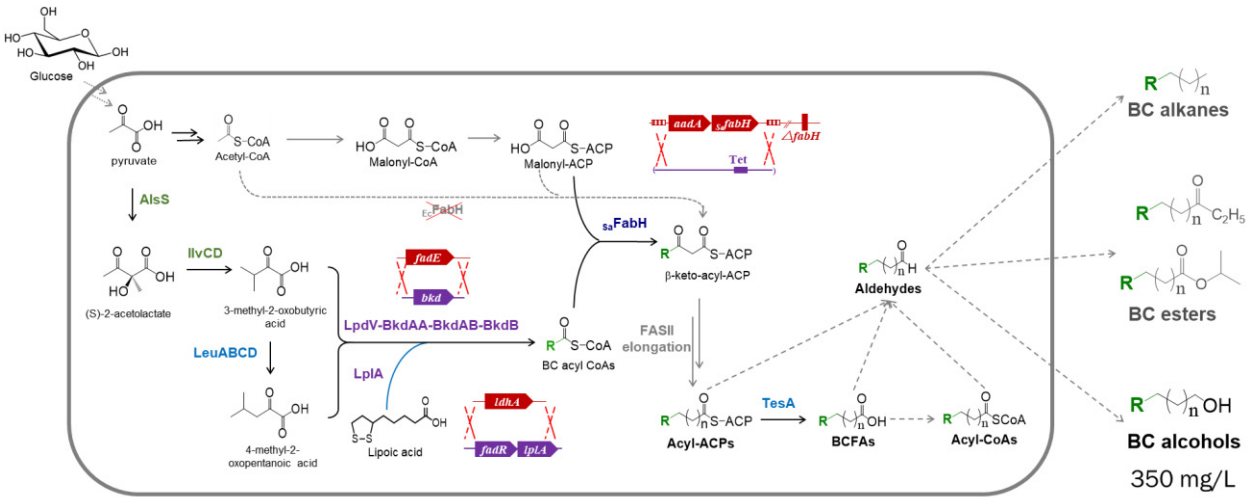
[2]Division of Biological & Biomedical Sciences,

Washington University in St. Louis, Saint Louis, MO 63130, USA

Project Goals: Extensive research has focused on engineering the fatty acid (FA) biosynthetic pathway for biofuel production because free FAs can be converted to fatty acid esters and alkanes that have similar energy content and cetane number to that of petroleum-derived diesel fuel. Most biological systems naturally produce straight-chain fatty acids (SCFAs). However, diesel fuels derived from SCFAs have undesirable freezing points and cold flow, limiting their practical use in low temperatures. On the other hand, petroleum-derived diesel contains a significant amount of branched hydrocarbons, which offer better cold flow properties. The overall goal of this project is to develop microbial systems for the production of branched-chain fatty acids (BCFAs) and their derived biofuels with improved physical and combustion properties. Our previous work focused on testing novel BCFA pathways in model bacterium E. coli. Our current aim is to transfer the pathway to a non-model bacterial species *Rhodococcus opacus* to enable production of BCFA-derived advanced biofuels from lignocellulose.

Abstract text.

BCFAs are important precursors of advanced biofuels with improved cold-flow properties. We developed metabolic pathways and a series of strategies to produce BCFAs and their derived biofuels in *E. coli* in high percentage. We first replaced the acetyl-CoA-specific *E. coli* FabH with one of branched-chain-acyl-CoA specific FabHs. Screening the most active branched-chain-acyl-CoA specific FabHs resulted in a 81-fold enhancement in BCFA-production compared to a strain containing *E. coli* FabH[1]. Next, we found that the position of the branch can be controlled by changing the metabolic pool of branched-chain α-keto acids. Supplementing different α-keto acids allowed us to produce specific BCFAs in high purities[1]. Meanwhile, we discovered a key bottleneck in BCFA production: overexpression of BKD is toxic to *E. coli* due to the depletion of protein lipoylation capacity. Engineering a complementary protein lipoylation pathway alleviated the toxicity and improved BCFA production to 276 mg/L and 85% of total free fatty acids[2]. Finally, an α-keto acid biosynthetic pathway was engineered and coupled with the rest of the BCFA pathway, the resulting strain produced BCFAs from glucose at 181 mg/L and 72% of total FFA[2].



Using this system, we introduced and tested four different downstream pathways that convert BCFAs to branched long-chain fatty alcohols (BLFLs). Furthermore, by engineering the α-keto acid biosynthetic pathways and balancing the expression levels of three different modules, we achieved BLFL titers of up to 350 mg/L from glucose, and BLFL percentages up to 79%[3]. We are currently exploring our engineering strategy in *R. Opacus* PD630 to produce BCFA-derived advanced biofuels from lignocellulose.

References

1. Jiang W., Jiang Y., Bentley G.J., Liu D., Xiao Y., Zhang F., Enhanced production ofbranched-chain fatty acids by replacing β-ketoacyl-(acyl-carrier-protein) synthase III (FabH).*Biotechnol Bioeng*. 112, 1613, (2015).

2. Bentley G.J., Jiang W., Guamán L.P., Xiao Y., Zhang F., Engineering *Escherichia coli* to produce branched-chain fatty acids in high percentages. *Metab Eng*. 38, 148, (2016).

3. Jiang W., Qiao J.B., Bentley G.J., Liu D., Zhang F., Modular pathway engineering for the microbial production of branched-chain fatty alcohols. Biotechnology for Biofuels. 10:244, (2017)

This work is currently funded by the Department of Energy [DE-SC0018324].

Advances in Lignin Valorization via Biological and Catalytic Transformations

Gregg T. Beckham[1,2*](gregg.beckham@nrel.gov), Eric M. Anderson[2,3], Bryon S. Donohoe[1,2], Christopher W. Johnson[1,2], Sandra Notonier[1,2], Davinia Salvachúa[1,2], Michael L. Stone[2,3], Adam M. Guss[2], Yuriy Román-Leshkov[2,3], and **Gerald A. Tuskan**[2]

[1]National Renewable Energy Laboratory Golden, CO; [2]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN; and [3]Massachusetts Institute of Technology, Cambridge, MA

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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

In most biorefinery designs, lignin is slated to be burned for heat and power because its inherent heterogeneity and recalcitrance make it difficult to valorize. Despite nearly a century of lignin depolymerization research, most catalytic strategies to break down lignin yield a heterogeneous slate of aromatic compounds, which makes purification of single, high-yielding co-products from lignin quite daunting.

Reductive catalytic fractionation (RCF) has emerged as a leading biomass fractionation and lignin valorization strategy. We will highlight recent work using flow-through reactors to investigate RCF of poplar. A new flow-based RCF process enables the acquisition of intrinsic kinetic and mechanistic data essential to accelerate the design, optimization, and scale-up of RCF processes. We examined time-resolved product distributions and yields obtained from experiments with different catalyst loadings to identify and deconvolute events during solvolysis and hydrogenolysis. Multi-bed RCF experiments provided unique insights into catalyst deactivation, showing that leaching, sintering, and surface poisoning are causes for decreased catalyst performance. The onset of catalyst deactivation resulted in higher concentrations of unsaturated lignin intermediates and increased occurrence of repolymerization reactions, producing high molecular weight species. This initial study demonstrates the concept of flow-through RCF, which will be vital for scale-up of this promising approach as well as enabling the survey of many biomass samples to understand how lignin chemistry affects RCF yields, such as those from the Center for Bioenergy Innovation's Genome-Wide Association Study (GWAS) libraries, which will be pursued in future work.

Additionally, some microbes have evolved catabolic pathways that enable the utilization of lignin-derived aromatic molecules as carbon and energy sources. Aromatic catabolism most commonly occurs via Upper Pathways that act as a "biological funnel" to convert heterogeneous lignin-derived substrates to Central Intermediates, such as syringate, protocatechuate, or catechol. These compounds subsequently undergo oxidative ring cleavage and are further converted to central carbon metabolism. In the Center for Bioenergy Innovation, we employ the robust aromatic-catabolic microbe, *Pseudomonas putida* KT2440, to understand, harness, and expand these powerful metabolic pathways to convert both aromatic model compounds and heterogeneous, lignin-enriched streams to value-added compounds such as ß-ketoadipate. Here, we will present several recent insights into lignin depolymerization and aromatic catabolism by *P. putida* KT2440.

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.

Microbial Valorization of Lignin: Using Bacteria and Their Enzymes to Develop Biological Systems for Depolymerizing Lignin

Wayne S. Kontur[1,2*] (wskontur@wisc.edu), Daniel L. Gall[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 Goals: As part of the Great Lakes Bioenergy Research Center, we are working on developing biological systems for converting lignin from plant biomass into valuable chemicals. Lignin is a hetero-polymer of aromatic units that makes up ~25% of lignocellulosic plant biomass. While the individual aromatic compounds that make up lignin could have potential value for several industries, lignin is difficult to break down. One aspect of our work is studying bacteria and their enzymes for development into systems to help depolymerize lignin. In this work, we have identified the soil bacterium *Novosphingobium aromaticivorans* as being one of the best organisms thus far characterized at breaking the β-aryl ether bond, which is the most common linkage between aromatic units in natural lignin. We have used *N. aromaticivorans* to identify and characterize enzymes involved in breaking the β-aryl ether bond, and have used enzymes from *N. aromaticivorans* and other bacteria to develop an *in vitro* system for depolymerizing bona fide plant-derived lignin.

There is economic and environmental interest in using renewable resources as raw materials for production of chemicals that are currently derived from fossil fuels. Lignin, a hetero-polymer composed of several aromatic subunits, can make up ~25% (dry weight) of vascular plant cell walls (1), making it one of the most abundant renewable organic materials on Earth. In its polymeric form, lignin has limited economic uses- it is generally discarded or burned to generate energy- but the aromatic compounds that make up lignin could potentially be used in the biofuel, chemical, cosmetic, food, and pharmaceutical industries (2). However, due largely to its irregular, covalently bonded structure, lignin has historically been difficult to depolymerize. Consequently, intensive efforts are currently aimed at developing chemical, enzymatic, and hybrid methods for deriving simpler and lower molecular weight products from lignin (2).

While the aromatic units in lignin are linked together via several classes of covalent bonds, the β-aryl ether (β-O-4) bond typically constitutes >50% of all the linkages (3), making it an important target in developing systems for depolymerizing lignin. The β-etherase pathway, found in some sphingomonad bacteria, is a promising biological route for cleaving the β-aryl ether bond. In this metabolic pathway, the ether bond between aromatic units is broken by replacement with a thioether bond involving glutathione (2). Several species are known that can

perform this pathway, and many of the enzymes required for the reactions of the pathway have been identified and characterized.

We tested several additional bacteria and found that *Novosphingobium aromaticivorans* can cleave the β-aryl ether bond of a dimeric aromatic compound at one of the fastest rates thus far reported, making it a good candidate for using as a cellular chassis for lignin depolymerization systems. We have also used *N. aromaticivorans* to study various aspects of the β-etherase pathway, including identifying the enzymes involved in the pathway. Through this work, we identified a previously uncharacterized enzyme that is integral to the pathway: a glutathione-S-transferase that removes glutathione from both stereoisomers of one of the intermediates in the pathway. This work has important implications for not only lignin depolymerization, but also for understanding glutathione-S-transferases and the reactions that they can catalyze more broadly.

While *N. aromaticivorans* could be used to cleave the β-aryl ether bond of plant-derived lignin oligomers, there may be limits to the size of the oligomers that can be transported into the cells for cleavage. We have therefore also developed an *in vitro* system using the minimal set of purified enzymes necessary to break the racemic β-aryl ether bond. Along with β-etherase pathway enzymes from *N. aromaticivorans* and a similar bacterium (*Sphingobium* sp. SYK-6), our system includes a glutathione reductase from *Allochromatium vinosum*, which recycles the cofactors NAD^+ and glutathione required for the pathway, thus limiting the expense of the system by eliminating the need to continuously supply fresh cofactors. We tested this system on bona fide plant-derived lignin and successfully cleaved monomeric aromatic compounds from the lignin polymer. This work is important for the use and future optimization of these bacterial enzymes for industrial level biotechnological applications designed to derive high-value monomeric aromatic compounds from lignin.

References

1. U.S. DOE (2015) Lignocellulose Biomass for Advanced Biofuels and Bioproducts: Workshop Report, DOE/SC-0170. U.S. Department of Energy Office of Science. Available at: http://genomicscience.energy.gov/biofuels/lignocellulose/

2.Gall DL, Ralph J, Donohue TJ, Noguera DR (2017) Biochemical transformation of lignin for deriving valued commodities from lignocellulose. Curr Opin Biotechnol 45, 120‑126.

3. Adler E: Lignin chemistry - past, present and future. *Wood Science and Technology* 1977, **11**:169-218.

This work was supported by a U.S. Department of Energy (DOE) Great Lakes Bioenergy Research Center grant (DOE Office of Science BER DE-FC02-07ER64494).

SbCOMT (Bmr12) is involved in the biosynthesis of tricin-lignin in sorghum

Aymerick Eudes[1,2*](ageudes@lbl.gov), Tanmoy Dutta[1,3], Kai Deng[1,4], Nicolas Jacquet[1,5], Anagh Sinha[1,2], Veronica Teixeira Benites[1,7], Edward E. K. Baidoo[1,7], Scott E. Sattler[8], Trent R. Northen[1,2,9], Seema Singh[1,3], Blake A. Simmons[1,7], and **Dominique Loqué**[1,2]

[1]Joint BioEnergy Institute, EmeryStation East, 5885 Hollis St, 4th Floor, Emeryville, CA 94608, USA

[2]Environmental Genomics & Systems Biology, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

[3]Biomass Science and Conversion Technology Department, Sandia National Laboratories, Livermore, CA 94551, USA

[4]Biotechnology and Bioengineering Department, Sandia National Laboratories, Livermore, CA 94551, USA

[5]Laboratory of Biological and Industrial Chemistry, University of Liege, Gembloux Agro-Bio Tech, Passage des Déportés 2, B-5030 Gembloux, Belgium

[6]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

[7]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

[8]Wheat, Sorghum, and Forage Research Unit, USDA-ARS, Lincoln, NE 68583, USA

[9]Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek CA, 94958, USA

Project Goals: The most abundant organic material on earth is lignocellulosic biomass or non-food plant material. JBEI's mission is to convert biomass to biofuels and bioproducts. The goal is to provide the nation with clean, renewable transportation fuels and chemicals. Building a successful lignocellulosic biofuels industry depends, in part, on developing specialized bioenergy crops or feedstocks that are optimized for deconstruction and conversion. Understanding the molecular basis underlying the synthesis of lignocellulosic biomass is a prerequisite to achieve these goals.

https//www.jbei.org/

Abstract: Lignin in plant biomass represents a target for engineering strategies towards the development of a sustainable bioeconomy. In addition to the conventional lignin monomers, namely *p*-coumaryl, coniferyl and sinapyl alcohols, tricin has been shown to be part of the native lignin polymer in certain monocot species. Because tricin is considered to initiate the polymerization of lignin chains, elucidating its biosynthesis and mechanism of export to the cell wall constitute novel challenges for the engineering of bioenergy crops. Late steps of tricin biosynthesis require two methylation reactions involving the pathway intermediate selgin. It has

recently been demonstrated in rice and maize that caffeate *O*-methyltransferase (COMT) involved in the synthesis syringyl (S) lignin units derived from sinapyl alcohol also participates in the synthesis of tricin *in planta*. In this work, we validate in sorghum (*Sorghum bicolor* L.) that the *O*-methyltransferase responsible for the production of S lignin units (SbCOMT / Bmr12) is also involved in the synthesis of lignin-linked tricin. In particular, we show that biomass from the sorghum *bmr12* mutant contains lower level of tricin incorporated into lignin, and that SbCOMT can methylate the tricin precursors luteolin and selgin. Our genetic and biochemical data point toward a general mechanism whereby COMT is involved in the synthesis of both tricin and S lignin units.

PLoS ONE 12(6): e0178160. https://doi.org/10.1371/journal.pone.0178160

Funding statement.
This work was part of the DOE Joint BioEnergy Institute (http:// www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

Development of a Genetic Toolkit in *Rhodococcus opacus* PD630 for Lignin Valorization

Drew M. DeLorenzo[1*] (drew.delorenzo@wustl.edu), William R. Henson[1], Austin G. Rottinghaus[1], **Yinjie J. Tang**[1], **Marcus Foston**[1], **Gautam Dantas**[2], and **Tae Seok Moon**[1,3+]

[1] Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, MO, USA; [2] Department of Pathology and Immunology, Center for Genome Sciences and Systems Biology, Department of Biomedical Engineering, and Department of Molecular Microbiology, Washington University in St. Louis, St. Louis, MO, USA; [3] Division of Biology and Biomedical Sciences, Washington University in St. Louis, St. Louis, MO, USA

* Presenting author

+ Corresponding author

http://moon.eece.wustl.edu/

Project Goals: The overall goal of this project is to interrogate the metabolic networks and genetic regulation that control the utilization of and tolerance to thermochemically depolymerized lignin, focusing on phenolics, in *R. opacus*.

Lignocellulosic biomass is an abundant and renewable feedstock that can be converted into biofuels and commodity chemicals using microorganisms. An ideal organism for biomass conversion should possess capabilities for consumption of both cellulose-based sugars and toxic lignin-derived aromatic compounds. Additionally, a candidate microbe should be tolerant to growth inhibiting compounds generated during lignocellulose depolymerization, genetically tractable, and demonstrate a rapid growth rate. The Actinomycetales *Rhodococcus opacus* PD630 (*R. opacus*) is a gram-positive microbe, known for high levels of triacylglycerol accumulation, which meets these biomass conversion qualifications. To enhance its innate aromatic-degrading capacity, we applied adaptive evolution, a growth-based strain selection method, by sequentially sub-culturing cells in diverse combinations of lignin-derived aromatic compounds as sole carbon sources. Our adapted strains demonstrated higher growth rates and higher lipid accumulation compared to the wild type strain. Whole genome sequencing, RNA-seq, targeted metabolomics, and ^{13}C-fingerprinting analysis have identified possible aromatic tolerance and utilization mechanisms such as upregulation of degradation pathways and putative transporters for aromatic compounds.

Despite our increased understanding of aromatic tolerance and utilization mechanisms, few genetic elements and parts have been directly characterized in *R. opacus,* limiting its future

industrial applications. To enable *R. opacus* as a future chassis for biomass conversion, we have developed a suite of genetic tools for reliable and predictable gene expression: **1)** six fluorescent reporters in three distinct wavelength ranges for quantifying promoter output, **2)** a constitutive promoter library spanning a 45-fold change in fluorescence output, **3)** three chemically inducible promoters for tunable gene expression, **4)** a dynamic metabolite sensor that detects ammonium concentration, **5)** a collection of metabolite sensors that detect various aromatic compounds, and **6)** a recombinase-based system for integration of exogenous DNA into newly identified neutral sites within the genome. Additionally, a CRISPR interference system for targeted gene repression has been developed and a set of stable reference genes for RT-qPCR have been identified. Overall, this work expands the ability to control and characterize gene expression in *R. opacus* and is a critical first step towards future fuel and chemical production in this host.

Publications

1. DM DeLorenzo, WR Henson and TS Moon. Development of Chemical and Metabolite Sensors for *Rhodococcus opacus* PD630. *ACS Synth. Biol.* 6, 1973–1978 (2017)

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. 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)

This work was supported by U.S. Department of Energy grants: DE-SC0012705 and DE-SC0018324.

Engineering Anaerobic Gut Fungi for Lignocellulose Breakdown

John K. Henske[1], Sean P. Gilmore[1], St. Elmo Wilken[1], Igor Podolsky[1], Susanna Seppala[1], **Michelle A. O'Malley (momalley@engineering.ucsb.edu)[1]***

[1]Department of Chemical Engineering, University of California, Santa Barbara

http://omalleylab.com

Project Goals: The goal of this project is to engineer anaerobic gut fungi as novel platform organisms for biofuel production from plant material. To accomplish this goal, a panel of anaerobic fungi will be isolated from different herbivores and screened for their ability to degrade lignocellulose. The basic metabolic networks that govern lignocellulose hydrolysis within anaerobic fungi will also be determined, and models will be generated to describe how important enzyme groups are coordinated during breakdown. Using this information, genetic transformation strategies to manipulate gut fungi will be developed, which would endow them with enhanced functionality against a range of industrially relevant substrates. Collectively, this information will establish the molecular framework for anaerobic fungal hydrolysis, and will guide in the development of lignocellulosic biofuels.

Anaerobic fungi are the primary colonizers of biomass within the digestive tract of large herbivores, where they have evolved unique abilities to break down lignin-rich cellulosic biomass through invasive, filamentous growth and the secretion of powerful lignocellulolytic enzymes and enzyme complexes (cellulosomes). Despite these attractive abilities, considerably less genomic and metabolic data exists for gut fungi compared to well-studied anaerobic bacteria and aerobic fungi that hydrolyze cellulose. This presents a significant knowledge gap in understanding gut fungal function, substrate utilization, and metabolic flux, which has prohibited the genetic and functional modification of gut fungi. Our approach combines next-generation sequencing with physiological characterization to establish the critical knowledge base to understand lignocellulose breakdown by gut fungi.

We worked with collaborators at the Joint Genome Institute (JGI) to sequence genomes of the anaerobic fungi *Neocallimastix californiae*, *Anaeromyces robustus*, and *Piromyces finnis*, which were isolated from large herbivores. Our recent publication in Nature Microbiology released the first high-quality genomes of the anaerobic gut fungi (results available on Mycocosm). With the advent of PacBio sequencing, long reads averaging 10,000bp in length were sufficient to remedy the assembly of highly repeat-rich regions of the genomes that could not be resolved with Illumina-based approaches. These results also enabled further epigenetic analysis of the fungal genomes (ongoing collaboration with Igor Grigoriev's group at JGI), that highlighted regions of the genomes that are highly transcribed (Mondo, et al, Nature Genetics).

In previous years of this project, we established that gut fungi secrete a large number of cellulolytic enzymes, which form cellulosomes that may be physically attached to the cell. However, the enzyme components, modular assembly mechanism, and functional role of fungal cellulosomes during biomass breakdown remained unknown. Although the basic interaction of these complexes is a modular cohesion-dockerin binding similar to that observed in bacteria, the fungal dockerin and scaffoldin domains have no similarity to their bacterial counterpart. Previously obtained genomic and transcriptomic data for three fungal strains isolated by our lab enabled us to identify a "parts list" for fungal cellulosomes, including a set of novel scaffolding proteins that biochemically interact with dockerin-fused enzymes from fungi (Haitjema, 2017 *Nature Microbiology*). Further, our previous work has revealed that (i) only some of the scaffoldin proteins have putative transmembrane helix domains, (ii) many of the dockerin domain proteins (DDPs) do not identify as carbohydrate active enzymes (CAZy) and (iii) fungi still secrete up to 50% of their CAZymes as free enzymes (lacking a dockerin domain) (Haitjema, 2017 *Nature Microbiology*). Finally, comparative genomics against anaerobic bacteria and other strains of anaerobic fungi revealed that a number of CAZyme domains identified in the fungal genomes originated from bacteria. This finding indicates that fungi and bacteria that co-exist in the rumen of large herbivores may have exchanged genetic information, allowing for the optimization of biomass degradation in these anaerobic systems. This finding is surprising, as horizontal gene transfer in this respect is typically observed to occur within life forms that share the same kingdom of life – it is seldom described to occur from prokaryotes to eukaryotes.

Building from these insights, we developed a biphasic fermentation scheme that combines the lignocellulolytic action of anaerobic fungi with domesticated microbes for bioproduction. When grown in batch culture, anaerobic fungi release excess sugars from both cellulose and crude biomass due to a wealth of highly expressed carbohydrate active enzymes (CAZymes), converting as much as 49% of cellulose to free glucose. This sugar-rich hydrolysate readily supports growth of *S. cerevisiae*, which can be engineered to produce a range of value-added chemicals. Further, reconstruction of metabolic pathways from transcriptomic data reveals that anaerobic fungi do not catabolize all sugars that their enzymes hydrolyze from biomass, leaving other carbohydrates such as galactose, arabinose, and mannose available as nutritional links to other microbes in their consortium. Overall, these results suggest that anaerobic fungi provide a nutritional benefit to the rumen microbiome, which can be harnessed to design synthetic microbial communities that compartmentalize biomass degradation and bioproduct formation.

This Project is supported by the Office of Biological and Environmental Research through the DOE Office of Science (Early Career Program)

Impact of Hydration and Temperature History on the Structure and Dynamics of Lignin

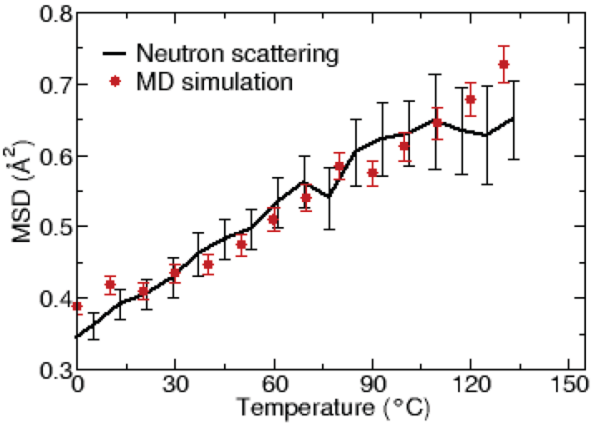
Loukas Petridis[1]* (petridisl@ornl.gov),[1] Derya Vural,[1] Hugh M. O'Neill,[1] Yunqiao Pu,[1] Sai Venkatesh Pingali,[1] Volker Urban,[1] Arthur J. Ragauskas,[1] Jeremy C. Smith,[1] and **Brian H. Davison**[1]

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

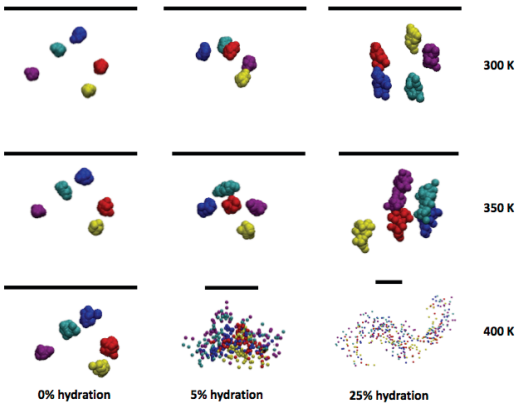
http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if cost-effective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

The full utilization of plant biomass for the production of energy and novel materials often involves high temperature treatment. Examples include the relocalization of lignin to increase the accessibility of cellulose for production of biofuels and the melt spinning of lignin for manufacturing low-cost carbon fiber. These temperature-induced effects arise from poorly understood changes in lignin flexibility. At the molecular scale this is achieved by enhancing the underlying atomic dynamics. Here, we combine molecular dynamics simulations with neutron scattering and dielectric spectroscopy experiments to probe the dependence of lignin dynamics on hydration and thermal history. Excellent agreement is found between the experiments and the simulations. We find a dynamical and structural hysteresis: at a given temperature, the lignin molecules are more expanded and their dynamics faster when the lignin is cooling than when heating. The structural hysteresis is more pronounced for dry lignin. The difference in dynamics, however, follows a different trend, being more significant at high temperatures and high hydration levels. The simulations also reveal syringyl units to be more dynamic than guaiacyl and the aliphatic chains to be more dynamic than the phenolic rings. The results provide an atomic-detailed description of lignin dynamics and suggest that the lignin glass transition temperature, at which the polymer softens, will be lower in the more hydrated environment of plant biomass than in purified lignin powder. This lignin glass transition is a critical stage in lignin reclocalization or in lignin aggregation. Furthermore, the dynamical hysteresis found here implies the softening temperature is lower when the lignin is cooled than when heated, therefore extending the cooling phase of processing may offer ways to lower the processing cost of lignin.



Comparison of lignin mean square displacement measured from quasi-elastic neutron scattering experiments and calculated from molecular dynamics simulations.



Snapshots of the simulations, showing lignin becoming more dynamic with increasing hydration and temperature.

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.

Engineering CRISPR/Cas9 platform to industrialize lignin modifying enzymes (LMEs) using *Aspergillus niger*

Laure Leynaud-Kieffer[1,2,3*], Vassily Hatzimanikatis[3], and **Blake Simmons [1,2]**

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Swiss Federal Institute of Technology, Lausanne, Switzerland.

http://www.jbei.org

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

The conversion of the biomass into advanced biofuels faces many challenges, one of which is finding the right organism for the job. The filamentous fungus *Aspergillus niger* has been chosen as a biocatalyst for cellulose, hemicellulose, and lignin degradation because it could secrete numerous hydrolytic enzymes, such as lignin modifying enzymes (LMEs), it is genetically tractable, and its genome sequence is available.

However, we currently lack efficient tools for editing and augmenting the *A. niger* genome. While genome editing techniques such as CRISPR/Cas9 editing function in *A. niger*, we are limited by the difficulty of making multiple mutations, restricted selection of markers, and inefficient, expensive and time-consuming methodologies for genome engineering.

Here I present progress towards developing a method for efficiently making multiple genomic mutations via Cas9/gRNAs without the use of permanent selective markers. This technique utilizes several approaches; 1) pyrG positive and negative selection by a transient donor DNA, 2) a sgRNA transcript in vitro for a more user-friendly method. Once complete, this strategy should (1) remove the need for laborious screening of colonies to identify mutants, (2) permit the rapid engineering of strains with multiple mutations without the need for multiple selection markers. Our objective is to first establish this method for genome engineering. Next, we will build a library of *A. niger* strains housing pathways for lignin degradation. Then we will optimize the technique of fermentation for *A. niger* in submerged bioreactor. The overall objective is to define the best strains and conditions for the productivity of LMEs in bioreactors at a pilot scale for industries.

Funding statement.

This work was part of the DOE Joint BioEnergy Institute (http:// www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

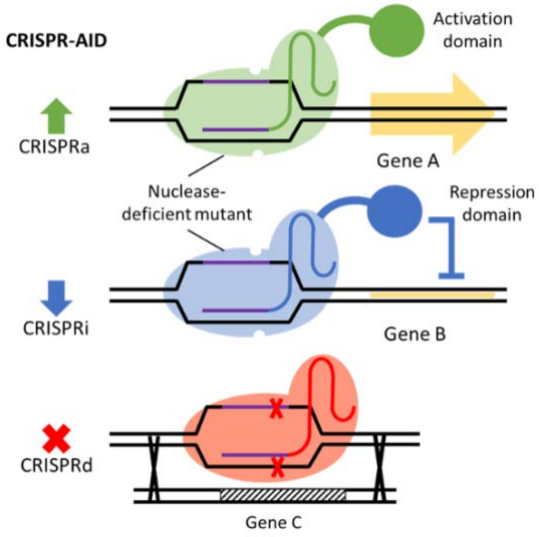
Combinatorial Metabolic Engineering Using an Orthogonal Tri-functional CRISPR System

Jiazhang Lian[1,2], Carl Schultz[1]* (jcschul2@illinois.edu), Mohammad HamediRad,[1] Sumeng Hu,[1] and **Huimin Zhao**[1]

[1]University of Illinois at Urbana-Champaign, Urbana IL 61801; [2]Zhejiang University, Hangzhou 310027, China

Project Goals: We aim to enable access to a complete range of expression profiles for any gene or combination of genes in the yeast *Saccharomyces cerevisiae* by using a tri-functional CRISPR/Cas system to perform simultaneous, multiplexed gene activation, interference, and deletion. This will enable metabolic engineering to systematically optimize phenotypes of interest through a combination of gain, decrease, and complete loss of functional mutations.

Designing an optimal microbial cell factory typically requires overexpression, knock-down, and knock-out of multiple gene targets. Unfortunately, such rewiring of cellular metabolism is often carried out sequentially and with low throughput. We report a combinatorial metabolic engineering strategy utilizing an orthogonal tri-functional CRISPR system that combines transcriptional activation, transcriptional interference, and gene deletion (CRISPR-AID) in the yeast *Saccharomyces cerevisiae*. This strategy enables multiplexed perturbation of the metabolic and regulatory networks in a modular, parallel, and high throughput manner.



To implement this system, three orthogonal Cas proteins were utilized: dLbCpf1 fused to a transcriptional activator, dSpCas9 fused to a transcriptional repressor, and SaCas9 for gene deletion. Deletion was accomplished by the introduction of a 20 bp frame-shift mutation using a homology donor on the guide RNA expression vector.

As a proof of concept, we demonstrate the application of CRISPR-AID to increase the production of β-carotene by 3-fold in a single step through the simultaneous activation of *HMG1*, interference of *ERG9*, and deletion of *ROX1*. Additionally, we effected a 2.5-fold improvement in the display of an endoglucanase on the yeast surface by optimizing a 15x18x6 matrix of metabolic engineering targets (activation, interference, and deletion, respectively) in a combinatorial manner.

Publications

1. Lian, J., Hamedirad, M., Hu, S., & Zhao, H. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nature Communications* **8**, 1-9 (2017).

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-SC0018260.

Construction of a Kinetic Model for Yeast Lipid Metabolism

Shekhar Mishra[1]* (smishr10@illinois.edu), Costas Maranas[2], **Huimin Zhao**[1,2]

[1]Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

[2]Department of Chemical Engineering, Penn State University, University Park, PA 16802, United States

[3]Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

Project Goals:
To construct a kinetic model to elucidate the metabolic and regulatory influences of sphingolipid metabolism in *Saccharomyces cerevisiae* on fatty acid pathways.

Abstract:
Kinetic models consisting of ordinary differential equations offer a comprehensive method to analyze and predict the dynamics of cellular metabolism. However, their use often poses a challenge due to uncertainties in parameter estimates and incomplete information on gene regulation processes. Here we develop a kinetic model of yeast lipid metabolism, specifically consisting of the pathways involved in sphingolipid metabolism and fatty acid synthesis and degradation. The model has been adapted from a previously published kinetic model (Savoglidis et al., 2016), and includes new pathways on fatty acid synthesis, as well as gene regulatory information wherever possible. A preliminary analysis of the parameter estimation process was used to highlight parameters with high uncertainty in estimated values. This can be used to aid in the design of perturbation experiments for further reducing the uncertainty in the model. Metabolic Control Analysis (MCA) was performed on the kinetic model to yield the pathway enzymes that exert strong influence on reaction fluxes and metabolite concentrations. Analysis of metabolic control can inform useful targets for metabolic engineering of target lipid overproduction.

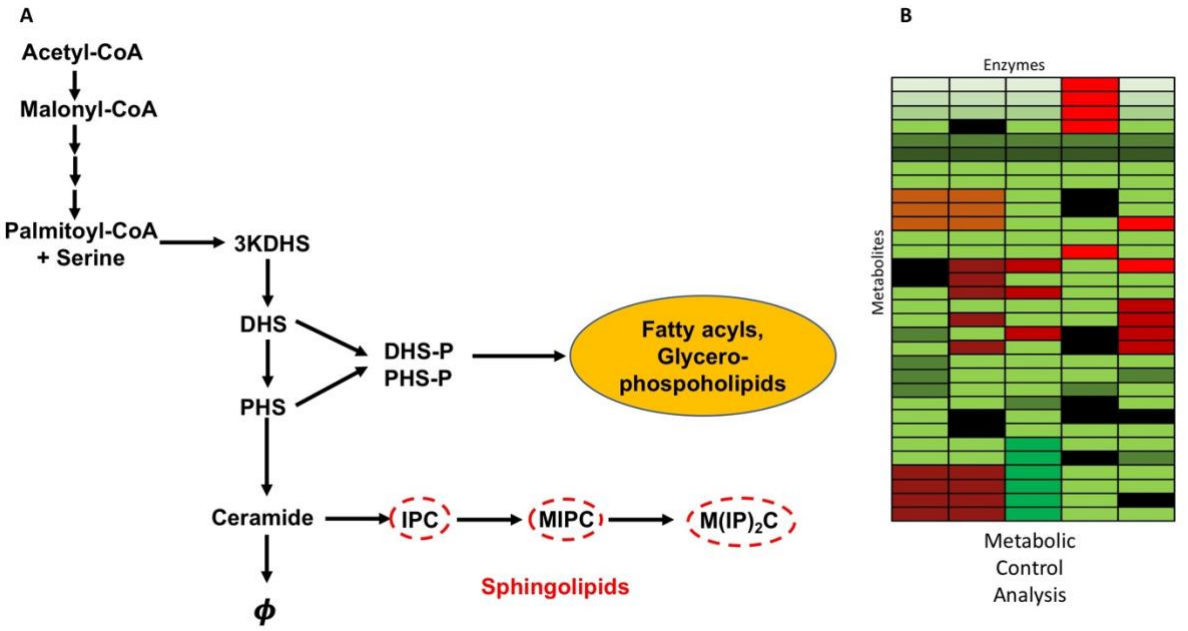


Figure 1. (A) Reaction map of sphingolipid metabolism depicting important compounds as well as pathways connecting sphingolipid intermediates to fatty acid and glycerophospholipid pathways. (B) Diagram illustrating the results of metabolic control analysis (MCA) as a heat map highlighting the enzymes that exert significant control over metabolites.

References:

Savoglidis, G., da Silveira Dos Santos, A. X., Riezman, I., Angelino, P., Riezman, H., Hatzimanikatis, V., 2016. A method for analysis and design of metabolism using metabolomics data and kinetic models: Application on lipidomics using a novel kinetic model of sphingolipid metabolism. Metab Eng. 37**,** 46-62.

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-SC0018260.

Construction of genome-scale metabolic models for non-model yeast organisms for biofuel and bioproduct engineering

Siu H. J. Chan (sxc554@psu.edu), Hoang Dinh, Saratram Gopalakrishnan, and Costas D. Maranas

Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802

http://www.maranasgroup.com/

Project Goals: This project aims to reconstruct the genome-scale metabolic models for two non-model yeast organisms, *Rhodosporidium toruloides* and *Issatchenkia orientalis*, which are industrially potential hosts for producing biofuels and organic acids. Using the metabolic model of the model yeast *Saccharomyces cerevisiae* as reference and multiple sources of data, the models for the two organisms were constructed following a rapid reconstruction workflow. The reconstructed metabolic models and the associated atom mapping will act as an important tool for kinetic modeling, ^{13}C metabolic flux analysis, identification of metabolic regulation and integration of physiological and omics data, facilitating the genome-scale analysis and engineering of the two yeast organisms for production of biofuels and bioproducts.

Yeast organisms are promising engineering targets for microbial cell factories with unique industrial advantages compared to bacteria such as higher tolerance against inhibitory compounds and contamination. However, most metabolic engineering and modeling efforts were focused on the model yeast *Saccharomyces cerevisiae* despite the discovery of other non-model yeast organisms with unique biochemical production capabilities. Among them, *Rhodosporidium toruloides* is able to ferment lignocellulose into lipids and can be a potential platform for the production of fatty-acid-derived biofuels. *Issatchenkia orientalis* is another yeast with industrial potential of high-level production of organic acids. We have reconstructed genome-scale metabolic models for the two yeast organisms following a previously established rapid reconstruction workflow [1]. Using *S. cerevisiae* as the reference organism, we identified the homologous genes between *S. cerevisiae* and the two organisms using bidirectional BLAST as well as the Yeast Genome Annotation Pipeline [2] which identifies homology based additionally on synteny conservation. The most recent yeast7 model for *S. cerevisiae* [3] has been updated with the corrected gene-protein-reaction (GPR) associations [4], completely balanced stoichiometries and standardized biomass reactions [5]. Backbone models for the two organisms were extracted from the updated yeast7 model by comparing the homologous genes and evaluating the GPR associations. Additional reactions not in the backbone models were identified from multiple sources of data, including annotations by InterPro [6], KEGG [7], Pfam [8, 9] and UniProt [10]. Gapfilling was performed using reactions from the yeast7 model to ensure that the models are able to produce biomass. Flux elucidation using 13C-MFA requires the availability of a genome-scale carbon mapping model. To this end, the mapping models corresponding to the constructed genome-scale models of *R. toruloides* and *I. orientalis* are constructed using the previously published mapping models for *E. coli* and *Synechocystis* PCC 6803 as the starting point. Elementary metabolite unit (EMU) decomposition is performed using 46 metabolite fragments from 15 central metabolites and 12 amino acids quantified using GCMS and LCMS techniques. The identified carbon transitions are compared and contrasted with *E. coli*. These findings will serve as the foundation for further genome-scale computational analysis, including 13C-MFA, kinetic modeling, identification of metabolic regulation, and integration of experimental data to facilitate the engineering of the two yeast organisms.

References

1. Mueller, T.J., Berla B.M., Pakrasi H.B., and Maranas C.D. (2013) Rapid construction of metabolic models for a family of Cyanobacteria using a multiple source annotation workflow. *BMC Syst Biol*, 7:142

2. Proux-Wéra, E., Armisén, D., Byrne, K. P., and Wolfe, K. H. (2012). A pipeline for automated annotation of yeast genome sequences by a conserved-synteny approach. *BMC Bioinformatics*, *13*: 237
3. Aung,H.W., Henry, S.A., and Walker, L.P. (2013) Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Ind Biotechnol*, 9:215–228.
4. Chowdhury R., Chowdhury A., and Maranas C.D. (2015) Using gene essentiality and synthetic lethality information to correct yeast and CHO cell genome-scale models. *Metabolites*, 5:536-570
5. Chan S.H.J., Cai J., Wang L., Simons-Senftle M.N., and Maranas C.D. (2017) Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics*, 33:3603-3609.
6. Finn, R.D., et al. (2017) InterPro in 2017 — beyond protein family and domain annotations**.** *Nucleic Acids Res*, 45, D190-D199.
7. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*, 44:D457-D462
8. Finn, R.D., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, 44:D279-D285.
9. Alborzi, S. Z., Devignes, M. D., and Ritchie, D. W. (2017) ECDomainMiner: discovering hidden associations between enzyme commission numbers and Pfam domains. *BMC bioinformatics*, 18*:*107.
10. The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res*, 45:D158-D169

Supported by funding from the U.S. Department of Energy by grant DE-SC0018260.

Genome-scale Engineering of *Saccharomyces cerevisiae* with Single Nucleotide Precision

Zehua Bao,[1] Pu Xue,[2]* (puxue2@illinois.edu), Mohammad HamediRad,[2] Han Xiao,[1,3] Ipek Tasan,[4] Ran Chao,[4] Jing Liang,[5] and **Huimin Zhao**[1,2,4,5,6,★]

[1]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; [2]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801; [3]Current address: State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, and Laboratory of Molecular Biochemical Engineering, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China; [4]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801; [5]Metabolic Engineering Research Laboratory, Science and Engineering Institutes, Agency for Science, Technology and Research, Singapore 138669, Singapore; and [6]Departments of Chemistry and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801

★Correspondence should be addressed to H.Z. (zhao5@illinois.edu). Phone: (217) 333-2631. Fax: (217) 333-5052.

http://scs.illinois.edu/~zhaogrp/

Project Goals:

To develop an efficient, iterative, genome-scale single nucleotide resolution editing tool to improve complex yeast phenotypes.

Abstract:

Through systematically exploring the fitness landscape in a genome-wide manner, genome-scale engineering can test multiple hypotheses in parallel and overcome our limited knowledge of biocomplexity. However, existing methods to create a genome-wide deletion library in yeast require years and many researchers. Here we report a CRISPR/Cas9 and homology-directed repair assisted genome-scale engineering (CHAnGE) method that enables generation of a genome-wide library of yeast gene disruption mutants each with a defined 8 nucleotides removal by a single researcher within a month. Such unprecedented precise and small deletions minimize the risk of disrupting overlapping genomic features and chromatin structures. We validated CHAnGE by identification of single disruption mutants conferring 5~7 fold improved tolerance to different growth inhibitors including acetic acid and furfural in a genome-wide manner (targeting 6459 genes). We further iterated CHAnGE for directed evolution of yeast genomes, achieving an additional 3~8 fold improvement. Finally, we demonstrated the genome-editing capability of CHAnGE at single-nucleotide resolution and provided preliminary evidence that CHAnGE will work in mammalian cells.[1]

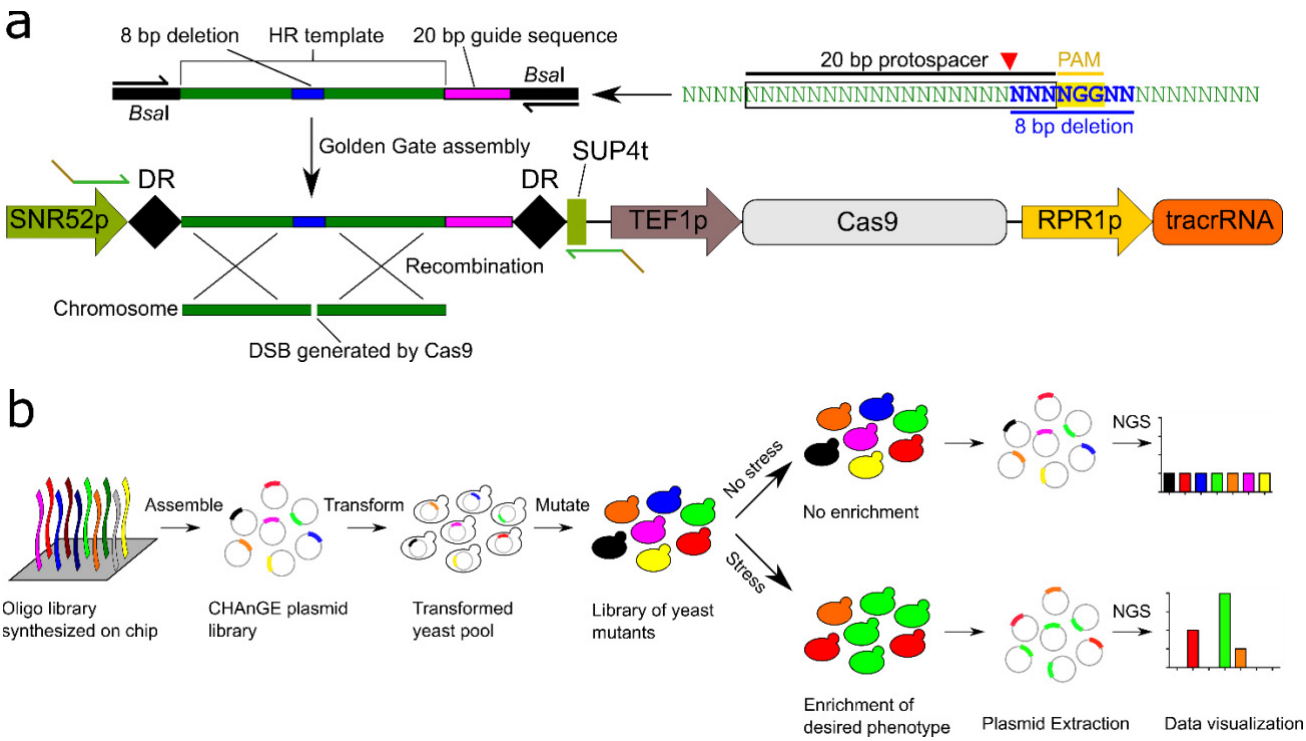


Figure 1. The CHAnGE workflow and validation. (a) Design of the CHAnGE oligonucleotide cassette. A 8 bp deletion near the Cas9 cleavage site (denoted by the red triangle) removes the PAM sequence and part of the protospacer sequence and introduces a frame shift mutation. A 100 bp HR template with 50 bp homology arms and the 8 bp deletion at the center was placed 5' of the guide sequence, constituting the CHAnGE cassette. The CHAnGE cassette was cloned into a Cas9 and tracrRNA expressing plasmid using Golden Gate assembly. After cleavage by Cas9, the target ORF was mutated by recombination with the HR template. (b) The CHAnGE workflow.[1]

References

1. Z. Bao, et al. "Genome-scale Engineering of *Saccharomyces cerevisiae* with Single Nucleotide Precision." *Nature Biotechnology* in press.

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-SC0018260.

Auxotrophy Prediction and Community Metabolic Modeling Reveals Functional Delegation in a Variety of Microbiome Systems

Pamela Weisenhorn*[2] (pweisenhorn@anl.gov), **Christopher Henry[2]**, Janaka Edirisinghe[2], **Adam P. Arkin[1]**, **Bob Cottingham[3]** and the KBase Team at the following institutions

[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 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.

KBase includes a pipeline that supports the development and analysis of community metabolic models starting from metagenomic data. These community models are valuable tools for improving our understanding of the molecular interactions and ecological principles that guide the behavior of microbial communities for numerous applications in the environment and industry. This new KBase pipeline integrates 11 analysis steps, including: metagenome assembly; binning of assembled contigs by species; assessment of assembly and genome quality and completeness; genome annotation; species metabolic model reconstruction; merging of models into a community model; mapping RNA-seq reads to individual species; model gapfilling; flux balance analysis; comparison of flux and expression profiles; and simulation of growth phenotypes. All eleven steps are available within a user-friendly point-and-click interface at http://narrative.kbase.us.

We applied this pipeline to understand the ecology and trophic interactions occurring within three microbiome-based datasets: (i) a lab-constructed community comprised of the cyanobacterium *Thermosynechococcus elongatus* supporting the heterotrophic bacterium *Meiothermus ruber* (Figure 1a); (ii) 120 naturally occurring soil isolates, many of which are comprised of 2-3 tightly coupled species (Figure 1b); (iii) a 13-species microbiome that performs electrosynthesis (Figure 1c); and (iv) a larger 18-species natural community comprising an epsomitic phototrophic microbial mat in Hot Lake, Washington. We demonstrate how the tools in KBase can integrate community Biolog profiles and transcriptomic profiles (Figure 1d) in our community models of these systems. Further, we show how our platform can predict the trophic interactions that occur among species in each of these systems. Overall, we find significant

trophic interactions in all communities, involving numerous essential metabolites, including amino-acids, vitamins, and cofactors. We also find numerous cases of metabolic handoff, particularly in our lab-constructed, electrosynthetic, and Hot Lake communities, which all include an autotrophic organism that performs carbon fixation functions for all other species in the community. We see how individual species form complex ecological networks of interactions where each species derives chemical energy from the nutrients in the environment, while stability is preserved through essential trophic interactions. We also see how changes in the nutrient content of the environment can disrupt these webs of trophic interactions.

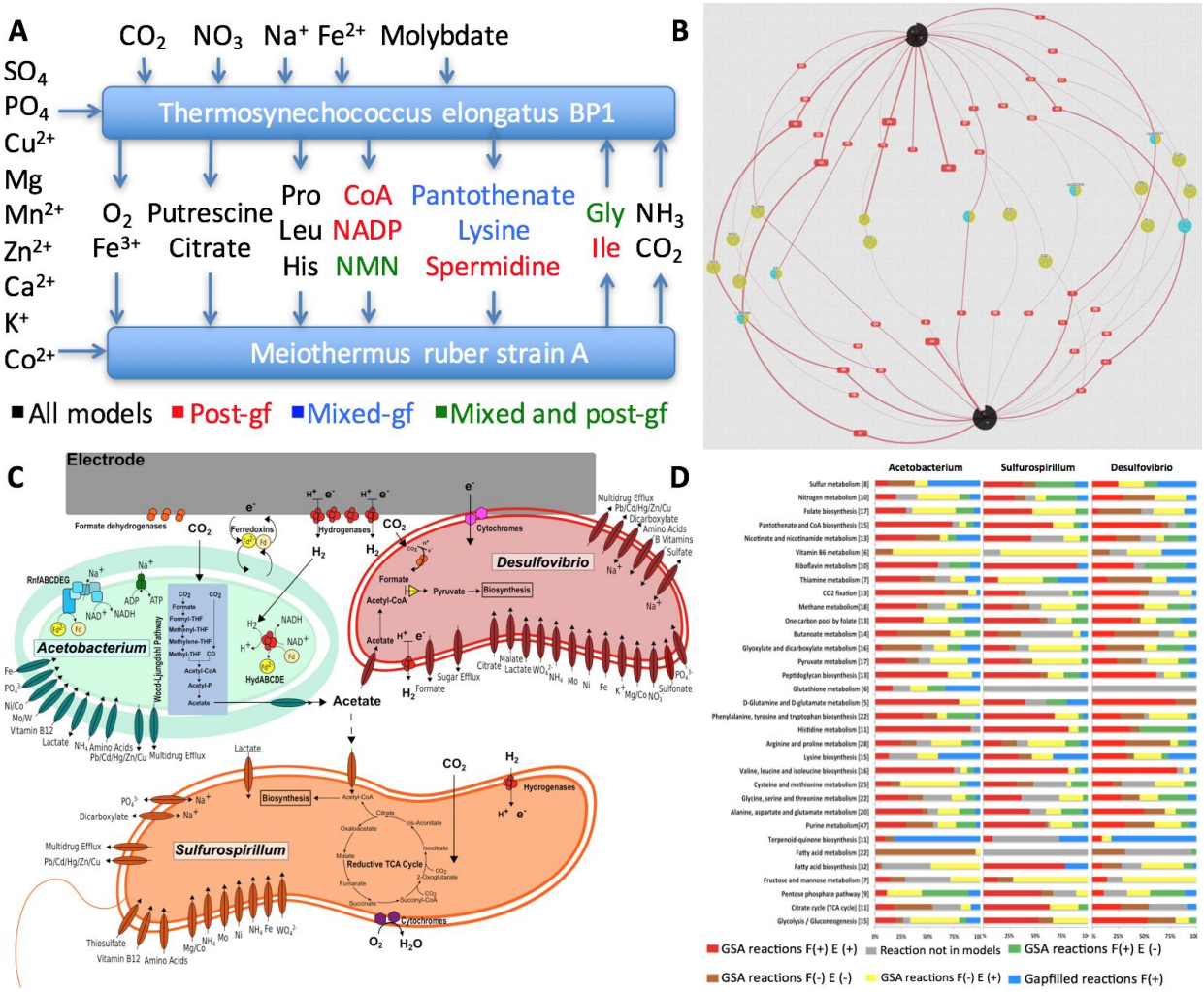


Figure 1. KBase was applied to predict interspecies interactions (A-C) with validation from transcriptomic data (D). The tools have been applied to a variety of systems, including a 2-species phototrophic consortia (A), numerous multi-species soil isolates (B), and a 13-species electrosynthetic microbiome (C).

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.

Systems Analysis and Engineering of Biofuel Production in *Chromochloris zofingiensis*, an Emerging Model Green Alga

Melissa S. Roth,[1,2] Sean D. Gallaher,[3] Daniel Westcott,[1,2] Tim Jeffers,[1,2] Daniela Strenkert,[3] Katherine B. Louie,[4] Benjamin P. Bowen,[4] Crysten E. Blaby-Haas,[5] Nanette R. Boyle,[6] Mary S. Lipton,[7] Trent R. Northen,[4] Sabeeha S. Merchant,[3] and **Krishna K. Niyogi**[1,2]* (niyogi@berkeley.edu)

[1]Howard Hughes Medical Institute, Berkeley, CA; [2]University of California, Berkeley, CA; [3]University of California, Los Angeles, CA; and [4]Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Brookhaven National Laboratory, Upton, NY; [6]Colorado School of Mines, Golden, CO; [7]Pacific Northwest National Laboratory, Richland, WA

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.* We will perform a comprehensive, large-scale systems analysis (transcriptome, proteome, metabolome, physiology) of *C. zofingiensis* during trophic transitions to understand and model how the energy metabolism of the cell is redirected based on the carbon source. We will integrate the systems data in a predictive model that will guide us in redesigning and engineering the metabolism of *C. zofingiensis* using cutting-edge synthetic biology and genome-editing tools to increase production of biofuel precursors (triacylglycerols) under photoautotrophic growth conditions.

Abstract:

As a core component of a sustainable bio-economy, microalgae have the potential to become a major source of biofuels and bioproducts without exacerbating environmental problems. These photosynthetic microbes utilize solar energy, grow quickly, consume CO_2, and can 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. Integrative systems biology and engineering of emerging model systems are needed to expand the possibilities of microbial production of biofuels and bioproducts. The unicellular green alga, *Chromochloris zofingiensis*, is one of the highest producers of the preferred lipid precursor for biofuel products, triacylglycerol (TAG), making it a promising biofuel feedstock. Moreover, *C. zofingiensis* exhibits an ability to dramatically switch its metabolism and divert much of its energy into making TAGs when a carbon source is added during growth. Unlike most algae, this accumulation of TAGs is in concert with accumulating biomass. However, during this metabolic switch, photosynthesis is also shut off. Recently, we have identified the molecular player hexokinase1 as a critical regulator of the photosynthetic and metabolic switch (Roth et al., in preparation). The overarching aim of this project is to design and engineer high-level production of TAGs in *C. zofingiensis.* To understand how energy metabolism is redirected

within the cell, we will perform a comprehensive, large-scale systems analysis integrated with a predictive model and using cutting-edge synthetic biology and genome-editing tools to increase production of biofuel precursors. To develop *C. zofingiensis* as an emerging oleaginous model system, we have sequenced and assembled a chromosome-level genome and high-quality transcriptome (1). The outcomes of the proposed research include a detailed, predictive, systems-level understanding of the regulation of metabolism for enhancing production of biofuels, development of tools and capabilities to enable manipulation of metabolism, and generation of engineered strains of *C. zofingiensis* to produce high levels of biofuel products. Advancing the basic science underlying regulatory 'switches' between growth and lipid production will likely be critical to developing economically viable algal biofuels.

Reference

1. Roth MS, Cokus SJ, Gallaher SD, Walter A, Lopez D, Erickson E, Endelman B, Westcott D, Larabell C, Merchant SS, Pellegrini M, Niyogi KK (2017). Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production. Proc Natl Acad Sci USA 114: E4296-E4305.

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.

Title:

Systems biology towards a continuous platform for biofuels production:

Engineering an environmentally-isolated *Bacillus* strain for biofuel production and recovery under supercritical CO_2.

Jason T. Boock[2], Adam J. E. Freedman[1,2], Geoffrey Tompsett[3], **Kristala L. J. Prather[2], Michael T. Timko[3]** and **Janelle R. Thompson*[1] (jthompson@mit.edu)**

[1]Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge MA; [2]Chemical Engineering, Massachusetts Institute of Technology, Cambridge MA; [3]Chemical Engineering, Worcester Polytechnic Institute, Worcester, MA

Project Goals: We are working towards the following goals (1) Develop a supercritical CO_2 tolerant strain of *B. megaterium* into a bioproduction host for biofuels (2) Engineer *de novo* pathways for biosynthesis of longer chain fuels in *B. megaterium* and (3) Develop and model a two-phase stripping chemostat for continuous biosynthesis and *in situ* extraction of biofuels using $scCO_2$ as a sustainable extractive solvent.

Abstract:

Supercritical carbon dioxide ($scCO_2$) is an attractive substitute for conventional organic solvents due to its unique transport and thermodynamic properties, its renewability and labile nature, and its high solubility for compounds such as alcohols, ketones and aldehydes. However, biological systems that use $scCO_2$ are currently limited to *in vitro* processes due to its strong inhibition of cell growth. Using the broad microbial lethality and solvent chemistry of $scCO_2$ to our advantage, we hypothesize that a dual-phase reactor of growth media and $scCO_2$ will simultaneously provide a sterile growth environment and the capacity to continuously strip off strain-produced biofuels, thus alleviating long standing bioprocess challenges of culture contamination and end-product toxicity. Towards this goal, environmental strain isolation provides an opportunity to discover organisms capable of growth in harsh environments, such as under $scCO_2$. Using a targeted bioprospecting approach by sampling fluid from a natural, deep-subsurface $scCO_2$ well, several species of gram-positive, endospore forming Bacilli were isolated that demonstrate consistent, robust growth in the presence of $scCO_2$. The species that showed the highest frequency and magnitude growth was identified as a strain of *Bacillus megaterium*, from here on referred to as SR7. The genome and plasmids of SR7 have been sequenced and annotated to determine the metabolic potential of this organism and compare it to related strains of *B. megaterium.* We have established optimal growth conditions and media by studying the single carbon preference of SR7. Spore germination was found to be crucial for achieving successful growth of SR7 under pressure and the addition of germination enhancers, such as L-alanine, significantly improved the growth frequency and magnitude of SR7 under $scCO_2$. High-level growth under $scCO_2$ enabled the measurement of the fermentation products lactate and acetate, representing the first biological products observed under $scCO_2$ and confirmation of active metabolism of SR7 cultured under high pressure.

Next we sought to develop SR7 as a biotechnologically relevant organism by implementing bioproduction pathways to generate products that would preferentially partition into $scCO_2$. Transformation of SR7 is possible using a protoplast-based method, which has permitted the identification of promoters (including two that have not been previously used in *B. megaterium*) capable of inducible heterologous protein expression in both aerobic and anaerobic conditions. Furthermore, the xylose-inducible promoter was evaluated under $scCO_2$ and found to have similar expression compared to anaerobic cultures. We engineered SR7 to produce isobutanol by introducing a two-enzyme (2-ketoisovalerate decarboxylase (KivD) and alcohol dehydrogenase (Adh)) pathway. A library of Adh proteins was screened to identify enzymes that rapidly convert the isobutyraldehyde intermediate in the pathway since this compound is expected to highly partition into the $scCO_2$ phase. Combining our recombinant biofuel strain with $scCO_2$ culturing, isobutanol production was observed, representing the first recombinant bioproduct generated from bacteria grown under $scCO_2$. For cultures that showed high metabolic activity under $scCO_2$, we found almost 50% conversion of the α-ketoacid substrate to biofuel product.

Extraction of alcohols into $scCO_2$ was measured using a custom-built, two-phase reactor/fermenter. We discovered that the efficiency of extraction as well as extraction rate is greater for longer chain length alcohols (*n*-hexanol > *n*-pentanol > *n*-butanol), which is the opposite of what is found for gas stripping of these molecules. Isobutanol extraction was compared to that of *n*-butanol, and was observed to be moderately faster and more efficient. We found that there is a strong dependence on efficiency of extraction with product concentration, necessitating the improvement of biological production for an economically viable $scCO_2$-based extraction strategy. Using the empirically collected data on extraction of alcohols as well as production titers observed in similar microbes, we have developed a process model for $scCO_2$ culturing and have found conditions that are comparable if not better than existing *in situ* extraction techniques such as gas stripping. Furthermore, we observed after collecting the alcohol-$scCO_2$-water mixture, upon intermediate depressurization, an alcohol-rich phase occurs, resulting in high purity product.

Currently we are working on building a genome scale model of our strain of *Bacillus*, inputting data from a transcriptomics study of SR7 grown in various environments. We are also using our transcriptomic data set to design a set of promoters to be used for future engineering of SR7, to understand the natural metabolic capacity of SR7 especially towards bioproducts of interest such as isopentanol, and to aid in on going work to elucidate the $scCO_2$ tolerance mechanism. Additionally, we are designing and optimizing metabolic pathways to make fuels directly from common carbon substrates using the known metabolism of SR7, and developing genomic integration/knockout protocols to enhance the metabolic engineering of this unique host. We have begun implementing a portion of the biofuel pathway to produce 4-methyl-pentanol, to evaluate the functionality of the carboxcylic acid reductase in SR7. Lastly, we are working to characterize this unique phase behavior of the alcohol-$scCO_2$-water mixture to better understand the process variables that maybe changed to optimize the $scCO_2$-based extraction strategy.

This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0012555).

Using Multi-Omic Data to Understand the Response of *Clostridium thermocellum* to Deletions of Genes for Lactate and Acetate Production

Evert Holwerda[1,3], Daniel G. Olson[1,3*] (daniel.g.olson@dartmouth.edu), David Stevenson [2,3], Daniel Amador-Noguez[2,3], Lee R. Lynd[1,2] and **Gerald A. Tuskan**[3]

[1]Dartmouth College, Hanover; [2]The University of Wisconsin, Madison; [3]The Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, Tennessee

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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Clostridium thermocellum is a promising candidate for the conversion of lignocellulosic feedstocks to biofuels due to its native ability to solubilize cellulose. Initial metabolic engineering efforts focused on improving ethanol production by deleting competing pathways for carbon flux (i.e., lactate and acetate production). Strains of *C. thermocellum* with deletions of both the lactate production pathway (*ldh* gene) and acetate production pathway (*pta* and *ack* genes) were constructed (A1 lineage). Subsequently, another set of strains was constructed with similar deletions (A2 lineage). Initially, ethanol production remained unchanged in both lineages. After a short period of adaptation, ethanol production improved in some strains, but not others. In these strains, carbon flux was redirected to pyruvate and amino acid production. As a means of understand the different fate of these parallel engineering strategies, we grew both sets of strains in chemostats. We identified three clusters of fermentation phenotypes. The first cluster is the wild-type phenotype, where cells produce a 1:1 ratio of ethanol and acetate. Strains with deletions of *hpt*, *spo0A* and *ldh* are in the first cluster. The second cluster phenotype involves production of ethanol, pyruvate and valine (i.e., flux was diverted from acetate to pyruvate and valine). Introducing the *pta* deletion to a strain in cluster 1 was sufficient to move it to cluster 2. The third cluster includes strains that produce mainly ethanol (about 50% of the maximum theoretical yield). Adaptation by serial transfer was sufficient to move strains from cluster 2 to cluster 3, at least in some cases. The resulting strains were resequenced and gene expression was analyzed by RNAseq. A mutation that causes a D494G mutation in the bifunctional alcohol and aldehyde dehydrogenase protein (AdhE) was observed to correlate with increased ethanol production (the change from cluster 2 to cluster 3 fermentation phenotype). To test the effect of this mutation, it was re-introduced into a strain without *ldh* or *pta* mutations and found to

reproduce about 90% of the increase in ethanol production. Ethanol production in *C. thermocellum* is limited by NADH availability rather than acetyl-CoA availability. Introducing a point mutation into the AdhE protein that changed its cofactor specificity allowed NADPH to be used for ethanol production, which increased flux through this pathway. There is no evidence that ethanol production in *C. thermocellum* is regulated at the transcriptional level. Information learned here will facilitate our understanding of our C4 alcohol and C6 ester fuels targets.

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

Development of an accelerated procedure for parameterizing kinetic metabolic models for *C. thermocellum*

Satyakam Dash[1]* (sud25@psu.edu), Saratram Gopalakrishnan[1]* (sxg375@psu.edu), Charles Foster[1] (cjf33@psu.edu), and **Costas D. Maranas**[1]

[1]Department of Chemical Engineering, The Pennsylvania State University of State, University Park, PA 16802

http://www.maranasgroup.com/

Project Goals: The goal of the project is to systematically construct dynamic models of *Clostridium thermocellum* by making use of Ensemble Modeling (EM) paradigm through integration of multiple omic information (metabolomic & fluxomic). These models will be instrumental in exploring genetic interventions for overproduction of biofuel products.

In this study, we describe the development of a metabolic model for a cellulolytic microbe, *C. thermocellum* which shows improved predictive capabilities upon incorporation of kinetic information. We further describe the limitations of existing parametrization procedures for kinetic model development and methods to overcome them using *E. coli* as a model system. For *C. thermocellum*, we constructed a core kinetic model k-ctherm118 to capture the regulatory impact of changes in metabolite pools on reaction fluxes using the Ensemble Modeling paradigm. k-ctherm118 was parameterized by using fermentation yield data in lactate, malate, acetate and hydrogen production pathways for 19 measured metabolites, 19 distinct single and multiple gene knockout mutants along with 18 intracellular metabolite concentration data for a *Δgldh* mutant and ten experimentally measured Michaelis-Menten kinetic parameters. k-ctherm118 captures metabolic perturbations caused by (i) nitrogen limitation leading to increased yields for lactate, pyruvate and amino acids, and (ii) ethanol stress triggering an increase in intracellular ammonia and sugar phosphate concentrations due to upregulation of cofactor pools. A secondary activity of ketol-acid reductoisomerase and possible regulation by valine and/or leucine pool levels were revealed by robustness analysis of k-ctherm118. k-ctherm118 also captures the growth inhibitory effect of pentose sugars by upregulating the non-oxidative pp-pathway genes. Overall, the *C. thermocellum* case study demonstrates that the developed kinetic model (k-ctherm118) provides greater insight into metabolic pathways and regulations than the stoichiometric model.

Kinetic models simulate genetic perturbations by modifying specific enzyme levels *a priori* based on the mutant genotype, which in turn modulates the concentration of all intracellular metabolites. We expand the scope of the developed kinetic models by incorporating a transcriptional regulatory layer which refines enzyme levels based on a linear combination of log-normalized changes in growth rate (global) and select intracellular metabolite pool (specific) levels. A major computational bottleneck in the kinetic parameterization process is the lack of a fast and efficient algorithm to identify the optimal set of kinetic parameters that minimizes deviations between predicted and MFA-derived steady-state flux distributions in response to genetic perturbations. The implementation of a gradient-based procedure is limited by failures in steady-state flux evaluations due to slow numerical integration. To this end, we have implemented an algorithm to compute steady-state flux distributions under multiple genetic perturbations for a given set of kinetic parameters that overcome the limitations of numerical integration. In conjunction with an efficient gradient-based scheme for updating kinetic parameters, we have developed a fast and automated algorithm for parametrization of a kinetic model that includes allosteric and transcript

regulations. We demonstrate the strength of this procedure using a recent study which provides a comprehensive metabolic flux characterization of wild-type *E. coli* and 22 knockouts of enzymes in the upper part of central carbon metabolism, including glycolysis, pentose phosphate pathway and ED pathway. We apply these flux datasets and the parametrization procedure to construct a core kinetic model of *E. coli* containing 138 model reactions, 93 metabolites, and 60 substrate-level regulatory interactions.

The work was supported by the genomic science grant from Department of Energy, USA (grant # DE-SC0012377).

A Metabolomic Based Approach to Identifying Bottlenecks in Biosynthetic Pathways

[1,2]**Edward E.K. Baidoo*** (eebaidoo@lbl.gov), [1,2]George Wang, [1,2]Chijioke Joshua, [1,2]Veronica Teixeira Benites, and [1,2,3,4]Jay D. Keasling.

[1]Joint BioEnergy Institute, Emeryville, CA, USA, [2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [3]Department of Bioengineering, University of California, Berkeley, CA, USA, [4]Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA, USA.

Contract url: https://science.energy.gov/bso/contract-management/

DOE Bioenergy Research Centers: http://genomicscience.energy.gov/centers/

Joint BioEnergy Institute url: https://www.jbei.org/

Project Goal: To demonstrate the utility and effectiveness of metabolomics, when combined with microbial physiology, in the identification of bottlenecks in biosynthetic pathways.

In the post-genomics era there has been a growing emphasis on understanding the functions associated with gene products. However, due to many genes not being under transcriptional control and the incomplete prediction of the proteome from the transcriptome owing to post-translational modifications, there has been a push in the fields of microbial metabolic engineering and synthetic biology to obtain phenotypic information from the metabolome, as it reflects more closely metabolic activities within a cell. While complete global metabolite profiles (comprehensive metabolomics) are not yet achievable due to the chemical diversity that exists within the metabolome, a focused strategy involving the measurement of localized metabolism (targeted metabolomics) has emerged as the favored approach. Targeted metabolomics illuminates those

aspects of metabolomics data that are truly meaningful to a research study. Thus, ensuring the identification of bottlenecks in biosynthetic pathways through the observation of accumulated pathway intermediates and cofactors. Furthermore, when combined with the knowledge and understanding of microbial physiology, we can accurately determine metabolism that is closely related to engineered biosynthetic pathways. Moreover, we can correlate the impact of the engineered pathway on central carbon metabolism and, hence, a microorganism's physiological state. Herein, we demonstrate the effectiveness of this strategy through the identification of bottlenecks in the heterologously expressed mevalonate pathway in *Escherichia coli*. This approach could be used to determine mechanisms employed by microorganisms to overcome genetic modification and can be directed towards engineering robust microbial cell factories for producing valuable renewable bioproducts.

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.

novoStoic: Pathway design using *de novo* steps through uncharted biochemical spaces

Lin Wang[1*] (lqw5322@psu.edu), Akhil Kumar[2], Chiam Yu Ng[1], and **Costas D. Maranas**[1]

[1]Department of Chemical Engineering, The Pennsylvania State University of State, University Park, PA 16802; [2]The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802

http://www.maranasgroup.com/

Project Goals: The goal of this project is to develop an optimization based pathway design tool that seamlessly integrates known reactions with reaction rules. The software tool (i.e., novoStoic [1]) allows for the incorporation of performance criteria for rank ordering pathway designs, thermodynamic feasibility constraints, limits on the number of novel reaction steps, and restrictions on the organism or class of organisms the reactions are chosen from. In addition, enzymes are flagged that will have to be re-engineered for an altered substrate specificity to enable the novel reaction steps.

Computational pathway design algorithms enumerate potential routes linking the two molecules, while often taking into consideration a multitude of criteria such as shortest route, the minimal number of non-native reactions, thermodynamic feasibility, and enzyme availability. Although most methods rely on the large number of enzymatic reactions catalogued already, there are also increasing number of tools that employ biotransformation rules derived from the existing reactions to discover novel pathways. The latter relies on the remarkable malleability of enzymes to accept a broad range of substrates as well as the potential of protein engineering for enzyme redesign. However, existing retrosynthesis tools generally traverse production routes from a source to a sink metabolite using known enzymes or *de novo* steps by graph-based methods. Generally, important considerations such as blending known transformations with putative steps, complexity of pathway topology, mass conservation, cofactor balance, thermodynamic feasibility, microbial chassis selection, and cost are largely dealt with in a posteriori fashion.

To address the problems in current computational methods, novoStoic [1] (see Figure 1) designs bioconversion routes while simultaneously considering any combination of the aforementioned design criteria. First, we augmented the MetRxn [2] (www.maranasgroup.com/metrxn) repository with a new dataset of elementally balanced reaction operators using the automated CLCA [3] based reaction rule extraction procedure termed rePrime. For each reaction, the rePrime procedure identifies and captures as reaction rules the molecular graph topological changes underpinning the substrate to product graph conversion. A reaction rule is a vector that captures the location of active reaction centers affected by the conversion of substrates to products. The reaction rules and known reactions are then operated upon a mixed integer linear programming (MILP) procedure, novoStoic then identifies a mass-balanced biochemical network that converts a source metabolite to a target while satisfying a multitude of constraints and optimizing the objective function. We demonstrate the use of novoStoic in bypassing steps in existing pathways through putative transformations, assembling complex pathways blending both known and putative steps towards biofuel and biorenewable molecules and postulating ways to biodegrade xenobiotics. novoStoic is available at Maranas group website (https://github.com/maranasgroup). Efforts are currently under way to integrate the tools within KBase.

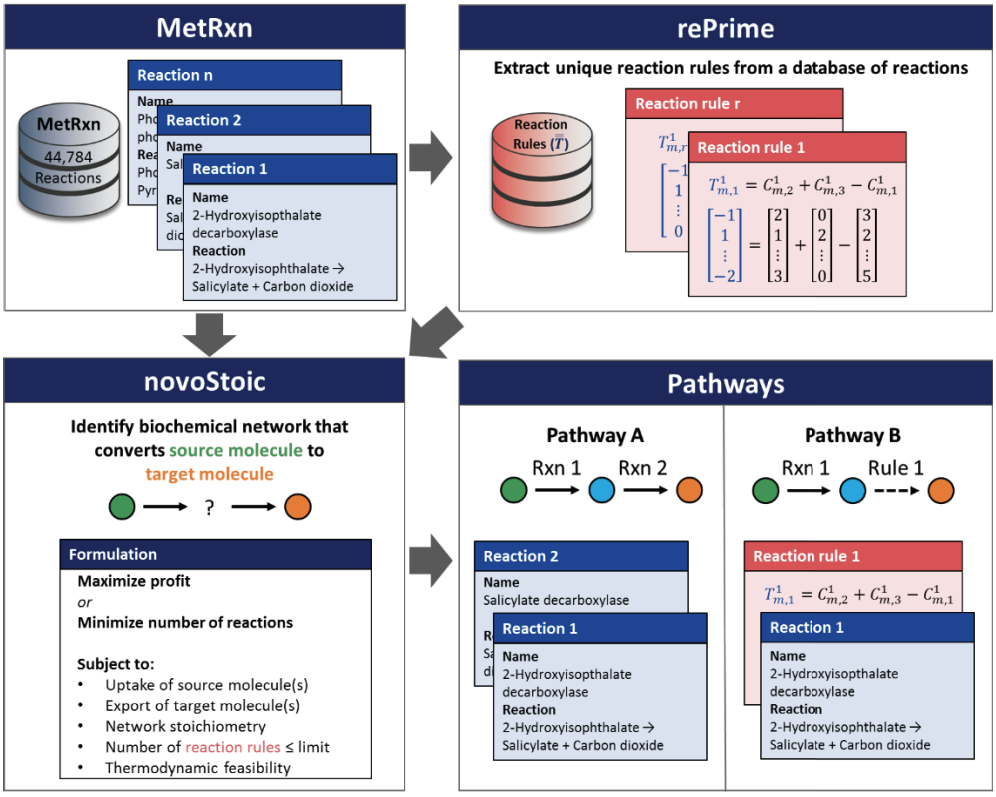


Figure 1: Schematic overview of the rePrime/novoStoic procedure. First, the rePrime procedure is used to pre-process the MetRxn database of reactions (blue boxes) to extract a unique set of reaction rules. The novoStoic procedure is then used to identify a series of intervening reactions and reaction rules that convert source molecule(s) (green circle) into target molecule(s) (orange circle) such that the profit can be maximized or the number of reactions in the pathway can be minimized.

References

[1] Kumar, A., Wang, L., Ng, C. Y., & Maranas, C. D. (2018). Pathway design using de novo steps through uncharted biochemical spaces. *Nature Communications*, *9*(1), 184.

[2] Kumar, A., Suthers, P. F., & Maranas, C. D. (2012). MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC bioinformatics*, 13(1), 6.

[3] Kumar, A., & Maranas, C. D. (2014). CLCA: maximum common molecular substructure queries within the MetRxn database. *Journal of chemical information and modeling*, 54(12), 3417-3438.

Supported by funding from the U.S. Department of Energy by grant DE-SC0008091 (completed).

Encoding new non-standard amino acid designs into proteins can address limitations of fluorescent proteins and facilitate genome editing approaches

Erkin Kuru[1]* (erkin_kuru@hms.harvard.edu) , Jonathan T. Rittichier,[1] Devon Stork,[1] Aditya Kunjapur,[1] and **George M. Church**[1]
[1] Harvard Medical School, Boston, MA

http://arep.med.harvard.edu

Project Goals: To engineer and demonstrate new cellular translation machinery in our reduced-codon *E. coli* strains to enable efficient expression of proteins containing multiple Non-Standard Amino Acids (NSAAs).

Improvements in imaging/labeling technology have often led to explosive growth of new knowledge in the basic sciences – for example, the development of fluorescent protein (FP) fusions as genetically encoded 'beacons'. However, as our ability to track the subcellular localization of single-molecules with nanometer/microsecond spatiotemporal resolutions soars, FP and FP-inspired protein labeling technologies are still lagging due to their large (>4nm) sizes and minutes-to-hours long maturation times.

Using designer fluorogenic amino acids (FgAAs) and nucleobase amino acids (NuAAs), we aim to develop new protein labeling strategies that can replace FPs in many cell biology applications. FgAAs, with their unique ability to become fluorescent only when incorporated into proteins, (i) are ~50-100 times smaller than most FPs; (ii) 'mature' instantaneously; (iii) are fluorescent in anaerobic conditions; and (iv) do not oligomerize. NuAAs will facilitate programmability of protein-DNA interactions, e.g. allowing multi-color protein imaging *via* DNA-PAINT, giving <10-nm spatial resolutions. Using growth- and translation-optimized *E. coli* C321.ΔA-based strains developed in our lab that lack UAG codons and stop factors, we are developing the ability to encode FgAAs and NuAAs into proteins of interest. These technologies promise to improve our understanding of sub-cellular organization and, *via* NuAAs, to confer DNA-like properties on proteins that will facilitate genome engineering.

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-FG02-02ER63445.

Engineering a reduced 57-codon genetic code in *Escherichia coli*

Nili Ostrov,[1] Alexandra Rudolph*[2] (alexandra_rudolph@g.harvard.edu), Maximilien Baas-Thomas,[2] Matthieu Landon,[3] **George Church**[1,4]

[1]Department of Genetics, Harvard Medical School, Boston, MA; [2]Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA; [3]Program in Systems Biology, Harvard University, Cambridge, MA; [4]Wyss Institute for Biologically Inspired Engineering, Boston.

http://arep.med.harvard.edu

Project Goals: Develop an *E. coli* strain that has been engineered to use only 57 codons, and demonstrate its ability to incorporate multiple non-standard amino acids, to enable tight biocontainment, and to resist phages.

We report progress towards assembly of a 3.97 Mb, 57-codon *Escherichia coli* genome in which seven codons were replaced with synonymous alternatives in all protein coding genes. In-house design software selected optimal synonymous alternatives for targeted codons, and the recoded genome was synthesized in 87 segments. We de-risked >3,000 recoded genes and established a rapid troubleshooting procedure for the 27 design exceptions identified. Here, we present our pipeline for the construction of a single recoded strain. We have developed an iterative process using CRISPR/Cas9-assisted λ-Red recombineering for scarless replacement of genomic sequences with recoded segments. Using this system, we constructed two parental strains (seg16 and seg25) and began parallelized, successive segment replacement. Importantly, this process is modular (in segment replacement order and choice of markers) and is compatible with hierarchical genome recoding. This work underscores the feasibility of rewriting genomes and construction of synthetic organisms.

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-FG02-02ER63445 and Defense Advanced Research Projects Agency grant BAA-12-64.

Exploring the Species Specificity of Lambda Red Recombination

Timothy Wannier[1*] (timothy_wannier@hms.harvard.edu), Gabriel Filsinger[1], Verena Volf[1], Helene Kuchwara[1], George Church[1]

[1]Harvard Medical School, Department of Genetics, Boston, MA

http://arep.med.harvard.edu

Project Goals: To extend high-efficiency recombineering beyond *E. coli* and closely related organisms.

The single-stranded annealing protein (SSAP) "Beta" enables a wide variety of genomic manipulations in *E. coli*, including easy generation of single base pair changes, deletions, insertions, memory storage using retrons, and recoding of large 50kb+ segments. However, this protein is not active in all organisms, limiting its use to applications involving *E. coli* and closely related organisms such as *Salmonella*. Finding SSAPs that are active in new organisms is laborious and requires screening of candidate protein variants in the target host organism. This method may not identify highly active SSAPs, as only proteins from closely related species would normally be screened. Ideally, high-activity SSAPs could be engineered to function effectively in any host organism. Here, we use *E. coli* and *L. lactis* as model systems to examine the species specificity requirements of single stranded annealing proteins. *E. coli* is a gram negative bacterium from the phylum Proteobacteria, while *L. lactis* is a gram positive bacterium from the phylum Firmicutes. Active SSAP variants have been identified for both species, but with much higher activity in *E. coli*. We screen hundreds of protein variants and thousands of SSAP fusions to elucidate the features enabling species-specific programming of single-stranded annealing proteins.

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-FG02-02ER63445.

Toward Recombinant Cellulolytic Bacteria: Stabilizing Displayed Cellulases on Vegetative *Bacillus subtilis* Cells

Grace L. Huang,[1] Jason Gosschalk[1]* (jaygosschalk@ucla.edu), Ye Seong Kim,[1] Xiyan Sun,[2] and **Robert Clubb**[1]

[1]UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, 611 Charles Young Drive East, Los Angeles, CA 90095, USA. [2]Chinese Academy of Sciences, Kunming Institute of Botany, Yunnan Kunming, China.

Project Goals: We are studying how eubacteria display proteins and polymers on their surface, toward the goal of creating surface engineered microorganisms that can degrade lignocellulose. To gain broad insight, our surface engineering efforts are concentrated on *Bacillus subtilis*, a model Gram-positive microorganism that is used industrially. For practical applications, surface-displayed cellulases need to be bound to the peptidoglycan at a high density and complexity, while remaining highly stable. In this study, we have systematically explored several well-established protease-deficient *B. subtilis* strains' ability to display cellulases and assessed the impact of surface-display on bacterial growth, cellular stress and morphology. The results of this work can be used to guide the construction of cellulase-displaying microbes that convert plant biomass into second generation biofuels, chemicals, and materials.

Lignocellulosic biomass is a promising feedstock from which to sustainably produce useful biocommodities, but its recalcitrance to hydrolysis limits its commercial utility. One attractive strategy to overcome this problem is to use consolidated bioprocessing (CBP) microbes that directly convert biomass into chemicals and biofuels. Several industrially useful microbes possess desirable consolidated bioprocessing characteristics, yet they lack the ability to degrade biomass. Engineering these microbes' surfaces to display cellulases and cellulosome-like structures could endow them with potent cellulolytic activity, enabling them to be used in CBP. To gain general insight into factors controlling protein display, we engineered the surface of the model Gram-positive bacterium *B. subtilis*. We constructed a *B. subtilis* protein display reporter system in which the *Clostridium thermocellum* Cel8A endoglucanase is fused to the LysM cell wall binding module. The effects of LysM positioning, extracellular proteases, and solution conditions on the copy-number and stability of the reporter protein were determined. We demonstrated that heterologous surface enzyme activity is rapidly lost, even when *B. subtilis* is genetically modified to eliminate all of its extracellular proteases (AprE, Epr, Bpr, Vpr, NprE, NprB, Mpr, WprA, HtrA and HtrB). This problem presumably occurs because the membrane's proton motive force (pmf) dissipates when nutrients are scarce, leading to autolysis and the concomitant release of cytoplasmic proteases that degrade the heterologous surface proteins. We overcame cellular autolysis using a two-step procedure in which the pmf is maintained by glucose or glycerol additives. This procedure enables the production of cellulase-coated *B. subtilis* cells that are stable for more than two days, as substantiated by whole-cell enzyme activity measurements and cell fractionation experiments in tandem with immunoblotting. We observed that cell morphology is affected by protein display in certain strains and that the secretory stress response is activated as a result of both Cel8A-LysM display and deficiency of the HtrA and HtrB proteases; we have made strides toward determining the ideal *B. subtilis* strain and growth conditions for CBP. The ability to produce stable enzyme-coated vegetative *B. subtilis* is a step toward their practical use in biotechnological applications, and lays the foundation for their further optimization.

This work was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421. GL Huang was supported by a Ruth L. Kirschstein National Research Service Award GM007185.

Design and engineering of native regulatory networks

Emily Freed[1*] (emily.freed@colorado.edu), Alaksh Choudhury[1], Rongming Liu[1], Liya Liang[1], Adam Arkin[2], Christopher Voigt[3], Carrie Eckert[1,4], and **Ryan Gill**[1]

[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

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 organism *E. coli* and then in DOE relevant non-model organisms.

Advances in DNA reading and writing technologies are driving adoption of new paradigms for engineering biological systems. 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 biofuels and bioproducts. Regulation plays a special role in the evolution of such combinatorial structures, where mutations affecting the structure and function of regulatory networks are known to account for much of life's diversity. For example, major evolutionary innovations often arise through mutations in regulatory networks and organisms with larger regulatory networks are both more evolvable and more robust. A challenge, however, is that very little is known about the more detailed design rules underlying these observations. This lack of understanding generally undermines efforts to forward design and engineer native regulatory networks, and as such prevents us from taking advantage of such networks in the same way evolution does to go after complex phenotypes.

To address these challenges, we previously developed a technology, CRISPR Enabled Trackable Genome Engineering (CREATE), for constructing and tracking (by barcode) 100,000's of designer mutations in microbial genomes. We used CREATE to construct a regulator library in *E. coli* that contains 170,000 designer mutations spanning 131 *E. coli* regulators. These regulators were sub-divided into the following groups: "global" regulators (50+ regulated genes), regulators involved in core transcriptional process, and "high-level" regulators (20-49 regulated genes). We designed mutations to target known and predicted functionally important sites. The functions include active site, ligand binding sites, protein dimerization interfaces and DNA/RNA binding sites.

We further designed a pipeline that uses bioinformatics mining, protein structural analysis and homology modelling to find sites with the highest likelihood of functional importance. Using the pipeline, we were able to identify ~8800 amino acid residues across the 131 proteins, with an average coverage of ~30% of protein sequence space. We then designed CREATE cassettes to generate site-saturation mutagenesis libraries of these sites. We exposed these libraries to growth selections in several environments relevant to biofuels production, in all cases rapidly identifying mutants with improved fitness. We similarly were able to identify regulators in *S. cerevisiae* that, when mutated, confer a fitness advantage in environments relevant to biofuels production.

In future work, we will expand upon these early studies to now target all regulators in the *E. coli* genome (~300). We expect this will require a library of ~500,000 mutants. In addition to substitution libraries targeting the DNA binding pocket of regulatory proteins, we also have the ability to modify the promoter regions of any gene in the genome by: i) inserting synthetic promoters to turn "on" or "off" downstream genes, ii) inserting DNA binding consensus sequences to introduce novel "on" of "off" native regulation, or (iii) removing DNA binding sequences to remove native regulation. We propose to do so in combination with quantitative measurements of global expression (e.g. RNAseq) and DNA binding patterns (e.g. ChIPseq) to not only decipher how selected regulatory protein mutants affect overall regulatory network function but also to codify this understanding into design criteria for improved engineering of native regulation.

References

1.Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* **433**, 481–487 (2005).

2. Payne, J. L. & Wagner, A. The robustness and evolvability of transcription factor binding sites. *Science* **343**, 875–877 (2014).

3. Isalan, M. et al. Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**, 840– 845 (2008).

4. Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. & Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314**, 1565–1568 (2006).

5. Garst, A. D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat Biotechnol* **35**, 48–55 (2017).

6. Payne, J. L. & Wagner, A. Mechanisms of mutational robustness in transcriptional regulation. *Front Genet* **6,** 322 (2015).

7. Wagner, A. The molecular origins of evolutionary innovations. *Trends Genet* **27,** 397–410 (2011).

8. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A. & Gill, R. T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28**, 856– 862 (2010).

This research is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Building a Bridge between Cell-Free Experimentation and Cellular Metabolic Engineering

Ashty Karim,[1,2,3]* (ashtykarim@u.northwestern.edu), Quentin Dudley[1,2,3], Alex Juminaga[6], Samantha Crowe[1,2,3], Jacob Heggestad[1,2,3], William Grubbe[1,2,3], Yongbo Yuan[6], Rasmus Jensen[6], Sean Simpson[6], Michael Koepke[6], **Michael C. Jewett**[1,2,3,4,5]

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL; [2]Chemistry of Life Processes Institute, Northwestern University, Evanston, IL; [3]Center for Synthetic Biology, Northwestern University, Evanston, IL; [4]Robert H. Lurie Comprehensive Cancer Center, Chicago, IL; [5]Simpson Querrey Institute, Northwestern University, Chicago, IL; [6]LanzaTech, Skokie, IL, United States.

http://jewettlab.northwestern.edu; http://www.lanzatech.com

Project Goals: We will address the challenge of designing, building, and optimizing biosynthetic pathways in cells in a new 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.

Speeding up design-build-test (DBT) cycles is a fundamental challenge facing metabolic engineering. We plan to address this challenge by establishing the clostridia Foundry for Biosystems Design (cBioFAB) (Figure 1). To this end, we report a new *in vitro* prototyping and rapid optimization of biosynthetic enzymes approach (termed iPROBE) to inform cellular metabolic engineering. In our approach, cell-free cocktails for synthesizing target small molecules are assembled in a mix-and-match fashion from crude cell lysates selectively enriched with pathway enzymes. This approach reconstructs pathways in two steps where the first step is enzyme synthesis via cell-free protein synthesis and the second step is enzyme utilization via substrate and cofactor addition to activate small molecule synthesis. We demonstrate that iPROBE can quickly study pathway enzyme ratios, tune individual enzymes in the context of a multi-step pathway, screen enzyme variants for high-performance enzymes, and discover enzyme functionalities. The rapid ability to build pathways *in vitro* using iPROBE allows us to generate large amounts of data describing pathway operation under several operating conditions. However, to date no easy method of analysis provides informative bridging of cell-free data to cellular metabolic engineering. In this work, we address this limitation by developing a

quantitative metric that combines titer at reaction completion, rate during the most productive phase of pathway operation, and enzyme expression as measured by protein solubility (TREE score). By reducing the complexity of available cell-free data to one value we can now quickly screen and rank pathways in the cell-free environment and provide useful information for cellular metabolic engineering. We demonstrate iPROBE and the use of the TREE score for the production of 3-hydroxybutyrate and *n*-butanol in *Clostridium*, an industrially relevant non-model organism. This work shows that iPROBE can be used for multiple enzymatic pathways and for non-model organisms. We anticipate that iPROBE will facilitate efforts to define, manipulate, and understand metabolic pathways for accelerated DBT cycles in the cell-free environment before engineering organisms.

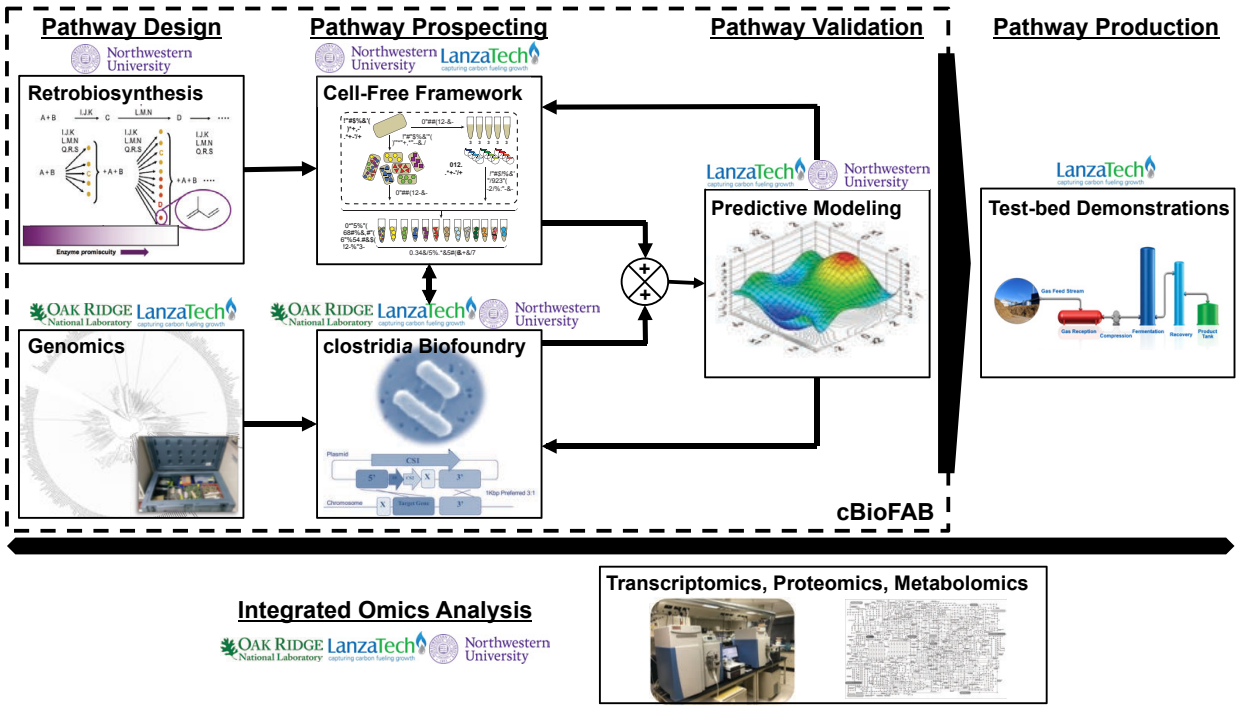


Figure 1. Proposed workflow for this project and participating units.

We acknowledge the Department of Energy grant DE-SC0018249 for funding of this project.

MinGenome: An *in silico* top-down approach for the synthesis of minimized genomes

Lin Wang[1*] (lqw5322@psu.edu), Costas D. Maranas[1], and **Himadri B. Pakrasi**[2]

[1]Department of Chemical Engineering, The Pennsylvania State University of State, University Park, PA 16802; [2]Department of Biology, Washington University, St. Louis, MO 63130

http://www.maranasgroup.com/

Project Goals: The goal of this project is to design a computation tool (i.e., MinGenome [1]) to support top-down genome minimization by shedding functions unnecessary for fast growth and other desirable traits. The hypothesis is that genome-minimized organisms will behave in a more predictable and controllable manner. MinGenome also aids in the identification of previously unknown essential functions whenever a new deleted stretch causes lethality.

Genome-scale metabolic models have led us to a greater understanding of cellular metabolism. However, many cellular and metabolic processes in a cell are still not well understood, placing limits on the extent to which the systems can be predictably engineered. Genome minimized strains are obtained by removing genome segments associated with genes or processes either detrimental or un-needed under bioproduction conditions. They offer advantages as production chassis [2] by reducing transcriptional cost, eliminating competing functions and limiting unwanted regulatory interactions. Existing approaches for identifying stretches of DNA to remove are largely *ad hoc* based on information on presumably dispensable regions through experimentally determined non-essential genes and comparative genomics. As more sophisticated genome editing tools (e.g., CRISPR) are becoming commonplace, the need for a computational aid that will help successively minimize genomes consistent with a set of performance criteria beyond simply growth rate is becoming more pressing.

Herein we present a versatile genome reduction algorithm MinGenome [1] (Figure 1) that implements a mixed integer linear programming (MILP) algorithm to iteratively identify the largest dispensable contiguous sequences without affecting the organism's growth or other desirable traits. Known essential genes or genes that cause significant fitness or performance loss are flagged and their deletion is thus prohibited. MinGenome also preserves needed transcription factors and promoter regions ensuring that retained genes will be properly transcribed while also avoiding the simultaneous deletion of synthetic lethal pairs. The potential benefit of removing even larger contiguous stretches of DNA if only one or two essential genes (to be re-inserted elsewhere) are within the deleted sequence is explored. We apply the algorithm to design minimized *E. coli* strain and *B. subtilis* strains and find that we are able to recapitulate the long deletions identified in previous experimental studies [3] and discover alternative combinations of deletions which have not yet been explored *in vivo*. MinGenome is a versatile computational tool to guide genome reduction. It can be accessed at Maranas group website (https://github.com/maranasgroup). Efforts are currently underway to integrate the tool within KBase.

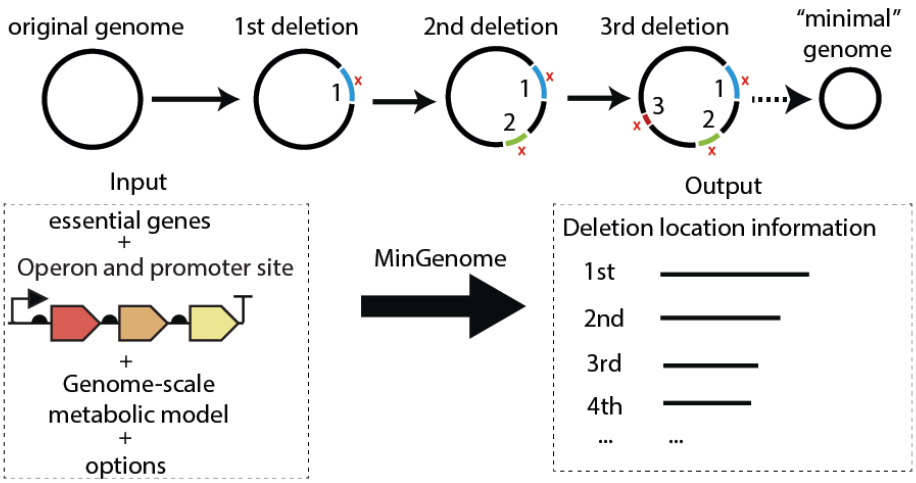


Figure 1. Schematic representation of the MinGenome algorithm. The MinGenome algorithm requires genome sequence information, gene annotation, the presence of a GSM model and information on essential genes, gene and promoter positions. It identifies the sequence of deletions starting with the largest dispensable region and proceeding monotonically to shorter ones.

References

[1] Wang, L., Maranas, C.D. (2017), MinGenome: An *in silico* top-down approach for the synthesis of minimized genomes. *ACS Synthetic Biology*.

[2] Vickers, C. E., Blank, L. M., & Krömer, J. O. (2010). Grand challenge commentary: Chassis cells for industrial biochemical production. *Nature chemical biology*, 6(12), 875-877.

[3] Juhas, M., Reuß, D. R., Zhu, B., & Commichau, F. M. (2014). *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiology*, 160(11), 2341-2351.

Supported by funding from the U.S. Department of Energy by grant DE-SC0012722.

Adaptive Laboratory Evolution Provides a Route to Improve Biofuel Production Under Ionic Liquid Stress

Thomas Eng (tteng@lbl.gov) [1], Philipp Demling[1], Robin A Herbert[1], Yan Chen[1], Veronica Benitez[1], Lars M Blank[3], Joel Martin[2], Anna Lipzen[2], Edward E K Baidoo[1], Christopher J Petzold[1], and **Aindrila Mukhopadhyay**[1]

[1] Joint BioEnergy Institute, Emeryville, CA; [2]Joint Genome Institute, Walnut Creek, CA;
[3] RWTH Aachen University, Aachen, Germany

Project Goals: To use adaptive laboratory evolution (ALE) to identify strains with the capacity produce high titers of biofuels in the presence of ionic liquids.

The production of biofuels in microbes needs to consider many upstream and downstream bottlenecks. Upstream bottlenecks include the toxicity inherent from crude biomass or pretreatment inhibitors, both which reduce the ability to consolidate fermentation processes. Downstream bottlenecks such as the toxicity of the target molecule also limit production titers. We hypothesized that lab adaptation could yield strains that are better able to reach high metabolic rates in biomass hydrolysates and potentially provide an improved background to express biosynthetic pathways and final products. Our adaptation studies led to the discovery of strains with the desired improved strain phenotype but tracked to mutations that functioned independently of any existing background mutants. Here, we report a hereditable mutation in an essential gene that bestows high IL tolerance. We confirmed that the mutant background not only provides improved IL tolerance but also results in increased production a final biojetfuel product, d-Limonene. The production of limonene production at 200 mg/L is the highest reported biogasoline production in the presence of typical levels of residual ILs in the culture medium. Our study suggests that the mutant strain mounts a distinct global physiological response which alleviates the cytotoxicity of [EMIM]OAc allowing for greater metabolic flux into the target gene pathway.

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, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

Assembling and Annotating Prokaryotic Genomes in KBase

Benjamin Allen*[3] (allenbh@ornl.gov), **Chris Henry**[2], **Adam P. Arkin**[1], **Bob Cottingham**[3] and the KBase Team at the following institutions

[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 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.

For microbiologists, assembling prokaryotic Next-Generation Sequencing reads and constructing an annotated genome with functional information about the coding DNA sequences are necessary first steps for comparative genomics, phylogenetic analysis, or metabolic modeling. In KBase, users can quickly and easily perform de novo assembly on prokaryotic DNA reads using multiple assemblers, and then run two separate annotation tools on the assemblies, calling genes and other genomic features and assigning biological functions, to generate an annotated genome. This genome can be used in downstream analysis within KBase or downloaded for use in other software platforms.

KBase has tools for performing sequencing quality assessment and control to ensure that high quality data is fed into the assemblers. For instance, a user can use the FastQC app on Illumina sequencing reads to determine if any trimming or adapter removal is necessary for raw sequence data, and then use the Trimmomatic app to trim the low quality reads on the ends of the reads and remove any adapters.

After performing quality assessment and control, users can assemble their reads with one of the many open source *de novo* assembly apps available in KBase. Two popular assembly apps in KBase are Velvet and SPAdes. Both Velvet and SPAdes are capable assemblers for prokaryotic reads, although SPAdes has been optimized for single cell data. Velvet is a de Bruijn graph based assembler that combines multiple algorithms for the construction, simplification, and error correction of assembly graphs. SPAdes is also a de Bruijn graph based assembler unique in its utilization of multisized de Bruijn graphs to generate a final assembly from multiple *k*-mers combined with error and mismatch correction tools. Running these assembly apps on a set of

reads generates an Assembly object, which contains the assembled contigs and can be used in downstream analyses or downloaded as a FASTA file. After performing multiple assemblies on a set of sequencing reads, it can be useful to compare the performance of each assembler by analyzing the quality of each assembly. QUAST is an open source assembly quality assessment tool that allow users to compare summary statistics of multiple assemblies. These statistics can be used to determine which assembly is optimal for feeding into downstream annotation pipelines.

Once the assembly has been created, users can annotate structural and functional features on the sequence to generate a genome. Annotating the structural and functional features of an organism's genome allows for comparative analysis against other organisms' genomes and provides a reference for understanding the flow of information within a biological system. Prokaryotic genome annotation in KBase is provided by two apps: Annotate Microbial Assembly and Annotate Assembly with Prokka. Annotate Microbial Assembly takes an assembly generated by one of the assembly apps or imported by the user and runs the sequence through a multi-step pipeline to predict gene locations within the assembly and perform functional annotation using the RAST (Rapid Annotations using Subsystems Technology) toolkit. This annotation pipeline assigns functions from the SEED Subsystems Ontology to genes using a fast *k*-mer based approach. Annotate Assembly with Prokka is a KBase app that calls the popular prokaryotic annotator Prokka. Prokka combines multiple open-source annotation tools in a quick and thorough annotation pipeline for prokaryotic sequences that calls gene annotations from UniProt and RefSeq, then queries the TIGRFAM and Pfam hidden Markov model databases for domain-specific annotations.

The final output of both annotation apps is a genome that can be analyzed for specific biological features called by annotation, such as proteins or regulatory elements. The contents can be explored in a tabular genome viewer that shows summary information about the Genome as well as a list of contigs and the genes that were annotated on each contig. The genome object can be downloaded as a GenBank file or used as input to KBase apps for comparative genomics, metabolic modeling, and more.

To learn more about KBase's tools for assembly and annotation, visit http://kbase.us/assembly-and-annotation/ or try the interactive Narrative tutorial, https://narrative.kbase.us/narrative/notebooks/ws.18188.obj.6.

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.

Pleiotropic and Epistatic Network-Based Discovery: Integrated Networks for Target Gene Discovery

Deborah Weighill[1,2], Piet Jones[1,2], Manesh Shah[2], Priya Ranjan[2], Wellington Muchero[2], Jeremy Schmutz[4,5], Avinash Sreedasyam[5], David Macaya-Sanz[6], Robert Sykes[3], Nan Zhao[7], Madhavi Z. Martin[2], Stephen DiFazio[6], Timothy J. Tschaplinski[2], Daniel Jacobson[1,2*] (jacobsonda@ornl.gov) and **Gerald A. Tuskan**[2]

[1]The Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, [2]The Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge, TN, [3]National Renewable Energy Laboratory, Golden, CO, [4]Department of Energy Joint Genome Institute, Walnut Creek, CA, [5]HudsonAlpha Institute for Biotechnology, Huntsville, AL, [6]Department of Biology, West Virginia University, Morgantown, [7]The University of Tennessee Institute of Agriculture, University of Tennessee, Knoxville

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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

Biological organisms are complex systems that are composed of functional networks of interacting molecules and macro-molecules. Complex phenotypes are the result of orchestrated, hierarchical, heterogeneous collections of expressed genomic variants. However, the phenotypic effects of genomic variants are the result of historic selective pressure, current environmental conditions and epigenetic signals, and, as such, their co-occurrence can be seen as genome-wide correlations in a number of different manners. Lignin is a complex polymer which is a significant component of plant cell walls and can be viewed as a polygenic phenotype of high importance to biofuels initiatives. This study makes use of data derived from the re-sequenced genomes from over 800 different *Populus trichocarpa* genotypes in combination with metabolomic and pyMBMS data across this population, as well as co-expression and co-methylation networks in order to better understand and identify target genes involved in lignin biosynthesis/deconstruction. A Lines-of-Evidence (LOE) scoring system is being developed to integrate the information in the different layers and quantify the number of LOE linking genes to lignin-related-phenotypes across the

network layers. The resulting Genome Wide Association Study networks, integrated with Single Nucleotide Polymorphism (SNP) correlation, co-methylation and co-expression networks through the LOE scores are proving to be a powerful approach to determine the pleiotropic and epistatic relationships underlying cellular functions and, as such, the molecular basis for complex phenotypes.

The Center for Bioenergy Innovation (CBI) 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

This research was funded by the BioEnergy Science Center (BESC) at the Oak Ridge National Laboratory (contract DE-PS02-06ER64304). BESC 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.

This research was also supported by the Department of Energy Laboratory Directed Research and Development funding (7758), at the Oak Ridge National Laboratory. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US DOE under contract DE-AC05-00OR22725.

This research used resources of the Oak Ridge Leadership Computing Facility (OLCF) and the Compute and Data Environment for Science (CADES) at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE- AC05-00OR22725.

The JGI Plant Gene Atlas project conducted by the U.S. Department of Energy Joint Genome Institute was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Full Gene Atlas data sets are available at http://phytozome.jgi.doe.gov.

Comparative Phylogenomics in KBase

Dylan Chivian*[1] (DCChivian@lbl.gov), Priya Ranjan*[3], **Adam P. Arkin**[1], **Bob Cottingham**[3], **Chris Henry**[2] and the KBase Team at the following institutions

[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 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.

Understanding the functional complements of genomes is essential to placing them in the correct ecological and evolutionary context. Interactions between species, their roles within environments, and the genetic underpinnings of individual and community phenotype greatly benefit from holistic views of the genomes of related lineages available using comparative phylogenomic methodologies. To this end, KBase is developing a suite of tools for phylogenetic and pangenomic analyses that allow the user to analyze their uploaded or KBase-assembled genomes, including microbial reference genomes from NCBI's RefSeq database.

A foundational KBase App for such work is "Build Species Tree", which places a genome or genomes of interest into a Species Tree. Instead of using a 16S tree, KBase takes advantage of the universal protein-coding phylogenetic marker genes to build more reliable species trees. After identifying related species in RefSeq, the user can then tailor a GenomeSet that combines the user's genomes with the desired reference genomes for subsequent analyses. Such work includes "hands-on" analyses such as searching the genomes for gene matches with BLAST or HMMER, building and refining multiple sequence alignments with MUSCLE and Gblocks, and building gene trees with FastTree2.

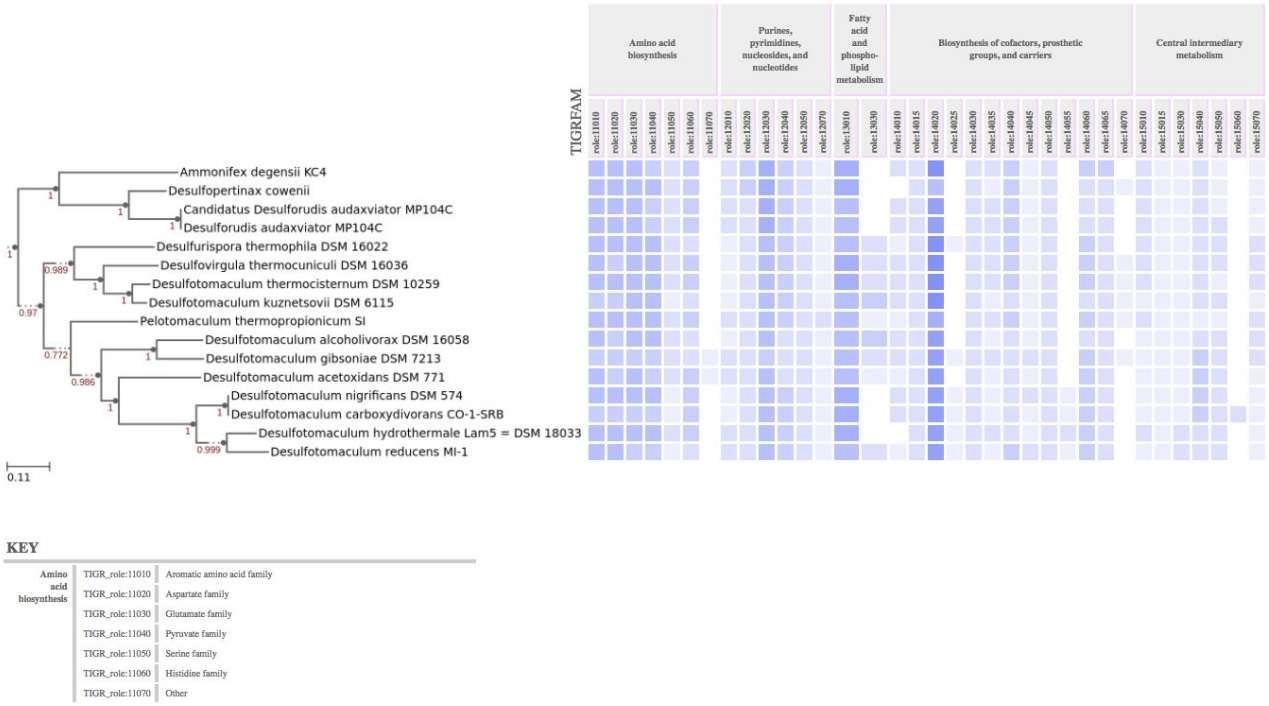


Figure 1. Functional Profiling with Gene Families

KBase has apps for computing pangenomes for a set of Genomes, such as OrthoMCL, as well as annotation with canonical protein-coding domain family content, such as COG, Pfam, and TigrFAMS. The results of these can be fed to downstream apps, including "Phylogenetic Pangenome Accumulation", gene-family-based "Functional Profiling" (Figure 1), and "Pangenome Circle Analysis", which shows visualizations of core, intermediate and singleton gene sets. These enable the user to focus in on functions, genes, and those genomes them. Additionally, KBase is developing the ability for a user to scan their genomes with custom gene family Hidden Markov Models, either created by the user or pre-developed ones, such as the dbCAN models for the CAZy database, or the antibiotic resistance gene family models of ResFam. We expect continued growth in the capabilities of the Comparative Phylogenomics apps suite as the community of KBase developers incorporates their own favorite tools as KBase apps using KBase's Software Development Kit.

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.

Connecting JGI and KBase

Shane Canon*[1] (scanon@lbl.gov), Kjiersten Fagnan[1], Harika Tandra[1], Chris Beecroft[1], **Adam P. Arkin[1]**, **Chris Henry[2]**, **Bob Cottingham[3]** and the KBase Team at the following institutions

[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 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.

To maximize the utility of KBase, users need to be able to easily import the data they are interested in exploring, analyzing, and sharing. For many researchers in the DOE Biological Science community, this data is initially generated at the Joint Genome Institute (JGI), and the ability to easily find and transfer data from JGI to KBase lowers the barrier to data analysis in KBase. JGI and KBase have collaborated closely over the past year and a half to connect KBase and JGI in a more seamless manner for our users. In this poster we will describe this effort and how we have integrated these two systems. This enhanced integration enables a user to seamlessly search JGI data from within KBase and easily import data sets of interest into private workspaces where they can conduct further analysis in KBase. This greatly improves on previous methods available to users that required them to jump between multiple web systems and had limited support for different data types.

In order to streamline and improve the data import from JGI to KBase, the two teams have heavily redesigned how the systems interconnect and created new interfaces to facilitate search and data transfers. These enhancements heavily leverage the JGI Archive and Metadata Organizer (JAMO) system which is used to store JGI data and maintain critical metadata about the data. JAMO is targeted as being the primary data tracking system for JGI and already tracks over 6 PB of data and 5.5 million files. In cooperation with KBase, JGI developed a new RESTful interface that the KBase platform uses to search data in JAMO and request transfers. This service layers on top of an ElasticSearch database that creates indexes of the data the enable rapid searching. KBase has developed an enhanced user interface (Figure 1) that enables users to specify query terms and filter on key fields like data type, Project ID, or PI. Once a user has found data of interest, they can easily transfer it to their KBase staging area (Figure 2) for import using KBase's existing suite of data importers.

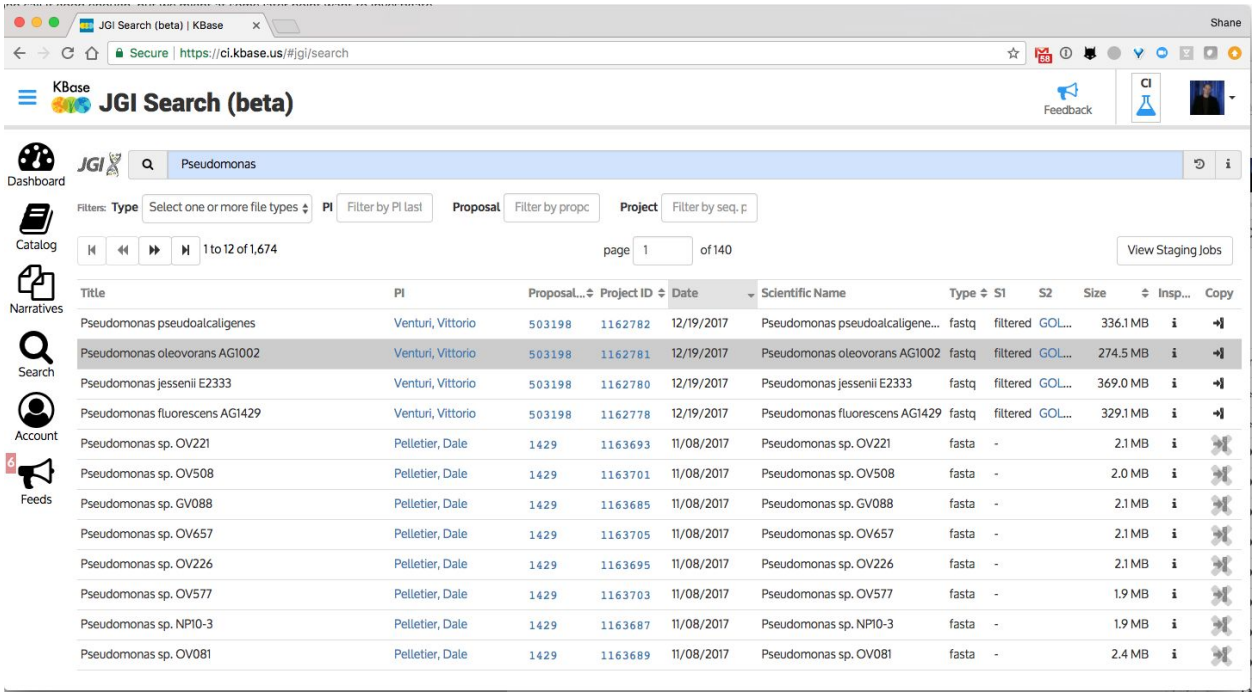


Figure 1. Improved interface for searching JGI data from inside KBase.

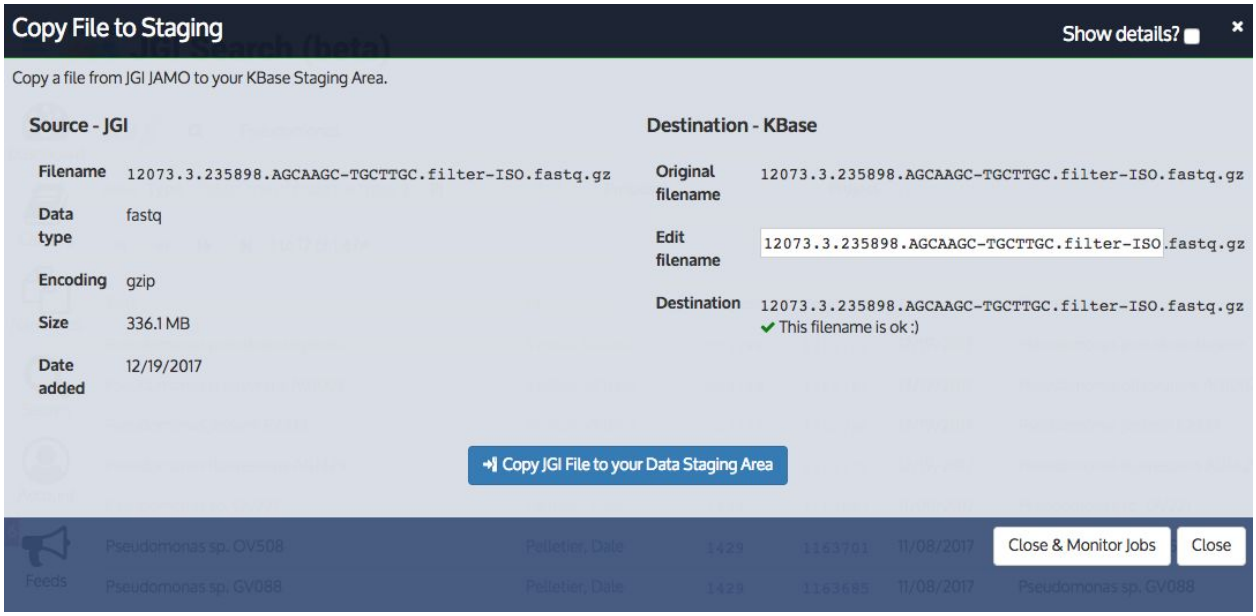


Figure 2. Staging interface that transfers JGI data to KBase for importing.

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.

Fungal Model Construction in KBase

Janaka N. Edirisinghe*[2] (janakaed@anl.gov), José P. Faria[2], **Christopher Henry[2]**, **Adam P. Arkin[1]**, **Bob Cottingham[3]** and the KBase Team at the following institutions

[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 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.

Fungal genome-scale metabolic models are an efficient way of predicting phenotypes across various environmental conditions. However, automating the construction of high-quality fungal models has been a challenge. Here we introduce a methodology to construct genome-scale fungal models in an automated fashion based on a curated set of reactions that are derived from published fungal metabolic models. As the basis for the method, we have produced a fungal model template that encompasses the biochemistry data from these published models and the structural annotations from the related genomes (Fig. 1).

The methodology 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. Once the orthologous protein families are determined, the related biochemistry data is propagated to construct a new draft metabolic model (Fig. 1). In addition, useful statistical data that was generated during the model construction process is presented as part of the method output report (Fig. 2a, Fig. 2b). This method is deployed in KBase as an app (currently still in beta) called "Build Fungal Model". The draft model that is generated by this app can be used for further analyses by utilizing a number of modeling related apps that are available in KBase.

Figure 1.

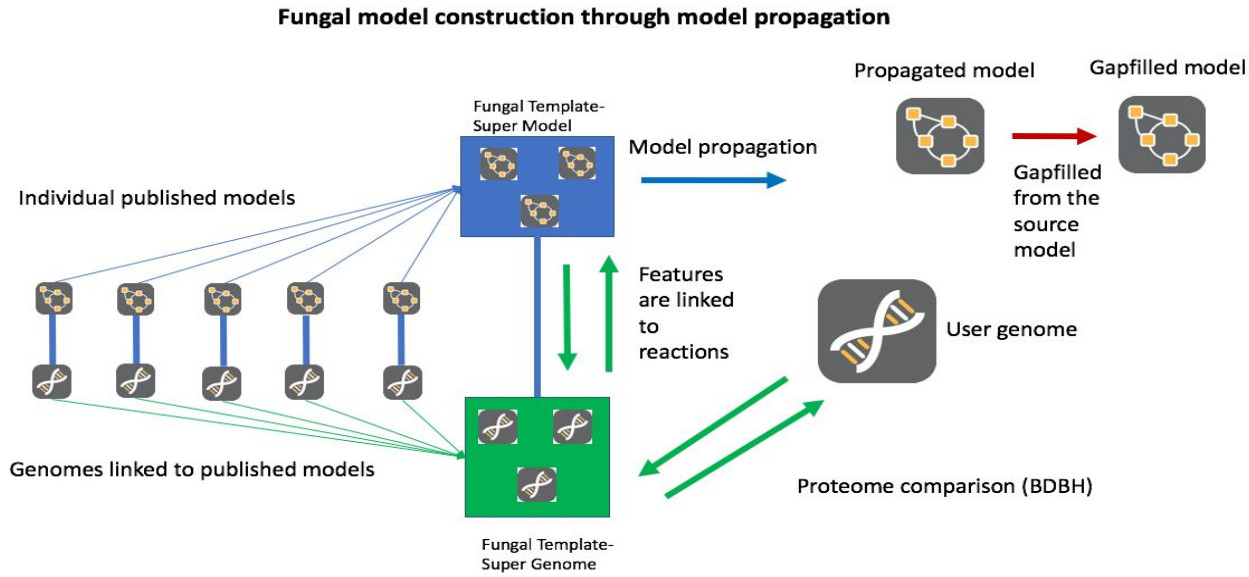


Figure 2a.

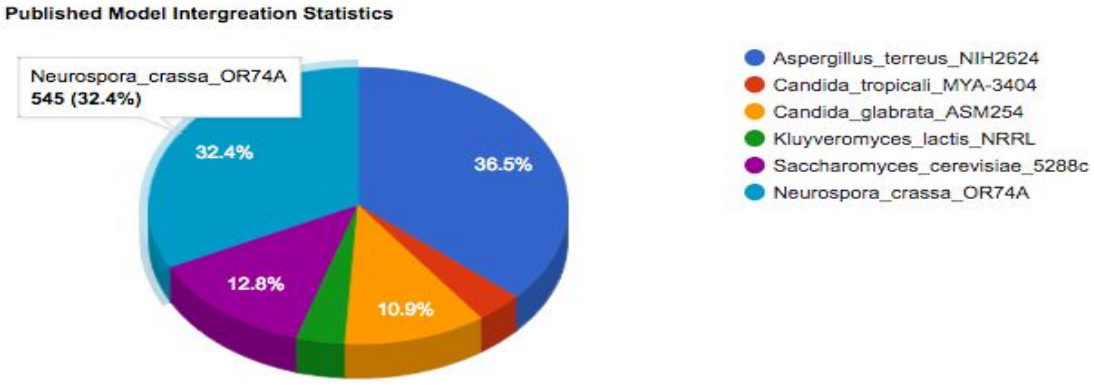
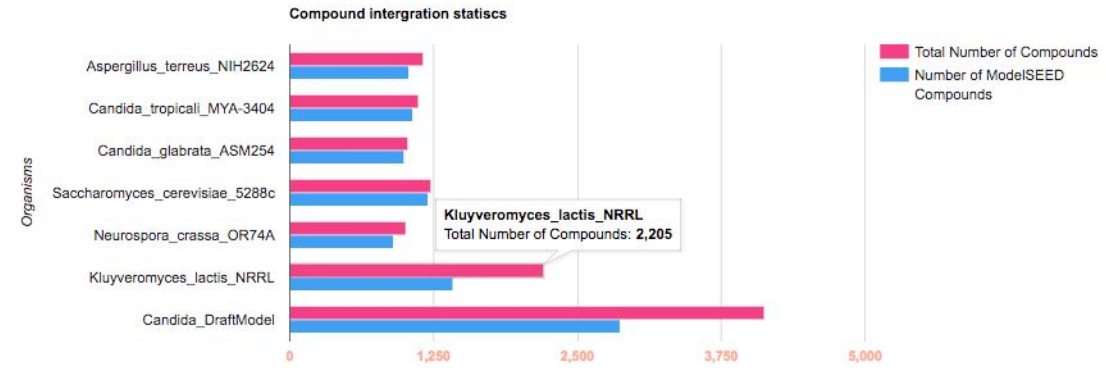


Figure 2b.



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: An Integrated Systems Biology Knowledgebase for Predictive Biological and Environmental Research

Adam P. Arkin[1], **Robert Cottingham***[3] (cottinghamrw@ornl.com), **Chris Henry**[2], Nomi Harris[1], Ben Allen[3], Jason Baumohl[1], Shane Canon[1], Stephen Chan[1], John-Marc Chandonia[1], Dylan Chivian[1], Paramvir Dehal[1], Meghan Drake[3], Janaka Edirisinghe[2], Jose Faria[2], Uma Ganapathy[4], Annette Greiner[1], Tian Gu[2], James Jeffryes[2], Marcin Joachimiak[1], Roy Kamimura[1], Keith Keller[1], Vivek Kumar[5], Sunita Kumari[5], Miriam Land[3], Sean McCorkle[4], Arman Mikaili[2], Dan Murphy-Olson[2], Arfath Pasha[4], Erik Pearson[1], Gavin Price[1], Priya Ranjan[3], William Riehl[1], Samuel Seaver[2], Alan Seleman[2], James Thomason[5], Doreen Ware[5], Shinjae Yoo[4], Qizhi Zhang[2], Diane Zheng[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 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.

The U.S. Department of Energy (DOE) has invested substantially in environmental and biological system science research to investigate the complex interplay between biological and abiotic processes that influence soil, water, and environmental dynamics of our biosphere. The community that has grown around these efforts has recognized the need to lower the barrier to accessing computational tools, data, and results, and to work collaboratively to accelerate the pace of their research. The DOE Systems Biology Knowledgebase (KBase, kbase.us) is a software platform designed to provide these needed capabilities.

KBase currently has over 160 analysis tools (see https://narrative.kbase.us/#appcatalog) that offer diverse scientific functionality for (meta)genome assembly, contig binning, genome annotation, sequence homology analysis, tree building, comparative genomics, metabolic modeling, community modeling, gap-filling, RNA-seq processing, and expression analysis (see Figure 1). Users can build and share sophisticated workflows by chaining together multiple apps–for example, one could predict species interactions from metagenomic data by assembling raw reads, binning assembled contigs by species, annotating genomes, aligning RNA-seq reads, and reconstructing and analyzing individual and community metabolic models.

Computational experiments in KBase are saved in the form of *Narratives*. A finished Narrative represents a complete record of everything the authors did to complete their analysis. This recording of a user's KBase activities within a sharable Narrative is a central pillar of KBase's support for reproducible transparent research, simplifying the re-purposing, re-application, and extension of scientific techniques.

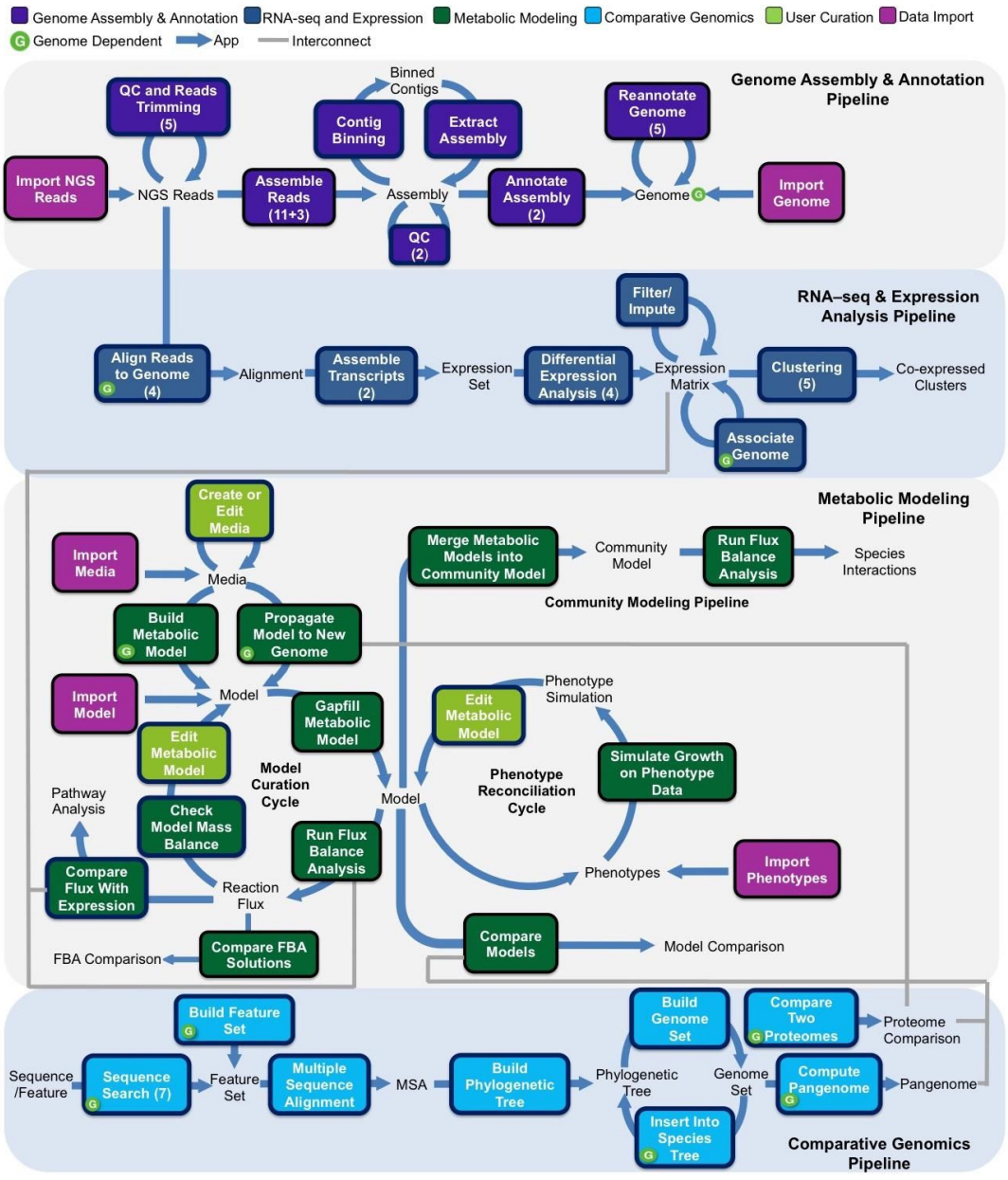


Figure 1. Outline of the major workflows and datatypes in KBase. See http://kbase.us/apps for more information.

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's App Catalog

William Riehl*[1] (wjriehl@lbl.gov), **Adam P. Arkin**[1], **Chris Henry**[2], **Bob Cottingham**[3], Nomi Harris[1] and the KBase Team at the following institutions:

[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 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.

KBase apps are analysis tools that you can use in KBase. Currently, KBase has over 160 apps (see narrative.kbase.us/#appcatalog) offering diverse scientific functionality for (meta)genome assembly, contig binning, genome annotation, sequence homology analysis, tree building, comparative genomics, metabolic modeling, community modeling, gap-filling, RNA-seq processing, and expression analysis. Apps interoperate seamlessly to enable a range of scientific workflows (see http://kbase.us/apps/). For example, a user could predict species interactions from metagenomic data by assembling raw reads, binning assembled contigs by species, annotating genomes, aligning RNA-seq reads, and reconstructing and analyzing individual and community metabolic models.

The KBase **App Catalog** lists all of the currently available apps. Each app links to a reference page (which includes technical details about the inputs and outputs) called an App Details Page, and (when available) to a tutorial that walks step by step through an example. You can click the star at the lower left of any app to add it to your "favorites." The options in the "Organize by" menu let you sort the apps by My Favorites, Run Count, Category (the default sort), and more. A "Version" menu enables toggling between those apps that are in production, and those that are currently under development by members of the community.

The App Catalog is also available in an abbreviated form in the Narrative Interface. This allows users to directly add apps to their workflow and execute them in place, while also browsing or searching the list of apps and managing their favorites.

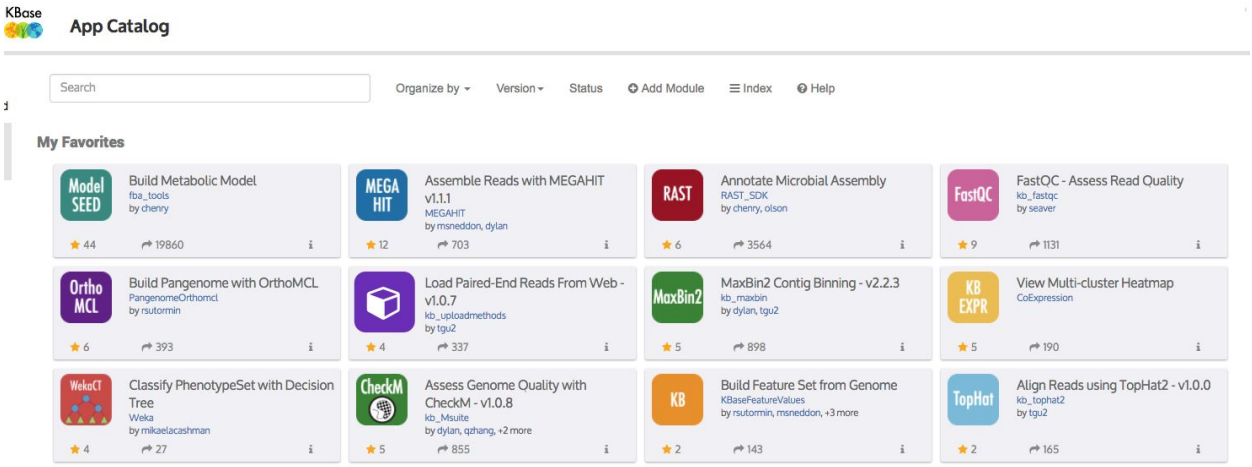


Figure 1. A portion of the KBase App Catalog **(**narrative.kbase.us/#appcatalog) showing the apps that a particular user has "favorited".

KBase was designed to be an extensible community resource. This extensibility is supported by the KBase Software Development Kit (SDK), which is a set of command-line tools and a web interface (part of the App Catalog) that allow any developer to build, test, register, and deploy new or existing software as KBase apps, thereby extending the platform's scientific capabilities. The number of apps available in KBase is expected to increase rapidly as members of the community use our SDK to add their analysis tools to the KBase platform.

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's Narrative Interface: A User Interface for Creating Reproducible Systems Biology Workflows

Nomi Harris*[1] (nlharris@lbl.gov), William J. Riehl[1], Erik Pearson[1], **Adam P. Arkin[1]**, **Chris Henry[2]**, **Bob Cottingham[3]**, and the KBase Team at the following institutions:

[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 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.

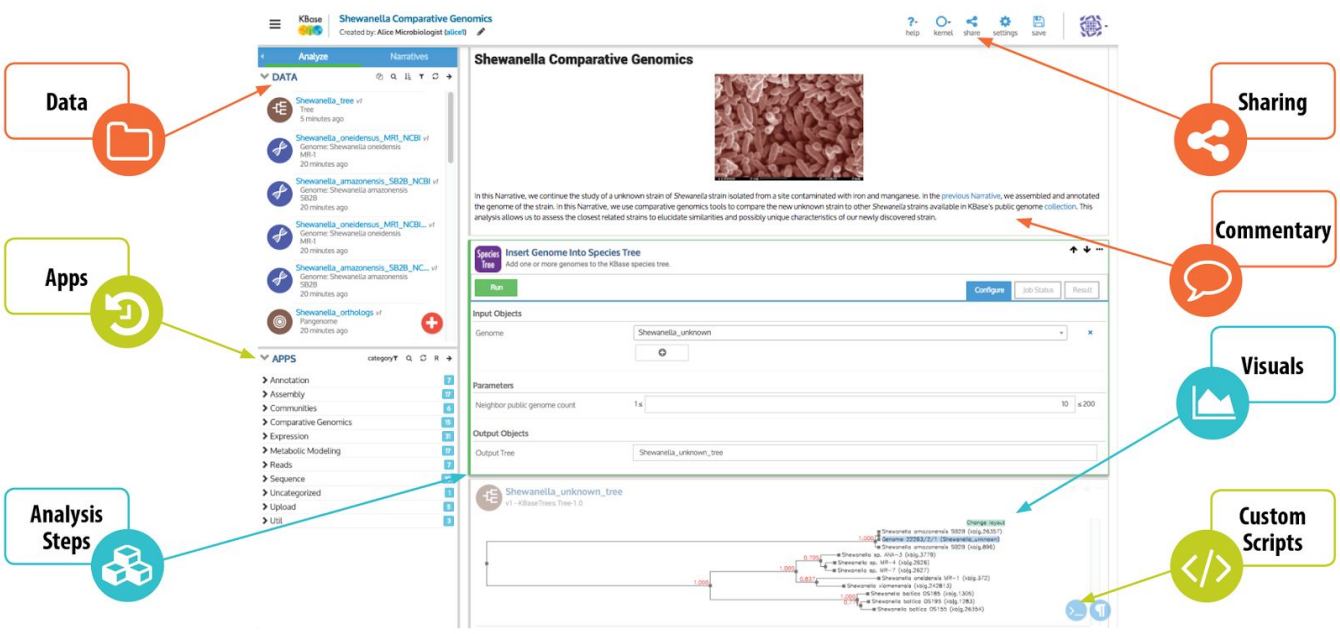


***Figure 1**. A KBase Narrative is an interactive, dynamic, and persistent document created by users that promotes open, reproducible, and collaborative science.*

KBase's Narrative User Interface (Fig. 1) enables users to (i) upload their private data; (ii) search and retrieve extensive public reference data; (iii) access data shared with them by others; (iv) share their own data with others; (v) select and run applications on their data; (vi) view and analyze the results from those applications; and (vii) record their thoughts and interpretations along with the analysis steps.

All of these activities take place within a point-and-click "notebook" environment—the Narrative interface is built on top of the Jupyter Notebook platform. When a user begins a new computational experiment in KBase, they create a new "notebook" to hold this experiment, which we call a "Narrative". Every action performed by a user shows up as a new Narrative "cell". App cells show the chosen input parameters for the application and the results of the analysis. Apps interoperate seamlessly to enable a range of scientific workflows. In addition to running apps, users can create and run blocks of code within a Narrative using "code cells", which allows them to, for example, run large-scale studies in KBase (e.g., building thousands of models at once). Users can also leverage the flexibility of code cells to add custom analysis steps that are not yet available as KBase apps. Markdown cells allow users to add formatted text and figures to a Narrative to annotate and describe the thought process behind the scientific workflow being crafted.

A finished Narrative represents a complete record of everything the authors did to complete their computational experiment. Once a Narrative has been shared (or made public), other users can copy the Narrative and rerun it on their own data, or modify it to suit their scientific needs. Thus, public Narratives serve as resources for the user community by capturing valuable data sets, associated computational analyses, and scientific context describing the rationale behind a scientific study. This recording of a user's KBase activities within a sharable Narrative is a central pillar of KBase's support for reproducible transparent research, simplifying the re-purposing, re-application, and extension of scientific techniques.

The kbase.us website has a number of resources to introduce you to the Narrative Interface. The Narrative Interface guide provides a detailed explanation of the features of the interface. Video tutorials, such as "KBase Quickstart", are available on our YouTube channel. Narrative tutorials, such as "Assembly and Annotation of Prokaryotic Genomes", are interactive tutorials that you can copy and re-run on the example data or your 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.

Learning to Use KBase to Accelerate Your Research

Meghan Drake*[3] (drakemm@ornl.gov), Nomi Harris[1], **Adam P. Arkin[1]**, **Chris Henry[2]**, **Bob Cottingham[3]** and the KBase Team at the following institutions:

[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 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.

KBase's integration of data and tools and the ease of creating and running large-scale analysis workflows have the potential to empower scientists in a broad range of application areas for systems biology, including environmental analysis and biosystems design. KBase's sharing capabilities amplify this potential by enabling scientists with differing expertise to work together and leverage each other's work. Science done within KBase has been published in over 30 peer-reviewed publications (see http://kbase.us/publications), including (1) reconstruction of >8000 models of core metabolism across the microbial tree of life; (2) reconstruction of semi-curated metabolic models for 773 gut microbes; (3) predicting trophic interactions within a microbial community; and (4) reconstruction of regulons from expression data. The Narratives associated with many of these publications can be found in our Narrative Library.

If you've never used KBase before, we suggest you start with the "New to KBase?" page (Fig. 1). The Narrative Interface guide and the Data Search guide will teach you about those key KBase interfaces. A printable four-page KBase brochure is also available. Video tutorials, such as "KBase Quickstart", are available on our YouTube channel. Narrative tutorials, such as the one about Assembly and Annotation of Prokaryotic Genomes, are step-by-step walkthroughs that demonstrate how to use KBase to perform bioinformatics analysis workflows. Not only can you see an example of how to assemble and annotate data in KBase; you can make your own copy of the Narrative tutorial and run the steps yourself, even changing the parameters or uploading your own datasets to analyze.

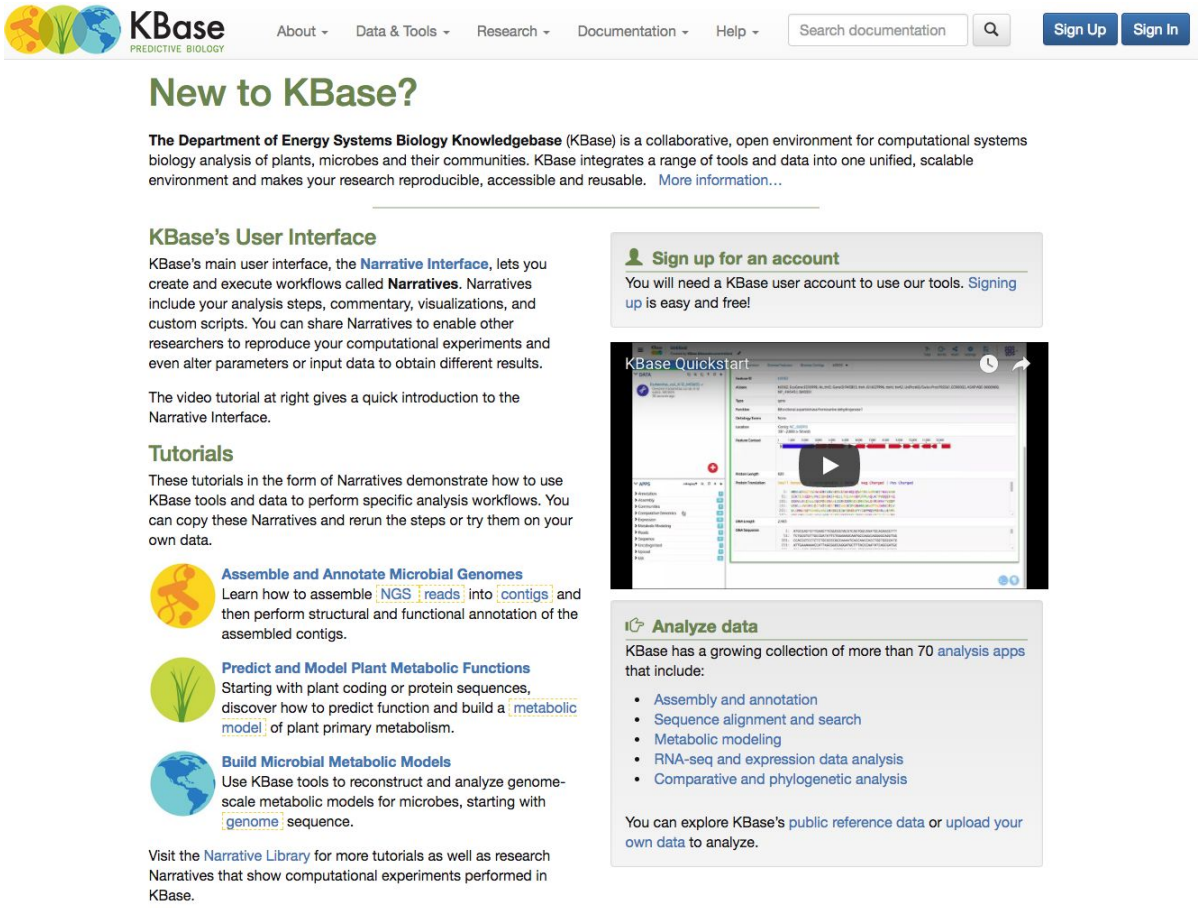


Figure 1: The "New to KBase" web page has pointers to videos, Narrative tutorials and other training materials that will help you learn how to use KBase to accelerate your research.

The KBase team presents at various conferences and conducts workshops to train users and developers. If you would like to host a KBase workshop at your organization, please contact us. We also offer webinars to reach a wider audience.

If you have questions about KBase, or you want to report a bug or request a new feature, you can join the KBase Help Board. In the Help Board, you'll be able to:

- Submit bug reports, questions and suggestions without revealing your email address.
- Engage in a two-way dialog with KBase staff as they track down and resolve your issue.
- Search the issues submitted by other users. If someone else has already reported your problem or suggested the same new feature, you can add a "Me too!" comment.
- Increase your productivity in KBase by seeing other users' suggestions and workarounds.

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.

Microbiome Genome Extraction and Metabolic Modeling of Species Interactions in KBase

Dylan Chivian*[1] (DCChivian@lbl.gov), José P. Faria*[2] (JPLFaria@anl.gov), **Adam P. Arkin[1]**, **Bob Cottingham[3]**, **Chris Henry[2]**, and the KBase Team at the following institutions

[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 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.

KBase was designed to enable systems biology analysis of communities of microbes and/or plants. KBase has powerful tools for metabolic modeling and comparative phylogenomics of microbial genomes that can be used for developing mechanistic understanding of functional interactions between species in microbial ecosystems. Essential to this process is obtaining high-quality genomes to annotate, either via cultivation or genome extraction from metagenome assembly. KBase is developing 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, community member interaction covariation with SparCC, 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 genome completeness and contamination assessment with CheckM. The apps that have been incorporated to date represent a single offering in each of these analysis steps; we expect additional apps will be incorporated into KBase that offer alternate algorithms to allow users to obtain variation in their results, such as alternate binning tools.

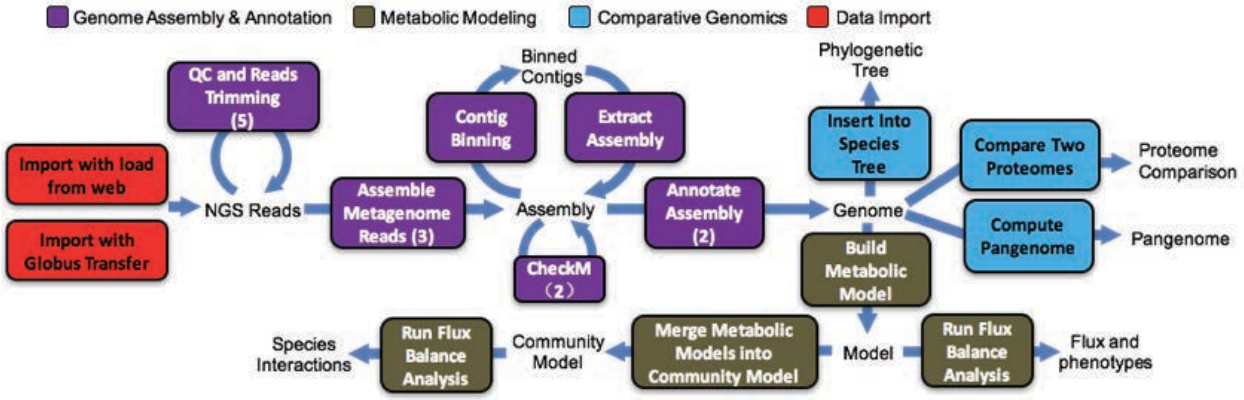


Figure 1. KBase's microbiome analysis pipeline.

KBase's microbiome analysis pipeline (see Fig. 1) begins with the upload of reads files. Next, the user can apply one of the metagenome assembly apps and bin the assembled contigs so that individual genomes can be extracted from the bins. Once the high-quality individual genomes are extracted, they 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 KBase reference genomes, metabolic modeling, and RNA-seq alignment. After generating metabolic models from the genomes assembled from a metagenome, individual metabolic models can be combined into a community metabolic model, which can be applied with the Flux Balance Analysis app to predict trophic interactions between species. Users have applied these tools to study: (i) interactions between plants and microbes in soil; (ii) why some microbes form stable communities; (iii) how a microbial community cooperates to produce a specific product; and (iv) how a community of heterotrophic species can feed on byproducts from an autotroph to grow autotrophically.

KBase's current microbiome analysis pipeline is just a starting point. It works best on microbial communities with limited diversity, and presently can only be applied to shotgun metagenomic data. In the future, we plan to support analysis of 16S amplicon data, as well as enabling further direct taxonomic and functional profiling of shotgun metagenomic reads.

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.

RNA-Seq Analysis in KBase

Sunita Kumari*[1] (kumari@cshl.edu), Vivek Kumar[1], Priya Ranjan[2], Doreen Ware[1], **Bob Cottingham[2], Chris Henry[3] Adam P. Arkin[4]**, and the KBase Team at the following institutions

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

http://kbase.us

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.

RNA-seq analysis is emerging as one of the most powerful approaches for assessing differential gene expression. RNA-seq uses next-generation sequencing to account for all the transcripts in one or more biological samples at a particular time. It can be used for a variety of applications such as transcriptome assembly, gene discovery/annotation, and detection of differential transcript abundances between tissues, developmental stages, genetic backgrounds and environmental conditions. The overarching goal of the RNA-seq pipeline in KBase is to create differential expression estimates and use these to inform metabolic models and to perform functional analysis of genes with similar expression patterns.

RNA-seq analysis typically consists of (i) mapping short sequence reads to the reference genome; (ii) assembling the transcripts into full-length transcripts and expression quantification; and (iii) differential analysis of the gene expression. KBase provides a set of apps that allow users to run the tools from the popular Tuxedo RNA-seq suites to generate the normalized full and differential expression matrix of the reads obtained from Illumina sequencing platforms using the reference prokaryotic and eukaryotic genome. The RNA-seq apps in KBase can be combined into multiple workflows, allowing users to select their choice of reads aligner and assembler for the differential gene expression analysis. For alignment of the reads to the reference genome, transcriptome profiling, and identification of differentially expressed genes, the original Tuxedo suite uses the tools TopHat2, Cufflinks, and Cuffdiff respectively; the new Tuxedo suite uses HISAT2, StringTie and Ballgown.

All of these tools are available as KBase apps; detailed usage instructions can be found at http://kbase.us/transcriptomics-and-expression-analysis/ .

There are three Narrative tutorials that demonstrate how to use the KBase RNA-seq pipeline end-to-end on plant and microbial reads. You can copy and re-run them to become acquainted with building RNA-seq analysis workflows in KBase.

- Arabidopsis RNA-seq Analysis using Original Tuxedo Suite: https://narrative.kbase.us/narrative/ws.19393.obj.1
- Arabidopsis RNA-seq Analysis using New Tuxedo Suite: https://narrative.kbase.us/narrative/ws.19391.obj.1
- E. coli RNA-seq Analysis https://narrative.kbase.us/narrative/ws.19340.obj.16

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 KBase Platform

Shane Canon*[1] (scanon@lbl.gov), **Adam P. Arkin[1]**, **Chris Henry[2]**, **Bob Cottingham[3]** and the KBase Team at the following institutions

[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 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.

In order to deliver on KBase's ambitious goals, the project has developed a robust, open-source, highly-extensible platform that leverages various cutting edge technologies. This service-oriented platform consists of several interconnected subsystems that handle authentication, data storage and access, job execution, user interaction, and SDK services. These services provide the key functionality to enable reproducibility and sharing, and provide the foundation for supporting knowledge propagation. In this poster we will describe the overall architecture and some of its innovative features.

The KBase platform consists of a collection of interacting distributed services backed by multiple databases, storage components, and computational resources. We divide the system into core services that provide the low-level infrastructure and software development kit (SDK) modules that capture the specific science functionality and data models. The core services include components such as low-level data services, SDK support services, execution engine, user interface services, and (in progress) knowledge services. Figure 1 shows the schematic layout of the architecture.

Two key innovations of KBase are the Narrative Interface and the SDK. The Narrative Interface builds on top of the popular Jupyter Notebook platform and provides the key user interface for KBase. KBase's enhancements allow users to easily interact with data, perform drag-and-drop analysis on their data, and share and publish results. Another key innovation is the KBase SDK which enables external developers to easily add functionality to KBase by adding new applications or services that are integrated into the platform. The KBase SDK heavily leverages

the Docker Container technology to allow these extensions to be compartmentalized and reliably executed.

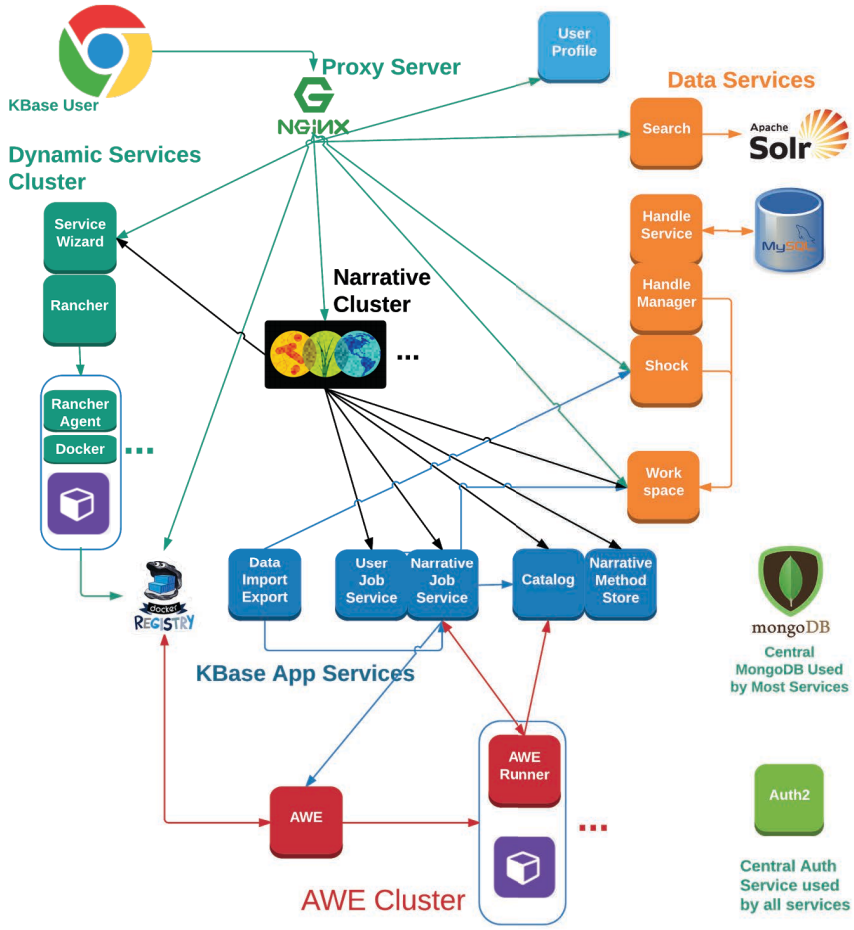


Figure 1: Schematic overview of the KBase platform architecture.

KBase leverages HPC resources provided by the DOE ASCR Facilities like NERSC and the Leadership Computing Facilities at Oak Ridge and Argonne. These resources open the door to running large-scale analysis and provide additional capacity to complement the limited dedicated compute resources in KBase. To enable this, the KBase execution engine and SDK have been extended to offer basic support for HPC. These enhancements enable SDK developers to create optimized applications that can run at large scales by KBase users to tackle large scale assembly, comparative analysis and other grand challenge problems.

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 KBase Software Development Kit Makes KBase an Extensible Community Resource

Paramvir Dehal*[1] (psdehal@lbl.gov), Shane Canon[1], **Adam P. Arkin[1]**, **Christopher Henry[2]**, **Bob Cottingham*[3]**, Nomi Harris[1] and the KBase Team at the following institutions

[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 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.

KBase was designed to be an extensible community resource. This extensibility is supported by the KBase Software Development Kit (SDK), which enables external developers to incorporate third-party open source tools as KBase apps. Such third-party apps are beginning to appear in the system, and the pace of development is increasing.

The KBase SDK is a set of command-line tools and a web interface that enable any developer to build, test, register, and deploy new or existing software as KBase apps. All SDK apps and any underlying tools that are registered in the KBase system must adhere to a standard open-source license (opensource.org/licenses). Information about the app developer is maintained in the documentation for that app so credit can be given to the contributor. Data provenance, job management, usage logging, and app versioning are handled automatically by the platform, allowing developers to wrap new scientific tools quickly with minimal KBase-specific training. Other existing platforms offer similar support for third-party development, but KBase's data model provides the additional benefit of improving interoperability of third-party applications by imposing a single data format and specification on all data types consumed or produced by each app.

A number of external developers have used the SDK to add new KBase apps. These include:

- "Classify PhenotypeSet with Decision Tree" (see Fig. 1), which builds a decision tree in order to classify results of a phenotype set to understand the effect the media compounds have on any defined phenotype (Mikaela Cashman, University of Nebraska)

- SBMLTools, tools to import and integrate published metabolic models (Felipe Liu and Sonia Rocha, University of Minho)
- optStoic (in development), which optimizes metabolic networks for the production of chemical products and biofuels (Costas Maranas and Anupam Chowdhury, Pennsylvania State University)

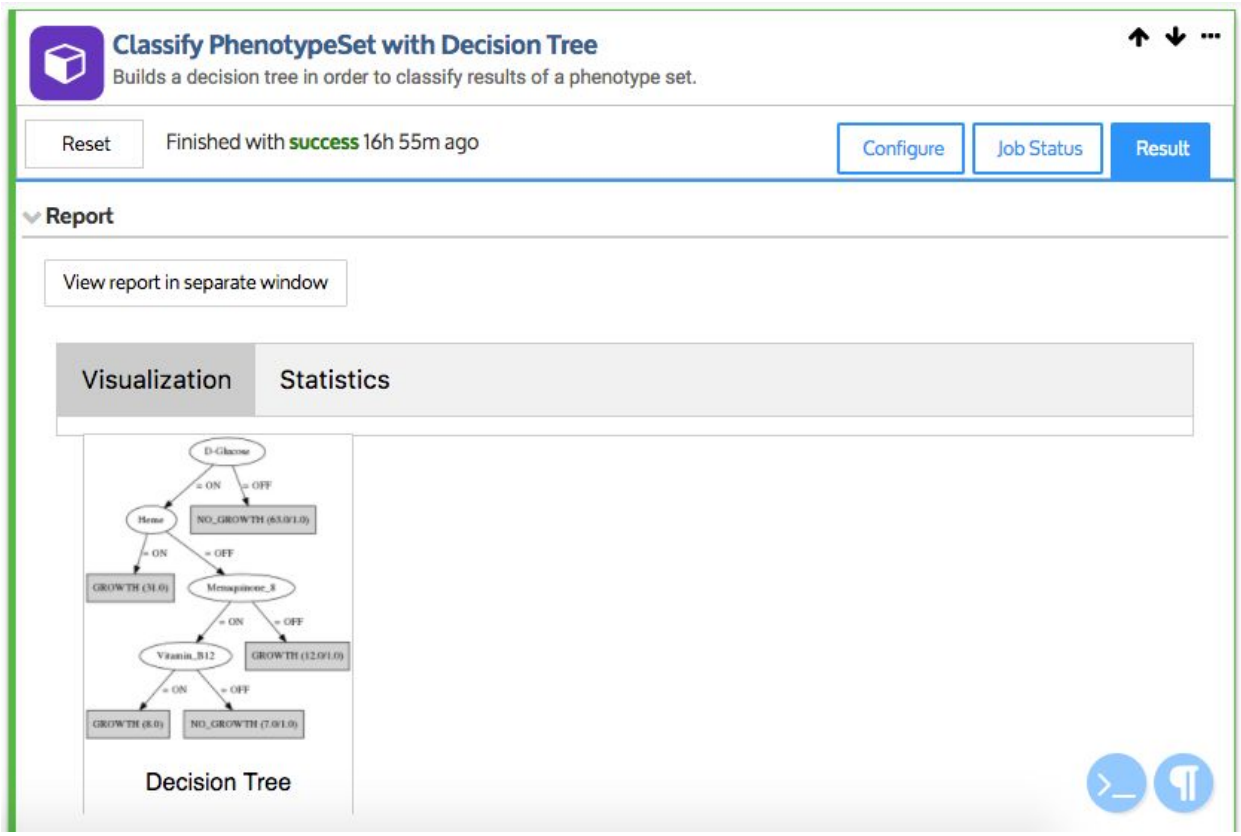


Figure 1. The app called "Classify PhenotypeSet with Decision Tree" was added by third-party developer Mikaela Cashman using the KBase SDK.

These third-party apps, along with many others wrapped by KBase-internal and third-party developers with the SDK, can be found in the KBase App Catalog (https://narrative.kbase.us/#appcatalog). More information about the KBase SDK is available at https://github.com/kbase/kb_sdk/blob/master/README.md.

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.

Tools and Data for Metabolic Modeling in KBase

José P. Faria*[2] (jplfaria@anl.gov)**, Christopher Henry[2]**, Janaka Edirisinghe[2], **Adam P. Arkin[1]**, **Bob Cottingham[3]** , Nomi Harris[1] and the KBase Team at the following institutions

[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 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.

KBase offers "one-stop shopping" for a growing range of integrated analysis tools and datasets that enable assembly, annotation, construction, and comparison of genome-based metabolic models. One of the most powerful and popular workflows in KBase starts with genomic sequence reads, goes through pipelines for assembly and annotation, and culminates in metabolic models that can be used to evaluate an organism's metabolic capabilities by simulating growth under different conditions. KBase's metabolic modeling toolkit supports the reconstruction, prediction, and design of metabolic networks in microbes and plants (and their communities). These tools can help advance efforts to optimize microbial production of a certain biofuel, find the minimal media conditions under which that fuel is generated, or predict species interactions within a soil or industrial microbiome.

In KBase, genome-scale metabolic models are primarily reconstructed from functional annotations produced by the KBase annotation apps (see figure). When a genome is functionally annotated, its metabolic genes are mapped onto biochemical reactions. This information is integrated with data about reaction stoichiometry, subcellular localization, biomass composition, estimation of thermodynamic feasibility (directionality of reactions) to produce a detailed stoichiometric model of metabolism.

KBase users have applied the system to address a range of scientific problems. Some examples of published studies that used KBase's metabolic modeling toolkit include finding patterns in the presence of core metabolic pathways across the microbial phylogenetic tree; identifying inconsistent growth conditions in Klebsiella KPPR1; and several studies that analyzed the interactions within communities of microbes and plants. The Narratives that researchers have

chosen to share publicly (see http://kbase.us/narrative-library) can be viewed, copied, and re-run, with varying parameters or new datasets.

Figure 1. KBase's metabolic modeling tools (green) as well as some other analysis tools that interoperate with them.

Resources for learning more about KBase's metabolic modeling suite include:

- Home page about modeling in KBase: http://kbase.us/metabolic-modeling-in-kbase/
- Narrative tutorial: "Microbial Metabolic Model Reconstruction and Analysis" (https://narrative.kbase.us/narrative/ws.18302.obj.61). You can copy the Narrative and rerun the steps yourself.
- Video tutorial: visit the KBase YouTube channel (https://www.youtube.com/channel/UCTX9Pn_WCCCpOER9-3AGL_A)
- Modeling FAQ: http://kbase.us/metabolic-modeling-faq/
- Narrative Library: http://kbase.us/narrative-library

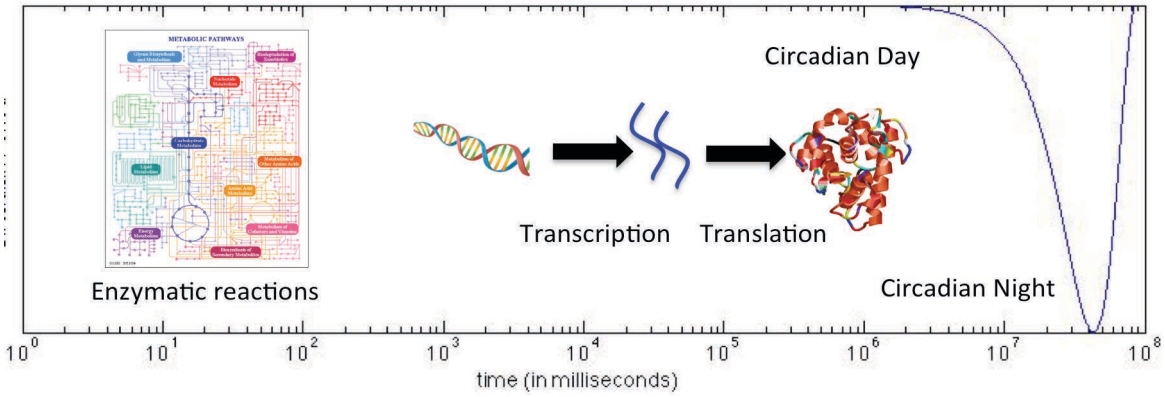
KBase is funded by the Genomic Science program within the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

Multi-scale Modeling of Circadian Rhythms: From Metabolism to Regulation and Back

William Cannon*[1] (bill@pnnl.gov), Jennifer Hurley[2], Meaghan Jankowski[2], Hannah De Los Santos[2], Sam Fordyce[2], Jeremy Zucker[1], Neeraj Kumar[1], Christina Kelliher[3], and **Jay Dunlap**[3]

[1]Pacific Northwest National Laboratory, Richland, WA; [2]Rensselaer Polytechnic Institute, Troy, NY; [3]Geisel School of Medicine, Dartmouth College, Hanover, NH

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 10^{0} s^{-1}) to gene and protein regulation (~20 minutes) to circadian rhythms (24 hours).



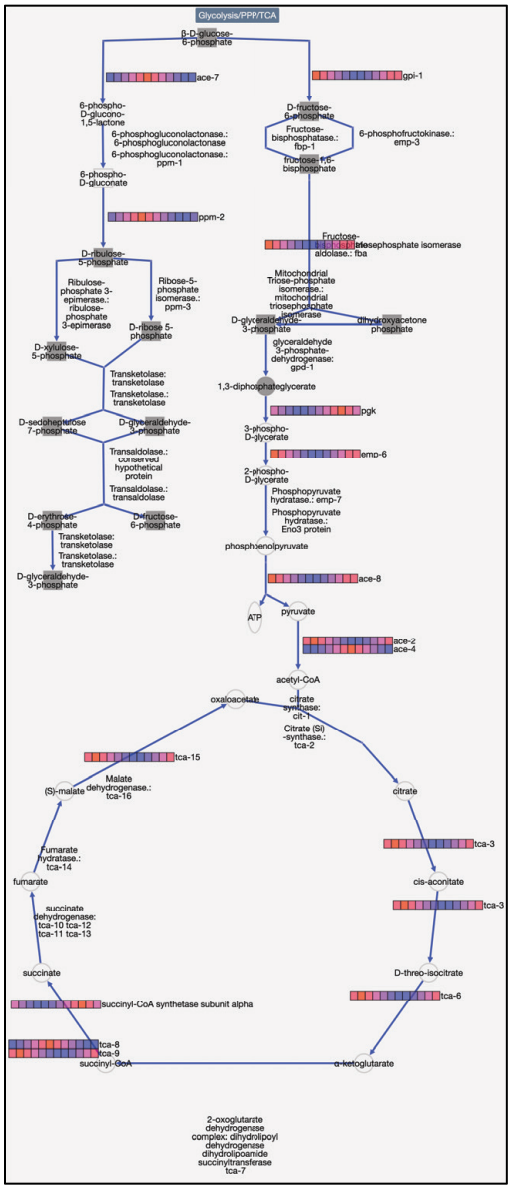
Timescales that the simulations using statistical thermodynamics will cover. Enzymatic reactions occur on the millisecond to second timescale while gene and protein expression occur on the minute to ~30 minute scale and the circadian rhythm occurs over a period of 24 hours.

To accomplish this, we are:

- Implementing an approach to the law of mass action that uses chemical potentials rather than rate constants. This approach involves a rescaling of the fast degrees of freedom, resulting in a compression of the time-dependence to fewer relative scales. Steady state processes can be 'telescopically' modeled to address the scale of interest while collapsing faster scales.
- Using the new method to understand the relationship between central metabolism and circadian rhythms in *Neurospora crassa* by using a multi-scale model of metabolism that will include regulation of the circadian clock.

Abstract: Cell metabolism is modeled using fundamental principles from which necessary kinetic parameters and regulation points can be derived when experimental data is not available [1]. The principle of maximum entropy production, a consequence of the second law of thermodynamics, is used to infer rate parameters for simulating the mass action kinetics of metabolism. Simulation predictions of metabolite levels of central metabolism of *Neurospora crassa* then allows for inference of post-translational enzyme regulation. Subsequent simulations

with post-translational regulation provide predictions of metabolite levels that are comparable to experimental measurements. The simulation of glycolysis + TCA cycle, and separately the pentose phosphate and TCA cycle pathways (right) provide free energy maps of metabolic pathways, fluxes and power characteristics of pathways and individual reactions. Importantly, flux through the former increases NADH concentration while flux through the latter increases NADPH concentration needed for growth.



Importantly, environmental controls are known to extensively modulate cellular metabolism in fungi and lead to pre-translational regulation as well [2]. *Neurospora* is commonly used in the study of metabolic regulation and it is also the principal fungal model system for the study of the effects of light and circadian clocks. Our work has found that the clock extensively regulates metabolism via protein expression and may, in turn, be regulated by metabolism itself. Our initial assessment of the expression of central metabolism enzymes suggests that glycolysis and the TCA cycle, generating ATP and NADH, are most active during the circadian dusk, while the pentose phosphate pathway, generating NADPH, is most active during the circadian dawn (right, protein expression indicated by heatmaps). Yet to be determined is whether metabolism oscillates independently between glycolysis and the pentose phosphate pathways, or whether metabolism is controlled by regulation of expression at the protein level [3].

References

1. Cannon, W.R. and S.E. Baker, *Non-steady state mass action dynamics without rate constants: dynamics of coupled reactions using chemical potentials.* Phys Biol, 2017. **14**(5): p. 055003.
2. Hurley, J.M., et al., *Analysis of clock-regulated genes in Neurospora reveals widespread posttranscriptional control of metabolic potential.* Proc Natl Acad Sci U S A, 2014. **111**(48): p. 16995-7002.
3. Hurley, J.M., J.J. Loros, and J.C. Dunlap, *The circadian system as an organizer of metabolism.* Fungal Genet Biol, 2016. **90**: p. 39-43.

Funding statement. 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.

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, extremely nitrogen (N) -poor soils in the arid Chilean Andes. 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" Chilean species and compare their genomes 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 "triplicate species" will be used to fuel a novel phylogenomic analysis that will identify the genes that support the evolutionary divergence of the extreme survivors in Chile 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, extremely Nitrogen (N) poor soils in the arid Chilean Andes. It uses a previously validated phylogenomic pipeline we developed called PhyloGeneious [1], which can identify genes that provide positive support to species divergence. By applying this phylogenomic pipeline to the gene sequences of these "triplicate species", we can identify the genes that distinguish these "extreme survivors" in Chile from their related species adapted to similarly dry regions in California (CA) not constrained by N and/or water availability. These "extreme survivor" species 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.

Key to our phylogenomic approach is the "triplicate species" sampling strategy. To maximize our ability to separate the trait-relevant signature from overall speciation events, our "triplicate species" sampling will cover 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 now extend this phylogenomic approach to the study of "marginal survivor strategies" as follows:

Aim 1. **Species collection and deep transcriptome sequencing**: Sample deep transcriptomes of 28 triplicate sister species pairs. (Chile "marginal survivors" and their closest relatives from CA) (NYU, NYBG, Chile). ***Progress:*** We performed deep RNA sequencing on all 28 Chilean species and 39 of their CA relatives (Table 1), with resulting average gene coverage of 87%, based on the BUSCO conserved single-copy orthologs existence assessment.

Aim 2. **Phylogenomic Analysis**: Phylogenomic analysis of 82 "triplicate species" to identify genes that repeatedly support nodes that distinguish the extreme survivors in Chile from their sister species in CA (AMNH, NYU). ***Progress:*** We have optimized our PhyloGeneious phylogenomic pipeline for volume and speed to enable the analysis of this large taxa set (90 species, including the outgroups and the model plant species for functional annotations).

Aim 3. **Network Analysis**: Combine phylogenomics (protein sequence) and gene networks (gene expression) to identify genes and network modules associated with adaptations to marginal, low-N soils (NYU, Chile). ***Progress:*** To exploit a comparative analysis of gene regulatory networks, we are currently developing a new module called **PhyloExpress** that extends the PhyloGeneious pipeline to include gene expression data.

Aim 4. **Functional Validation**: 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 most promising candidates from our preliminary analysis using our Chilean set and their closest sequence available sister species.

Table 1. Extreme survivor species in Chile (green) and their "sister" species in CA (Drought - yellow; Mesic - blue). *Our project studies 28 triplicates of species from Marginal (Dry +low-N, Chile), Dry (Dry, California), and moist (Mesic, California) soils. All the Chilean species and 39/54 of the California species have already been sequenced while 15/54 California species are being collected.*

Chile Marginal Survivor Species (Drought + low-N)	California Species (Drought)	California Species (Mesic)
Atriplex imbricata	*Atriplex lentiformis*	*Atriplex watsonii*
Mulinum crassifolium	*Sanicula crassicaulis*	*Conium maculatum*
Ambrosia artemisioides	*Ambrosia chamissonis*	*Ambrosia psilostachya*
Baccharis boliviensis and Baccharis tola	*Baccharis glutinosa*	*Baccharis salicifolia*
Trichocline caulescens	*Cirsium occidentale*	*Cirsium fontinale*
Chuquiraga atacamensis	*Coreopsis douglasii*	*Rudbeckia californica*
Parastrephia quadrangularis	*Grindelia hirsutula*	*Eriophyllum confertiflorum*
Senecio puchii	*Senecio californicus*	*Senecio mikanioides*
Tagetes multiflora	*Pectis papposa*	*Pluchea sericea*
Phacelia pinnatifida	*Eriodictyon tomentosum*	*Phacelia nemoralis*
Pycnophyllum bryoides	*Cerastium viride*	*Cerastium beeringianum*
Adesmia spinosissima	*Amorpha californica*	*Amorpha fruticosa*
Lupinus oreophilus	*Lupinus nana*	*Lupinus arboreus*
Lupinus subinflatus	*Lupinus hirsutissimus*	*Lupinus latifolius*
Aristida adscensionis	*Danthonia unispicata*	*Danthonia californica*
Bouteloua simplex	*Bouteloua curtipendula*	*Muhlenbergia filiformis*
Calamagrostis crispa	*Calamagrostis rubescens*	*Calamagrostis breweri*
Calamagrostis cabrerae	*Festuca californica*	*Festuca subuliflora*
Munroa decumbens	*Munroa squarrosa*	*Munroa utilis*
Nassella nardoides	*Nassella cernua*	*Nassella manicata*
Jarava frigida	*Stipa coronata*	*Stipa kingii*
Chorizanthe conmisuralis	*Chorizanthe palmeri*	*Rumex crispus*
Exodeconus integrifolius	*Lycium cooperi*	*Physalis lancifolia*
Fabiana denudata	*Nicotiana glauca*	*Petunia parviflora*
Solanum chilense	*Datura wrightii*	*Solanum douglasii*
Acantholippia deserticola	*Aloysia wrightii*	*Phyla nodiflora*
Fagonia chilensis	*Fagonia laevis*	*Tribulus terrestris*

References

1. Lee E et. al., A functional phylogenomics view of the seed plants. PLoS Genet 7(12):e1002411.
2. Delaux et. Al., Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. PloS Genet 10(7):e1004487.

The EvoNet grant is supported by the Genomics Science program within the Office of Biological and Environment Research in the DOE Office of Science.

Using iDIRECT (Inference of Direct and Indirect Relationship with Efficient Copula-based Transitivity) to improve network modelling of rhizosphere microbial assemblages

Naijia Xiao,[1*] (naijia.xiao@ou.edu), Daliang Ning,[1] Aifen Zhou,[1] Megan Kempher,[1] Zhou Jason Shi,[2,3] Benjamin Zhou,[4] Shengjing Shi,[1,5] Erin Nuccio,[5,6] **Mary Firestone[5,7], and Jizhong Zhou[1,7,8]**

[1]University of Oklahoma, Norman; [2]University of California, San Francisco; [3]Chan Zuckerberg Biohub, San Francisco, CA; [4]Stanford University, Stanford, CA; [5]University of California, Berkeley, [6]Lawrence Livermore National Laboratory, Livermore, CA; [7]Lawrence Berkeley National Laboratory, Berkeley, CA; and [8]Tsinghua University, Beijing, China

Project Goals: The overall goals of our project are to advance our understanding of the complex interactions controlling C flow in the rhizosphere by addressing two fundamental topics: 1) How multi-trophic interactions control soil C dynamics, and 2) How changing precipitation regimes alter these interactions, and thus impact flow and fate of soil C.

Networks are a vital tool for understanding and modeling the interactions of complex systems in science and engineering, and direct and indirect interactions are pervasive in all networks including those that occur in rhizosphere soil. However, quantitatively discriminating direct and indirect relationships in association networks remains a formidable task. Here, we present iDIRECT (**I**nference of **D**irect and **I**ndirect **R**elationship with **E**ffective **C**opula-based **T**ransitivity), a novel statistical approach for inferring direct dependences on association networks. Specifically, we first developed new algorithms using Copula-based transitivity, which overcomes several challenging mathematical problems inherent in previous network inference approaches, such as singularity or ill conditioning, self-looping and outside the natural range of [0, 1] in association data. By analyzing simulated data from microbiomes as benchmark examples, an average increase of 11% in prediction accuracy were obtained, demonstrating the effectiveness and robustness of iDIRECT. Compared to several existing approaches, iDIRECT have great advantages with 24-79% up to 2600% improvements in inferring direct relationships in terms of network topology, network size, accuracy and robustness. iDIRECT outperforms previous state-of-the-art approaches for inferring direct relationships. We applied this approach to previously reported bacterial networks that evaluate16S bacterial succession data from *Avena sp.* rhizosphere soil; this analysis shows a change in the composition of dominating modules in several complex networks. We also applied iDIRECT to real gene expression data from *Escherichia coli* and yeast to infer gene regulatory networks. Our results indicated that iDIRECT consistently improves the prediction by up to 12% as compared to the best performing approaches in DREAM5. The iDIRECT approach is clearly applicable to inferring direct relationships in plant root, bacterial, fungal, and mesofaunal association networks.

References

1. Alipanahi, B., & Frey, B. J. (2013). Network cleanup. Nature biotechnology, 31(8), 714.

2. Barzel, B., & Barabási, A. L. (2013). Network link prediction by global silencing of indirect correlations. Nature biotechnology, 31(8), 720-725.
3. Feizi, S., Marbach, D., Médard, M., & Kellis, M. (2013). Network deconvolution as a general method to distinguish direct dependencies in networks. Nature biotechnology, 31(8), 726.
4. Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. PLoS computational biology, 11(5), e1004226.

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 to UC Berkeley and the University of Oklahoma.

ARTENOLIS: Automated Reproducibility and Testing Environment for Licensed Software

Laurent Heirendt[1]& Sylvain Arreckx[1], Christophe Trefois[1], Yohan Jarosz[1], Maharshi Vyas[1], Venkata P. Satagopam[1], Reinhard Schneider[1], Ines Thiele[1], and **Ronan M.T. Fleming**[1]*(ronan.mt.fleming@gmail.com)

[1] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Project Goals: The aim of this project is to develop a continuous integration framework to automatically test changes to code, evaluate their impact on an open-source code base with licensed dependencies, and ensure the stability, integrity, and cross-platform compatibility of the COBRA Toolbox.

Automatically testing changes to code is an essential feature of continuous integration. For open-source code, without licensed dependencies, a variety of continuous integration services exist. The COnstraint-Based Reconstruction and Analysis (COBRA) Toolbox [1] is a suite of open-source code for computational modelling with dependencies on licensed software. It is distributed as open-source code, but as it is dependent on the licensed software MATLAB (The Mathworks, Inc.). A novel automated framework of continuous integration in a semi-licensed environment is required for the development of the COBRA Toolbox and related tools of the COBRA community. Here, we present ARTENOLIS [2], a freely accessible under http://artenolis.lcsb.uni.lu, a general-purpose infrastructure software application that implements continuous integration for open-source software with licensed dependencies. It uses a master-slave framework, tests code on multiple operating systems, and multiple versions of licensed software dependencies. ARTENOLIS ensures the stability, integrity, and cross-platform compatibility of code in the COBRA Toolbox and related tools.

References

[1] Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastian N. Mendoza, Anne Richelle, Almut Heinken, Hulda S. Haraldsdottir, Sarah M. Keating, Vanja Vlasov, Jacek Wachowiak, Stefania Magnusdottir, Chiam Yu Ng, German Preciat, Alise Zagare, Siu H.J. Chan, Maike K. Aurich, Catherine M. Clancy, Jennifer Modamio, John T. Sauls, Alberto Noronha, Aarash Bordbar, Benjamin Cousins, Diana C. El Assal, Luis V. Valcarcel, Inigo Apaolaza, Susan Ghaderi, Masoud Ahookhosh, Marouen Ben Guebila, Andrejs Kostromins, Hoai M. Le, Ding Ma, Yuekai Sun, Lin Wang, James T. Yurkovich, Miguel A. P. Oliveira, Phan T. Vuong, Lemmer P. El Assal, Scott Hinton, William A. Bryant, Francisco J. Aragón Artacho, Francisco J. Planes, Egils Stalidzans, Alejandro Maass, Santosh Vempala, Michael Hucka, Michael A. Saunders, Costas D. Maranas, Nathan E. Lewis, Thomas Sauter, Bernhard O. Palsson, Ines Thiele, and Ronan M. T. Fleming. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *Nature Protocols (accepted)*, 2018.

[2] Laurent Heirendt, Sylvain Arreckx, Christophe Trefois, Yohan Yarosz, Maharshi Vyas, Venkata P. Satagopam, Reinhard Schneider, Ines Thiele, and Ronan M. T. Fleming. ARTENOLIS: Automated Reproducibility and Testing Environment for Licensed Software. *arXiv:1712.05236 [cs]*, December 2017. arXiv: 1712.05236.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429, the National Centre of Excellence in Research, and the Luxembourg National Research Fund (FNR) ATTRACT program (FNR/A12/01) and OPEN program (FNR/O16/11402054) grants.

Creation and Analysis of Biochemical Constraint-based Models: the COBRA Toolbox v3.0

Laurent Heirendt[1] & Sylvain Arreckx[1], Thomas Pfau[2], Sebastián N. Mendoza[3,18], Anne Richelle[4], Almut Heinken[1], Hulda S. Haraldsdóttir[1], Jacek Wachowiak[1], Sarah M. Keating[5], Vanja Vlasov[1], Stefania Magnúsdóttir[1], Chiam Yu Ng[6], German Preciat[1], Alise Žagare[1], Siu H.J. Chan[6], Maike K. Aurich[1], Catherine M. Clancy[1], Jennifer Modamio[1], John T. Sauls[7], Alberto Noronha[1], Aarash Bordbar[8], Benjamin Cousins[9], Diana C. El Assal[1], Luis V. Valcarcel[10], Iñigo Apaolaza[10], Susan Ghaderi[1], Masoud Ahookhosh[1], Marouen Ben Guebila[1], Andrejs Kostromins[11], Nicolas Sompairac[22], Hoai M. Le[1], Ding Ma[12], Yuekai Sun[12], Lin Wang[6], James T. Yurkovich[13], Miguel A.P. Oliveira[1], Phan T. Vuong[1], Lemmer P. El Assal[1], Inna Kuperstein[22], Andrei Zinovyev[22], H. Scott Hinton[14], William A. Bryant[15], Francisco J. Aragón Artacho[16], Francisco J. Planes[10], Egils Stalidzans[11], Alejandro Maass[3,18], Santosh Vempala[9], Michael Hucka[17], Michael A. Saunders[12], Costas D. Maranas[6], Nathan E. Lewis[4,19], Thomas Sauter[2], Bernhard Ø. Palsson[13,21], Ines Thiele[1], **Ronan M.T. Fleming**[1]* (ronan.mt.fleming@gmail.com)

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, 6 avenue du Swing, Belvaux, L-4367, Luxembourg**;** [2]Life Sciences Research Unit, University of Luxembourg, 6 avenue du Swing, Belvaux, L-4367, Luxembourg; [3]Center for Genome Regulation (Fondap 15090007), University of Chile, Blanco Encalada 2085, Santiago, Chile; [4]Department of Pediatrics, University of California, San Diego, School of Medicine, La Jolla, CA 92093, USA; [5]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, CB10 1SD, United Kingdom; [6]Department of Chemical Engineering, The Pennsylvania State University, University, University Park, PA 16802, USA; [7]Department of Physics, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA; Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA, USA; [8]Sinopia Biosciences, San Diego, CA, USA; [9]School of Computer Science, Algorithms and Randomness Center, Georgia Institute of Technology, Atlanta, GA, USA; [10]Biomedical Engineering and Sciences Department, TECNUN, University of Navarra, Paseo de Manuel Lardizabal, 13, 20018, San Sebastian, Spain; [11]Institute of Microbiology and Biotechnology, University of Latvia, Jelgavas iela 1, Riga LV-1004, Latvia; [12]Department of Management Science and Engineering, Stanford University, Stanford CA 94305-4026, USA; [13]Bioengineering Department, University of California, San Diego, La Jolla, CA, USA; [14]Utah State University Research Foundation, 1695 North Research Park Way, North Logan, Utah 84341, USA; [15]Centre for Integrative Systems Biology and Bioinformatics, Department of Life Sciences, Imperial College London, London, United Kingdom; [16]Department of Mathematics, University of Alicante, Spain; [17]California Institute of Technology, Computing and Mathematical Sciences, MC 305-16, 1200 E. California Blvd., Pasadena, CA 91125, USA; [18]Mathomics, Center for Mathematical Modeling, University of Chile, Beauchef 851, 7th Floor, Santiago, Chile; [19]Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego, La Jolla, CA 92093, United States; [20]Latvian Biomedical Research and Study Centre, Ratsupites iela 1, Riga, LV1067, Latvia; [21]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800 Kgs. Lyngby, Denmark; [22]Institut Curie, PSL Research University, Mines Paris Tech, Inserm, U900, F-75005, Paris, France.

https://opencobra.github.io/cobratoolbox

Project Goals: The aim of this project is to expand the functionality of the COBRA Toolbox to cover new modelling methods, provide a comprehensive and consistent documentation as well as a new suite of interactive tutorials.

The COBRA Toolbox [1] is a comprehensive software suite of interoperable COnstraint-Based Reconstruction and Analysis (COBRA) methods to model, analyse, and predict a variety of metabolic phenotypes using genome-scale biochemical networks. By design, its functions can be flexibly combined to implement tailored COBRA protocols for any biochemical network in biology, biomedicine, and biotechnology. Version 3.0 includes new methods for quality controlled reconstruction, modelling, topological analysis, strain and experimental design, network visualisation as well as network integration of cheminformatic, metabolomic, transcriptomic, proteomic, and thermochemical data. New multi-lingual code integration also enables an expansion in application scope via high-precision, high-performance, and nonlinear numerical optimisation solvers for multi-scale, multi-cellular and reaction kinetic modelling, respectively.

A dedicated effort has been made to ensure that all functions in the COBRA Toolbox 3.0 are consistently documented. A novel and comprehensive suite of more than 35 tutorials (opencobra.github.io/cobratoolbox/latest/tutorials/) has been developed to enable beginner, intermediate, and advanced users to practise a wide variety of COBRA

methods. The COBRA Toolbox 3.0 protocol and its tutorials may be adapted for the generation and analysis of a constraint-based model in a wide variety of molecular systems biology scenarios.

The code and the documentation are freely released as part of the openCOBRA project on github.com/opencobra/cobratoolbox.

References

[1] Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, et al. (2018) Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. Nature Protocols (accepted) .

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429. This project was funded by the National Centre of Excellence in Research and has received funding from the European Union's HORIZON 2020 research and innovation programme under grant agreement No 668738. In addition, this work was supported by the Fonds National de la Recherche grant O16/11402054, and the Luxembourg National Research Fund (FNR) ATTRACT program grant (FNR/A12/01). Nathan E. Lewis has been supported by NIGMS (R35 GM119850) and the Novo Nordisk Foundation (NNF10CC1016517). Anne Richelle has been supported by the Lilly Innovation Fellows Award.

Mathematical Properties Of Stoichiometric Matrices

Susan Ghaderi[1], Hulda S. Haraldsdottir[1], Masoud Ahookhosh[1], Sylvain Arreckx[1],
and **Ronan M.T. Fleming**[1]*(ronan.mt.fleming@gmail.com)

[1] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Project URL: https://stanford.edu/group/SOL/multiscale/

Project Goals: Characterise the mathematical properties of stoichiometric matrices and decompose a biochemical network into a set of moiety subnetworks.

Abstract Characterising biochemical network structure in mathematical terms is a fundamental component of mathematical biology. It enables the inference of functional biochemical consequences from network structure with existing mathematical techniques and spurs the development of new mathematical techniques that exploit the peculiarities of biochemical network structure [1, 2]. The structure of a biochemical network is specified by reaction stoichiometry, that is, the relative quantities of molecular species produced and consumed in each chemical reaction of a network. The stoichiometry for the set of reactions in a network can then be compiled into a stoichiometric matrix, where every row corresponds to a molecular species and every column corresponds to a reaction. Herein, we summarise the main mathematical properties of stoichiometric matrices and emphasise those that distinguish them from arbitrary rectangular matrices.

References

[1] Hulda S. Haraldsdóttir and Ronan M. T. Fleming. Identification of conserved moieties in metabolic networks by graph theoretical analysis of atom transition networks. *PLOS Computational Biology*, 12(11):e1004999, November 2016.

[2] German A. Preciat Gonzalez, Lemmer R. P. El Assal, Alberto Noronha, Ines Thiele, Hulda S. Haraldsdóttir, and Ronan M. T. Fleming. Comparative evaluation of atom mapping algorithms for balanced metabolic reactions: application to Recon3d. *Journal of Cheminformatics*, 9:39, 2017.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429.

Numerical Properties Of Stoichiometric Matrices

Laurent Heirendt[1], Sylvain Arreckx[1], Almut Heinken[1], Dmitry Ravcheev[1],
Ines Thiele[1], and **Ronan M.T. Fleming**[1]*(ronan.mt.fleming@gmail.com)

[1] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Project URL: https://stanford.edu/group/SOL/multiscale/

Project Goals: This project aims at determining the numerical characteristics of stoichiometric matrices, identifying ill-scaled models requiring high precision computations, and drawing biologically inspired conclusions and determine possible avenues to accelerate Constraint-Based Reconstruction and Analysis computations for multiscale microbial communities.

Over the last 20 years, Constraint-Based Reconstruction and Analysis (COBRA) reconstructions gained considerably in complexity. A challenge, especially for huge scale networks and communities of organisms, is to accelerate related COBRA analyses using the COBRA Toolbox [1] or distributedFBA, part of COBRA.jl [2]. Knowledge of the numerical properties of the underlying metabolic network represented as the stoichiometric matrix is key to tackle this challenge. We present the numerical characteristics of the 773 microbial reconstructions [3], and link the properties to practical considerations in biology and scientific computing. It is shown that key numerical characteristics are correlated to inherent biological traits. Hints are provided to accelerate the analysis of large and huge-scale biochemical networks, especially for microbial community models.

References

[1] Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastian N. Mendoza, Anne Richelle, Almut Heinken, Hulda S. Haraldsdottir, Sarah M. Keating, Vanja Vlasov, Jacek Wachowiak, Stefania Magnusdottir, Chiam Yu Ng, German Preciat, Alise Zagare, Siu H.J. Chan, Maike K. Aurich, Catherine M. Clancy, Jennifer Modamio, John T. Sauls, Alberto Noronha, Aarash Bordbar, Benjamin Cousins, Diana C. El Assal, Luis V. Valcarcel, Inigo Apaolaza, Susan Ghaderi, Masoud Ahookhosh, Marouen Ben Guebila, Andrejs Kostromins, Hoai M. Le, Ding Ma, Yuekai Sun, Lin Wang, James T. Yurkovich, Miguel A. P. Oliveira, Phan T. Vuong, Lemmer P. El Assal, Scott Hinton, William A. Bryant, Francisco J. Aragón Artacho, Francisco J. Planes, Egils Stalidzans, Alejandro Maass, Santosh Vempala, Michael Hucka, Michael A. Saunders, Costas D. Maranas, Nathan E. Lewis, Thomas Sauter, Bernhard O. Palsson, Ines Thiele, and Ronan M. T. Fleming. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *Nature Protocols (accepted)*, 2018.

[2] Laurent Heirendt, Ines Thiele, and Ronan M. T. Fleming. DistributedFBA.jl: high-level, high-performance flux balance analysis in Julia. *Bioinformatics*, 33(9):1421–1423, May 2017.

[3] Stefanía Magnúsdóttir, Almut Heinken, Laura Kutt, Dmitry A. Ravcheev, Eugen Bauer, Alberto Noronha, Kacy Greenhalgh, Christian Jäger, Joanna Baginska, Paul Wilmes, Ronan M. T. Fleming, and Ines Thiele. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 35(1):81–89, January 2017.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429, the National Centre of Excellence in Research, the Luxembourg National Research Fund (FNR) ATTRACT program grant (FNR/A12/01), and the Luxembourg National Research Fund (FNR) CORE program grant (C16/BM/11332722).

Quantitative metabolic modeling at the Joint BioEnergy Institute

Hector Garcia Martin[1,2,3,*](hgmartin@lbl.gov), David Ando[1,2], Zak Costello[1,2,3], Tyler Backman[1,2], Jennifer Gin[1,2], Jason Brice[1,2], William Morrell[3,4], Garrett Birkel[2,3], Teresa Lopez[2,3], Mark Forrer[3,4], Nathan Hillson[1,2,3], Christopher Petzold[1,2]

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Lawrence Berkeley National Laboratory, Berkeley, CA; [3]DOE Agile BioFoundry; [4]Biological and Material Science, Sandia National Laboratories, Livermore, CA;

Project Goals: The Quantitative Metabolic Modeling (QMM) group is devoted to developing predictive models of metabolism that can leverage high-throughput 'omics data and direct metabolic engineering efforts.

The advent of synthetic biology as an effective tool to engineer biological cells has produced numerous beneficial applications, including the production of renewable biofuels[1]. However, effective design of biological systems is precluded by our inability to predict their behavior. New tools like CRISPR-enabled genetic editing, and DNA synthesis productivity that improves as fast as Moore's law, allows us to engineer changes faster than ever but the end result on cell behavior is usually unpredictable. At the same time, there is an exponentially increasing amount of functional genomics data available to the experimenter in order to phenotype the resulting bioengineered organism. Furthermore, the miniaturization of these techniques and the progressive automation of laboratory work through microfluidics chips promises a future where data analysis will be the bottleneck in biological research. Unfortunately, the availability of all this data does not translate into better predictive capabilities for biological systems: converting these data into actionable insights to achieve a given goal (e.g. higher bioproduct yields) is far from trivial or routine.

Here we show the variety of methods created in the QMM group to leverage 'omics data and guide metabolic engineering. We have used **machine learning** approaches to predict pathway dynamics directly from time-series data, without having to rely on Michaelis-Menten or other closed form of kinetics. We have also developed mechanistic models based on **^{13}C Metabolic Flux Analysis** to find mechanistic insights that have been shown to improve methyl ketone production by 110%. Finally we have also developed software tools to: guide **polyketide synthase (PKS) engineering** (clusterCAD [2]), **acquire and store data** systematically (EDD [3]), and perform **^{13}C MFA for genome-scale models** (jQMM library [4] [5,6]).

References

1. Chubukov V, Mukhopadhyay A, Petzold CJ, Keasling JD, Martín HG (2016) Synthetic and systems biology for microbial production of commodity chemicals. npj Syst Biol Appl 2: 16009. doi:10.1038/npjsba.2016.9.

2. Eng CH, Backman TWH, Bailey CB, Magnan C, García Martín H, et al. (2018) ClusterCAD: a computational platform for type I modular polyketide synthase design. Nucleic Acids Res 46: D509–D515. doi:10.1093/nar/gkx893.

3. Morrell WC, Birkel GW, Forrer M, Lopez T, Backman TWH, et al. (2017) The Experiment Data Depot: A Web-Based Software Tool for Biological Experimental Data Storage, Sharing, and Visualization. ACS Synth Biol 6: 2248–2259. doi:10.1021/acssynbio.7b00204.

4. Birkel GW, Ghosh A, Kumar VS, Weaver D, Ando D, et al. (2017) The JBEI quantitative metabolic modeling library (jQMM): a python library for modeling microbial metabolism. BMC Bioinformatics 18: 205. doi:10.1186/s12859-017-1615-y.

5. Ando D, Garcia Martin H (2018) Two-Scale 13C Metabolic Flux Analysis for Metabolic Engineering. Methods Mol Biol 1671: 333–352. doi:10.1007/978-1-4939-7295-1_21.

6. Martín HG, Kumar VS, Weaver D, Ghosh A, Chubukov V, et al. (2015) A Method to Constrain Genome-Scale Models with 13C Labeling Data. PLoS Comput Biol 11: e1004363. doi:10.1371/journal.pcbi.1004363.

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.

Machine learning techniques help accelerate enzyme engineering: a case study with glycoside hydrolases

Sanjan TP Gupta [1,2*] (sgupta78@wisc.edu), Evan Glasgow[2,3], Brian G Fox[2,3], Parmesh Ramanathan[4], 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]Department of Biochemistry, UW Madison, WI – 53706; [4]Department of Electrical and Computer Engineering, UW-Madison, WI – 53706

http://reedlab.che.wisc.edu

Project Goals:

The field of protein engineering primarily aims at improving the functional properties of a protein such as thermo-stability, binding affinity, and/or catalytic activity and has become a vital step in engineering industrial enzymes. This project aims to accelerate the protein engineering workflow using machine learning based approaches. In this regard, the broad goals of the project are:

i) To build an empirical model of protein fitness landscape to accurately predict the functional properties of an enzyme based on its protein sequence
ii) To intelligently navigate the above protein fitness landscape to design novel synthetic proteins with superior functional properties.

Abstract:

In this work, we have developed a machine learning (ML) based approach called 'MLProScape' to build an accurate model of the protein fitness landscape (ProScape) and then, use this model to design synthetic protein designs with superior functional properties. Unlike approaches using directed evolution, which requires experimentally screening millions of protein variants, our method requires substantially lesser number of variants (on the order of tens to hundreds) to be screened experimentally, owing to the power of statistical inference. MLProScape consists of three major steps - i) numerically encode a protein sequence using amino acid based physio-chemical properties as features, ii) build a machine learning based model using a subset of the most highly informative features, and iii) identify synthetic designs with improved characteristics.

To demonstrate our proposed workflow, we applied MLProScape to engineer the catalytic activity of glycoside hydrolases (enzymes that can break down cellulosic and hemi-cellulosic polysaccharides). We used experimentally measured specific activities for a diverse set of glycoside hydrolases to train the ML statistical models. The resulting elastic net regression based models have a high predictive power (with correlation coefficient and R^2 values as high as 0.896 and 0.714, respectively, between the predicted and experimentally measured specific activities

under 5-fold cross validation) surpassing previously published ML based enzyme engineering studies [1, 2]. Moreover, the use of position specific features helps us identify amino acid positions distal to the active site that might play a key role in modulating the activity level. Lastly, this ML based workflow is capable of modeling complex design criteria, such as optimizing the protein sequence for hydrolyzing multiple substrates simultaneously, as well as, to account for other desirable traits such as high stability and better *in vivo* expression levels.

References/Publications:

1. Fox RJ, Davis SC, Mundorff EC, Newman LM, Gavrilovic V, Ma SK, et al. Improving catalytic function by ProSAR-driven enzyme evolution. Nat Biotechnol. 2007; 25: 338–344. doi:10.1038/nbt1286
2. Srinivasulu Y, Wang J-R, Hsu K-T, Tsai M-J, Charoenkwan P, Huang W-L, et al. Characterizing informative sequence descriptors and predicting binding affinities of heterodimeric protein complexes. BMC Bioinformatics. 2015; 16: S14. doi:10.1186/1471-2105-16-S18-S14
3. Gupta STP, Glasgow E, Fox BG, Ramanathan P, Reed JL. MLProScape: Machine learning guided approach for engineering enzymes faster. (*In preparation*)

Funding statement:

This work was funded by the U.S. Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

Management and Analysis of ENIGMA Data using KBase

Pavel S. Novichkov[*,1] (PSNovichkov@lbl.gov), John-Marc Chandonia[1], Alexey E. Kazakov[1] and **Paul D. Adams**[1,2]

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

http://enigma.lbl.gov

Project Goals: ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) 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. The overarching goal of data management in ENIGMA is to enable integration of data from complex high-throughput field and laboratory studies towards multi-scale, predictive and systems level understanding of complex microbial community structure, function and evolution. To that end, we must ensure that all data are shared internally between ENIGMA teams, and also with scientists involved in other DOE programs in order to facilitate collaborative research and increase the overall pace of scientific discovery.

To date, we have deposited ENIGMA data into a number of publicly accessible locations, including NCBI and MG-RAST. We have also built specialized repositories that are accessible to the public (e.g., our fitness data browser available at fit.genomics.lbl.gov). To enable even more powerful data management and analysis in the future, we have established a strategic collaboration with DOE Systems Biology Knowledgebase (KBase), a fast growing computational infrastructure that provides a unique solution for sharing both data and computational modules/pipelines, conforms to the SC digital data management policy, and supports private data, data sharing and provenance.

We have performed project-wide surveys of ENIGMA data types and computational tools, and used this information to prioritize integration of our data and tools into KBase. Several large datasets, including 267 sequenced isolates from ENIGMA campaigns, as well as 16S surveys of 100 ENIGMA wells, have already been uploaded into KBase and shared for further computational analysis by ENIGMA groups. Such data have already been used for analysis of the genetic potential of isolates. We are continuing to monitor new data types and tools as they are introduced by ENIGMA team members, in order to update our priorities appropriately. We will prioritize all data types as well as key ENIGMA computational tools (e.g., the RB-TnSeq pipeline) for integration into KBase according to metrics such as scientific impact, costs, feasibility of integration.

The ENIGMA data management team has found that approximately 2/3 of ENIGMA data types would be well represented in KBase using “Generic” data objects that contain experimental measurements, which could be linked in KBase to non-Generic data types that represent biological or environmental objects (e.g., isolates and samples). We prototyped Generic data types, uploaders, graphing tools, search tools, and ontologies in KBase, and are currently collaborating with the KBase project to harden and deploy these technologies for use by ENIGMA team members as well as all other KBase users.

In addition to these "Generic Data" technologies, we are developing appropriate data models in KBase to describe other key ENIGMA data, such as environmental samples and sampling locations. We are also using KBase to organize and link to data that are stored elsewhere; e.g., MG-RAST, or data that are shared internally within the project using Google Drive. The data management team continues to facilitate work by ENIGMA scientists in uploading and sharing their data into KBase and other appropriate public repositories (e.g., GenBank). We are also providing assistance with project management (e.g., using JIRA when appropriate in order to track complex tasks that involve multiple laboratories).

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.

Combining multiple functional annotation tools increases completeness of metabolic annotation

Marc Griesemer[1], Jeffrey A. Kimbrel[1], Carol Zhou[2], Ali Navid[1], **Patrik D'haeseleer[1]** (dhaeseleer2@llnl.gov)

[1]Lawrence Livermore National Laboratory, Livermore CA, USA

https://bio-sfa.llnl.gov/

Project Goals: The LLNL Bioenergy SFA seeks to support sustainable and predictable bioenergy crop production through a community systems biology understanding of microbial consortia that are closely associated with bioenergy-relevant crops. We focus on host-microbial interactions in algal ponds and perennial grasses, with the goal of understanding and predicting the system-scale consequences of these interactions for biomass productivity and robustness, the balance of resources, and the functionality of surrounding microbial communities. Our approach integrates 'omics measurements with quantitative isotope tracing, characterization of metabolites and biophysical factors, genome-enabled metabolic modeling, and trait-based representations of complex multi-trophic biological communities, to characterize the microscale impacts of single cells on system scale processes.

Genome-scale metabolic modeling is a cornerstone of our genome and omics-enabled computational analysis of microbial communities and algal-bacterial interactions. However, genome-scale systems biology modeling efforts are invariably based on very incomplete functional annotations. Annotated genomes typically contain 30-50% of genes with little or no functional annotation (Hanson *et al.*, 2009), severely limiting our knowledge of the "parts lists" that the organisms have at their disposal. In metabolic modeling, these incomplete annotations are often sufficient to derive a reasonably complete model of the core metabolism consisting of well-studied (and thus well-annotated) metabolic pathways sufficient for growth in pure culture. Modern metabolic modeling efforts, however, are moving beyond studying core metabolic pathways in a single organism towards multi-species models, real-world communities and ecosystems.

In Flux Balance Analysis (FBA), the issue of missing metabolic annotations is dealt with by "gap filling" - the addition of a set of metabolic reactions beyond those that were derived directly from the genome annotation. However, gap filling typically requires the addition of several dozen reactions to allow production of biomass in simple defined nutrient media (Henry *et al.*, 2010); the reactions added by different gap filling algorithms may have little or no overlap with each other (Krumholz *et al.*, 2015); and even after gap filling, enough pathway holes remain to block on average one-third of the reactions in each model. Clearly, a more complete identification and annotation of metabolic reactions would be preferable to the addition of dozens of poorly supported reactions just to patch the holes in the network. Recent genome-scale modeling of *Clostridium beijerinckii* NCIMB 8052 (Milne *et al*, 2011) demonstrated that the total number of genes and reactions included in the final curated model could be almost doubled by incorporating multiple annotation tools.

Here, we present results on a comprehensive reannotation of 27 bacterial reference genomes from BioCyc, focusing on enzymes with EC numbers annotated by KEGG, RAST, EFICAz, and the BRENDA enzyme database. Our analysis shows that in comparison to metabolic annotation by one of the widely used metabolic modeling platform such as KEGG and RAST, annotation using multiple tools results in a drastically larger metabolic network reconstruction, adding on average 40% more EC numbers, and 37% more metabolic genes. These results are even more pronounced for bacterial species that are more distantly related to well-studied model organisms such as *E. coli* and *B. subtilis*.

Accurate metabolic models also rely on determination of substrate transport between the bacterium and its environment. Because of the computational difficulty in predicting substrates, transporter annotations have been rarely used in genome-scale metabolic modeling. Instead, most metabolic modeling methods simply assume that a transporter exists for the cellular import and export of any necessary metabolite. Better prediction tools such as TransportDB's Transporter Automatic Annotation Pipeline (TransAAP) now allow us to generate substrate predictions that are sufficiently detailed to be included in metabolic pathways, and that could give insights into growth or metabolite exchange phenotypes that are not readily apparent from the metabolic enzymes present in the genome. Our analysis across 27 reference genomes shows that transporter annotation using TransAAP adds 4 to 8 times more transporters with sufficiently detailed substrate annotation to be included in metabolic modeling, compared to annotation by KEGG or RAST alone, adding to each genome on average 212 to 233 additional transport reactions respectively.

The combination of multiple metabolic annotation tools allows us to achieve a significantly more complete genome annotation and metabolic network reconstruction, especially for non-model organisms, and for non-core pathways. We expect that these enhanced modeling abilities will be essential to study newly sequenced algal symbionts in complex real-world interactions. We are developing a computational pipeline to allow us to integrate metabolic annotations from IMG, KEGG, RAST, EFICAz and TransportDB, using SRI's Pathway Tools platform. Applying this approach to draft genomes of algal symbiotic bacteria, we demonstrate the effect of this more comprehensive annotation on the size and completeness of the metabolic reconstruction, and on our increased understanding of algal-bacterial metabolic interactions. This analysis, coupled with growth assays on minimal media with defined carbon and nitrogen substrates, has led to the identification of candidate substrates metabolized by the bacteria, and potential growth factors they may be exchanging with the algae.

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.

Synthetic Chemical Probes for Studying Lignin Deconstruction and Analysis of Biofuel Molecules Using nanostructure-initiator Mass Spectrometry (NIMS)

Kai Deng,[1,2]*(kdeng@lbl.gov) Jijiao Zeng,[1,2] Kenneth L. Sale,[1,2] Blake A. Simmons,[1,3] Paul D. Adams,[1,3] Anup K. Singh,[1,2] **Trent R. Northen**[1,3]

[1]Joint BioEnergy Institute, Emeryville, CA

[2]Sandia National Laboratories, Livermore, CA

[3]Lawrence Berkeley National Laboratory, Berkeley, CA

https://www.jbei.org

Project goal: To develop analytical tools to study enzyme function during lignin depolymerization and evaluate strains for biofuel production by utilizing bioconjugation chemistry and mass spectrometry readout. Adaptation of these tools in our automation platform will enable high throughput quantitation of enzyme or microbial strain libraries.

Lignocellulosic biomass is primarily composed of two polysaccharides (cellulose, hemicellulose) and a phenylpropanoid polymer (lignin). The complexity of the biomass structure requires the development and optimization of effective and affordable enzyme mixtures for depolymerization of these substrates. In addition, a robust and rapid method to screen biofuel-producing strains for desired products is needed to support development and optimization of strains with high titer, rate and yield. In order to meet these crucial challenges, we are developing mass spectrometry based assays that are high-throughput, small sample volume, good sensitivity and importantly, the integration of these screening technologies to automated workstations that facilitate the quantitative annotation of enzyme or microbial strain libraries. Central to our approach is to use synthetic organic chemistry to prepare chemical probes that enhance nanostructure-initiator mass spectrometry (NIMS) based analysis. This includes model substrates suitable for screening the activities of cellulases, hemicellulases and ligninases. This poster will focus on our progress on the development of model substrates to study ligninases and our approach to detecting alcohol products (fatty alcohol et. al) from biofuel production strains.

References

1. "Rapid kinetic characterization of glycosyl hydrolases based on oxime derivatization and nanostructure-initiator mass spectrometry (NIMS)" Deng, K.; Taksuka, 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.A.; Singh, A.K.; Fox, B.G.; Northen, T.R. *ACS Chemical Biology* 2014, *9*, 1470-1479.
2. "Development of a high throughput platform for screening glycoside hydrolases based on Oxime_NIMS" Deng, K.; Guenther, J.M.; Gao, J.; Bowen, B.P.; Tran, H.; Sale, K.L.; Simmons, B. A.; Adams, P. D.; Singh, A.K.; Fox, B.G.; Northen, T.R. *Frontiers in Bioengineering and Biotechnology* 2015, *3*, 153.

Funding statement:
This work was funded by the DOE Joint BioEnergy Institute (DOE BER Office of Science, through contract DE-AC02-05CH11231).

An Automated Microfluidic System for Gene Editing Processes

Kosuke Iwai[1,2]* (KosukeIwai@lbl.gov), David Ando,[1,3] Peter W. Kim,[1,2] Kai Deng,[1,2] Trent R. Northen,[1,4,6] Hector Garcia Martin,[1,3,4] Nathan J. Hillson,[1,3,6] Paul D. Adams,[1,4,5,7] and **Anup K. Singh**[1,2]

[1]Joint BioEnergy Institute, Emeryville, CA; [2]Biological and Material Science, Sandia National Laboratories, Livermore, CA; [3]Biological Systems and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA; [4]Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA; [5]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [6]DOE Joint Genome Institute, Walnut Creek, CA; and [7]University of California, Berkeley, CA

https://www.jbei.org/research/divisions/technology/microfluidic-assays/

Project Goals: The JBEI mission is to conduct basic and applied research to enable economically-viable conversion of lignocellulosic biomass into biofuels to provide the nation with clean, renewable transportation fuels identical to gasoline, diesel and jet fuel. The goal of this project, performed in the Microfluidic Assays group in the Technology Division at JBEI, is to deliver the robust and easy-to-use microfluidic platforms to automate the genetic engineering processes for advancing synthetic biology applications including biofuels development.

In recent years, synthetic biology has dramatically grown and became significantly important for both of scientific researches and industrial applications such as biofuel and pharmaceutical applications. However, multiple genetic engineering steps required for synthetic biology are often time-consuming and labor-intensive with repetitive pipetting and plating. Therefore, automated and efficient processes to perform molecular biology assays have been long desired. Microfluidic assays and devices with aqueous droplets (microliter to picoliter in volume) suspended in oil phase as compartmentalized reaction chambers have attracted a significant attention for performing biochemical reactions and analysis as they provide drastic improvements over their macroscale counterparts with various benefits such as faster reaction time, less volume of reagent consumption required, better control of experimental environment, and higher throughput with multiplexed processes.

We are involved in developing innovative microfluidic assays and integrated devices for many biofuel research applications including enzyme screening, enzyme evolution and synthetic biology. Our hybrid microfluidic platforms utilize continuous-flow (analog) microfluidics that manipulate the droplets by controlling the hydrodynamic force, and digital microfluidics (DMF) that utilize surface tension from electrowetting on dielectric with arrayed electrodes. The systems can handle large numbers of droplets at once as well as actively manipulate target 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, hydrodynamic

capture and array of the droplets, electric-field driven merge and split of the droplets to add specific amount and concentrations of various reagents, on-chip electroporation, and incubation process with localized temperature control. Specifically, multiple pairs of electrodes are designed and placed at each chambers to apply voltages to the arrayed droplets for on-chip electroporation. This configuration allows us to modify and program the electroporation conditions at each droplet for multiplexed DNA transformation processes, and it also enables us to easily scale-up the numbers of reactions for high-throughput transformation processes simply by designing the same structures in an array. In addition, we integrate optical fibers in the microchannels to add on-chip capability for fluorescence-based detection of encapsulated cells and enzymatic activities in the discrete droplets, and for triggering sorting of droplets. We utilize our microfluidic methodologies for automating CRISPR/Cas9 based gene editing processes such as recently established CRMAGE for *Escherichia coli* or cloning-free tool kit for *Saccharomyces cerevisiae*.

Unlike conventional microtiter plate based reactions, our analog-digital microfluidic platforms with on-chip electroporation and fluorescence detection allow completely automated genetic engineering steps using 10-100-fold lower amounts of reagents and can be useful for application requiring high throughput screening and reactions.

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. A. Reider Apel, L. d'Espaux, M. Wehrs, D. Sachs, R. A. Li, G. J. Tong, M. Garber, O. Nnadi, W. Zhuang, N. J. Hillson, J. D. Keasling, and A. Mukhopadhyay, "A Cas9-Based Toolkit to Program Gene Expression in *Saccharomyces Cerevisiae*", *Nucleic Acids Res.*, 2017, **45**, 496-508.

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.

A general, high-throughput platform for molecular prototyping of microbial cell factories via optically guided mass spectrometry

Tong Si[*,1,2], Pu Xue[1,2], **Huimin Zhao**[1,2], **Jonathan V. Sweedler**[1,3]

[1]Center for Advanced Bioenergy and Bioproducts (CABBI), University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA
[2]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA
[3]Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA
[3]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, USA

Project Goal:

To develop a general platform for high-throughput screening of microbial cell factories at the molecular level using advanced mass spectrometry approaches.

Abstract:

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) mass spectrometry (MS) imaging has been widely used for molecular analysis of biological samples, but manual procedures of sample preparation are rate-limiting. Recently, we developed automatic, optically guided MALDI-ToF MS analysis as a label-free method to enable high-throughput engineering of microbial cell factories[1]. Coupled with machine vision, automatic MALDI-ToF MS acquisition from randomly distributed colonies simplifies procedures for chemical phenotyping without liquid handling. This method was successfully applied to engineer and screen mutant strains that produce new antibiotic analogs or mixtures of glycolipid congeners with desirable composition. For both cases, large populations of colonies were rapidly surveyed at the molecular level, providing information-rich insights not easily obtained with traditional screening assays. Computational algorithms were also developed to process and visualize the resulting mass spectral data sets, whereby colorimetric readouts were overlaid with optical images to facilitate mutant recovery. After MALDI-ToF MS screening, follow-up analyses using high-resolution MS and tandem MS were readily performed on the same sample target. In our CABBI project, this platform will be applied to screen large libraries of engineered yeasts to synthesize lipid-derived products of custom chemical profiles with high efficiency. For preliminary results, when mutations were introduced into fatty acid synthetase genes, altered membrane lipid compositions in resulting yeast strains were robustly detected using optically guided MALDI MS. Utilizing standard microbiological techniques with routine microscopy and MALDI-ToF MS instruments, this simple yet effective workflow is applicable for a wide range of screening campaigns improving microbial cell factories.

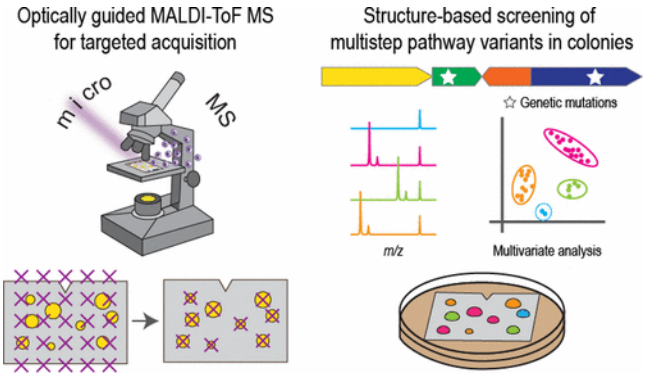


Figure 1. Schematic illustration of optically guided MALDI-ToF MS for automatic chemical profiling of microbial colonies and its application in biosynthetic pathway engineering.

References

1. T. Si, et al. "Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry." *Journal of the American Chemical Society*, 2017, 139(36), 12466-12473

Funding statement

We gratefully acknowledge financial support from the NIH GM077596 (H.Z.) and NIH AI113219 and NSF CHE 16-067915 (J.V.S.). In addition, 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-SCxxxxxx (H.Z. and J.V.S. through CABBI).

Identification and Optimization of CRISPR Systems for Genome Editing in *Pseudomonas putida* and *Clostridium thermocellum*

Jacob Fenster[1], Audrey Watson[1], Julie Walker[1], Emily Freed[1], Sandra Notonier[2], George Peabody[3], Anthony Lanahan[4], Gregg Beckham[2], Adam Guss[3], Dan Olson[4], Jeff Cameron[1], Ryan Gill[1], Lee Lynd[4], Carrie Eckert[1,2*] (carrie.eckert@nrel.gov), and **Gerald A. Tuskan**[3]

[1]Renewable and Sustainable Energy Institute, University of Colorado, Boulder CO; [2]National Renewable Energy Laboratory, Golden, CO; [3]Oak Ridge National Laboratory, Oak Ridge, TN; [4]Dartmouth College, Hanover, NH.

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 will address strategic barriers to the current bioeconomy in the areas of: 1) high-yielding, robust feedstocks, 2) lower capital and processing costs via consolidated bioprocessing (CBP) to specialty biofuels, and 3) methods to create valuable byproducts from the lignin. CBI will identify and utilize key plant genes for growth, composition and sustainability phenotypes as a means of achieving lower feedstock costs, focusing on poplar and switchgrass. We will convert these feedstocks to specialty biofuels (C4 alcohols and C6 esters) using CBP at high rates, titers and yield in combination with cotreatment or pretreatment. And CBI will maximize product value by *in planta* modifications and biological funneling of lignin to value-added chemicals.

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. Although understanding of complex biological systems and "design rules" for metabolic engineering of model organisms have progressed significantly, we are still unable to effectively engineer many desired properties necessary for a sustainable and efficient bioprocessing platform in non-model microbes and plant feedstocks. Therefore, it is essential that we expand genetic engineering capabilities in novel systems with desired traits for bioprocessing to accelerate advancement of the U.S. bioeconomy.

Advances in DNA synthesis and powerful genome editing tools such as CRISPR (**C**lustered **R**egularly-**I**nterspaced **S**hort **P**alindromic **R**epeats) have led to a rapid expansion of genome engineering in model and now increasingly in non-model organisms. To enable these and other cutting-edge genome editing technologies in CBI microbes, our first goal is to develop efficient CRISPR-mediated genome editing systems in both *Pseudomonas putida* and *Clostridium thermocellum*.

Genome editing utilizing Cas9 from *Streptococcus pyogenes* (a Type II CRISPR/Cas system) has recently been demonstrated in *P. putida* KT2440 (1). The editing efficiency is low, so we are also investigating a number of strategies to improve CRISPR-based editing and recombination efficiency.

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 *C. thermocellum* (55-60^{o}C). Diverse CRISPR systems exist amongst Archaea and Bacteria, and the genome of *C. thermocellum* DSM1313 encodes two Type III and one Type I CRISPR/Cas systems. Unlike Type II systems (e.g., Cas9), which require only a single Cas protein for targeting (via the guide RNA, or gRNA),

Type I and III systems require a multi-subunit Cas complex. Since Type III systems generally target RNA instead of DNA, we are characterizing the native Type I system as a tool for genome editing in *C. thermocellum.* In addition, two thermophilic Cas9 systems have recently been characterized (2,3) and are currently being evaluated for genome editing in *C. thermocellum*.

We expect that CRISPR systems can be harnessed to not only accelerate precise genome editing in CBI microbes, but also to create targeted, *trans*-acting regulatory systems as has been demonstrated in *E. coli* and cyanobacteria (4,5). Therefore, we are currently evaluating a CRISPR interference (CRISPRi) system in *P. putida* to allow for targeted knockdown of genes of interest to begin to determine gene-to-trait attributes towards desired phenotypes. Preliminary data on these evaluations will be presented.

Ultimately, we aim to utilize these CRISPR/Cas systems for a rapid, HTP method for phenotype-to-genotype discovery in *P. putida* and *C. thermocellum*. **CR**ISPR **E**n**A**bled **T**rackable Genome **E**ngineering (CREATE) couples multiplexed, high efficiency CRISPR-based genome editing with massively parallel oligomer synthesis. This method enables generation of hundreds of thousands of designer modifications (via pooled oligo microarrays) and, by means of a trans-acting barcode, the simultaneous sequence-to-activity mapping of all of such modifications (6). In conjunction with a selectable and/or screenable phenotype (e.g., improved growth, substrate tolerance, fluorescence, etc.), this technique can be applied to determine genotype-to-phenotype relationships for: 1) rational protein engineering, 2) complete residue substitution libraries, 3) pathway optimization and 4) discovery of new gene functions by genome-wide targeting strategies. This method will expand and accelerate the canonical the "Design, Build, Test" cycle for gene-to-trait discovery in support of CBI research needs.

References

1. Aparicio T, de Lorenzo V, Martínez-García E. CRISPR/Cas9-Based Counterselection Boosts Recombineering Efficiency in Pseudomonas putida. Biotechnol J. 2017 Oct 23
2. Mougiakos I, Mohanraju P, Bosma EF, Vrouwe V, Finger Bou M, Naduthodi MIS, et al. Characterizing a thermostable Cas9 for bacterial genome editing and silencing. BioRxiv. 2017 Aug 18.
3. Harrington L, Paez-Espino D, Chen JS, Ma E, Staahl BT, Kyrpides NC, et al. A thermostable Cas9 with increased lifetime in human plasma. BioRxiv. 2017 May 16.
4. Luo ML, Leenay RT, Beisel CL. Current and future prospects for CRISPR-based tools in bacteria. Biotechnol Bioeng. 2016 113(5):930–943.
5. Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemann MB, Pfleger BF. CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium Synechococcus sp. strain PCC 7002. Metab Eng. 2016 38:170–179.
6. 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 35(1):48–55.

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.

Optimization of Genome Engineering in the Model Pennate Diatom *Phaeodactylum tricornutum* using CRISPR-CAS9

Mark Moosburner[1,2], Andrew E. Allen[1,2*] (aallen@jcvi.org)

[1]Scripps Institution of Oceanography, UC San Diego, La Jolla, CA; [2]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA

Genetic engineering can change an organism into a biological factory for producing industrially valuable products. CRISPR-CAS9, known as molecular scissors, alters an organisms' genome by a CAS9-based riboendonuclease complex that is guided to a gene target by a programmable "guide" RNA called an sgRNA. CAS9 engineering of marine diatoms have recently been demonstrated and could potentially enhance the synthesis of naturally occurring high-valued metabolites such as fatty acids, pigments, and lipids for biodiesel production. In the Allen laboratory, we've been optimizing CRISPR-CAS9 genome engineering for multiple applications in the model marine diatom *Phaeodactylum tricornutum.* Primarily, CRISPR-CAS9 has been developed to be introduced to *Phaeodactylum* by bacterial-mediated conjugation DNA transformation. Targeting the *urease* gene resulted in 55% targeting efficiency compared to 17% targeting efficiency observed when using biolistics-based CRISPR-CAS9 delivery. However, a growth assay screen for *urease*, where mutants die when grown with urea, was crucial for screening transformed colonies prior to evaluating the targeted genomic loci. Nevertheless, a majority of *Phaeodactylum* genes do not have an associated phenotype to screen for. Also, conjugation here resulted in a loss of the CAS9 coding sequence in 90% of transformed *Phaeodactylum* which severely impacts the rate of producing false positive cell lines that do not contain a functional CAS9 enzyme. Recent efforts towards improving the bacterial-mediated conjugation methodology for introducing the CRISPR-CAS9 system have been made to boost CAS9 targeting efficiency. To ensure high-CAS9 expression in all conjugated *Phaeodactylum*, a cell line was produced harboring a CAS9 expression cassette in the genome. In order to continually track CAS9 expression, a fused eYFP was expressed with CAS9. Flow cytometry was used to quantify CAS9-eYFP expression for multiple cell lines. Two commonly used promoters, FcpB and pNR, were used to constitutively express CAS9 and induce CAS9 expression, respectively. A CAS9 cell line was chosen for high eYFP abundance within the culture and comparable growth to a wild-type *Phaeodactylum*. Two out of three lines could grow comparably to wild-type and expression of CAS9 widely varied between all picked lines, which indicates that CAS9 delivery via biolistics can impact CAS9 expression and cell physiology in unpredictable ways. Using a CAS9-eYFP *Phaeodactylum* cell line, sgRNAs will be introduced on a replicating episome. An episome has been constructed conferring resistance to the antibiotic *nourseothricin* and containing an sgRNA expression cassette. The episome has also been designed to allow the expression of multiple sgRNAs by harnessing a hierarchal Golden-Gate cloning scheme. The combination of a CAS9 cell line and bacterial conjugation delivery of sgRNAs will be used for the targeting of individual gene and also multiplexed targeting of multiple genes simultaneously. Lastly, the sgRNA episome has also been optimized for the introduction of a library of sgRNAs. Genome-wide mutagenesis of all coding gene or a "cherry-picked" set of genes will be performed using the sgRNA episome and a CAS9 *Phaeodactylum* cell line. Targeted mutagenesis of a massive number of genes in parallel will greatly accelerated

the identification and characterization of genes involved in the production of high-valued metabolites in *Phaeodactylum tricornutum.*

Funding Statement

Research supported by United States Department of Energy Genomics Science program grants DE-SC0008593 and DE-SC0018344.

Development of Resources and Tools to Improve Oil Content and Quality in Pennycress

Fan Yang[1], Tyler Swanson[2], Seyed H. Mousavi Avval[1], Luis A. Huezo Sanchez[1], Erich Grotewold[3], Ajay Shah[1], and **Ana Paula Alonso[2]*** (alonso.19@osu.edu),

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

Project Goals: The main objective of this research is to apply a biodesign strategy for improving oil content in a promising alternative source of jet-fuel, pennycress. To advance towards this goal, we are: 1) Investigating pennycress natural variation to identify candidate genes and biomarkers associated with oil accumulation and fatty acid composition; 2) Identifying targets to improve oil content and composition, and 3) Establishing metabolic engineering targets and develop community resources.

Bioenergy crops, which have potential for jet fuel production and do not compete with food crops, are urgently needed. Our strategy to address this fundamental challenge is to develop pennycress (*Thlaspi arvens*e), a member of the Brassicaceae, as a bioenergy crop, taking advantage of its ability to produce seed oil that is ideally suited as a renewable source of biodiesel and aviation fuel[1]. Moreover, pennycress performs well on marginal land, has a short maturity time and grows off-season, serving as a winter cover crop, and complementing the production of commodity food crops. Essential pennycress molecular and genetic resources can be rapidly developed by leveraging the fully sequenced genome and research tools of its close relative *Arabidopsis thaliana*. Furthermore, a draft genome[2] for *Thlaspi arvense*, as well as transcriptome[3] and metabolome[4] information have been made recently available. However, for this plant to become an economically viable and sustainable source of jet fuel, molecular and genetic resources need to be developed, and integrated with multi-variable techno-economical analyses to guide strategies for increase oil production through breeding and/or genetic manipulation. These are the gaps that this project intends to fill.

Specifically, we are: ***1) Investigating pennycress natural variation to identify candidate genes and biomarkers associated with oil accumulation and fatty acid composition.*** For this purpose, we are currently determining variation in genome-wide gene expression (derived from RNA-Seq) and intracellular metabolites (derived from metabolomics) in seeds from pennycress natural accessions. ***2) Identifying targets to improve oil content and composition.*** Towards this goal, we developed new mass spectrometric methods to follow the ^{13}C-labeling in intracellular compounds[5], and we are now generating a flux map of carbon partitioning in developing pennycress embryos. We will overlay metabolic maps with levels of transcripts and intracellular compounds to identify metabolic bottlenecks in oil accumulation. Finally, we will use ^{13}C-based metabolic flux analysis to validate bottlenecks in two accessions with contrasting oil contents. ***3) Establishing metabolic engineering targets and develop community resources.*** For this purpose, we are analyzing techno-economics of pennycress based agronomic and supply systems that will provide targets for future efforts aimed at increasing seed oil production through rational metabolic

engineering and breeding. We are working with the Arabidopsis Biological Resource Center to develop a public seed collection of pennycress mutants and transgenic lines, facilitating community synergy and accelerating research towards the established production goals.

Taken together, the knowledge and resources generated through this interdisciplinary project will facilitate rational breeding and metabolic engineering of pennycress and related alternative bioenergy crops.

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 Bioenerg.* **55**, 87-100 (2013).
2. Dorn, K.M., Fankhauser, J.D., Wyse, D.L. & Marks, M.D. A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Res* **22**, 121-31 (2015).
3. Dorn, K.M., Fankhauser, J.D., Wyse, D.L. & Marks, M.D. De novo assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant J.* **75**, 1028-1038 (2013).
4. Tsogtbaatar, E., Cocuron, J.C., Sonera, M.C. & Alonso, A.P. Metabolite fingerprinting of pennycress (*Thlaspi arvense* L.) embryos to assess active pathways during oil synthesis. *J Exp Bot.* **66**, 4267-4277 (2015).
5. Cocuron, J.C., Tsogtbaatar, E. & Alonso, A.P. High-throughput quantification of the levels and labeling abundance of free amino acids by liquid chromatography tandem mass spectrometry. *J Chromatogr A.* **1490**, 148-155 (2017).

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0016490.

Imaging Native Structure of Plant Cell Wall Cellulose Microfibril

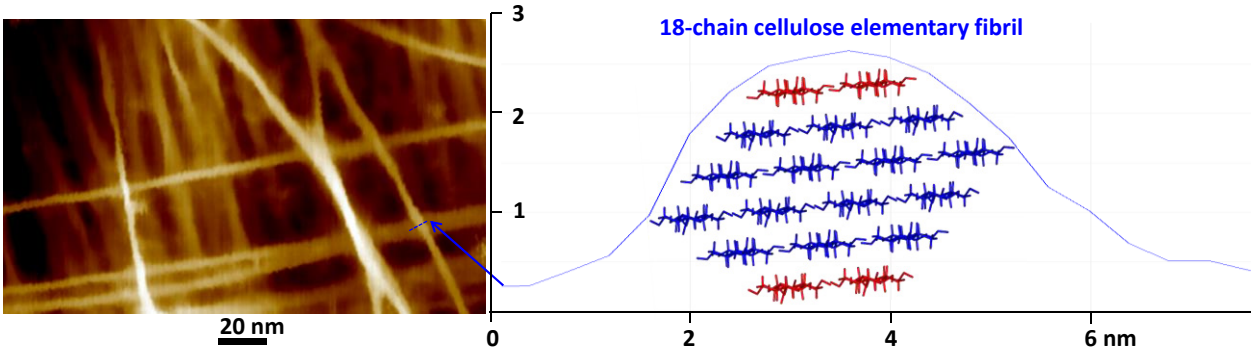
Wei Shen[1,2*] (weishen@msu.edu), Bo Song[2], and **Shi-You Ding**[1,2]

[1]Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, 48824 USA;
[2]Department of Plant Biology, Michigan State University, East Lansing, MI, 48824 USA

https://www.glbrc.org/

Project Goals: To elucidate the molecular mechanisms that confer resistance to catalytic deconstruction of biomass to sugars by means of developing emerging microscopic techniques in GLBRC to enable real-time imaging capabilities at the cellular and molecular scales, and thereby to provide critical knowledge needed to improve the processes to produce biofuels and bio-products.

Plants provide essential materials of our daily life, such as wood, fibers, and recently feedstock for producing biofuels and biomaterials. Cellulose is central to plant life and the global carbon cycle. Plants use rigid cellulose microfibrils together with non-cellulosic matrix polymers to build cell walls. Cellulose is known to comprise linear β(1,4)-glucan chains that are organized through hydrogen-bonding networks and van der Waals forces, however, in plants the number of chains and how they arrange in the microfibril remains elusive. This study uses atomic force microscopy (AFM) to directly image under water the cell walls from living maize (*Zea mays*) plants, which allows us to precisely measure the near-native structure of the microfibril at the sub-nanometer scale. We find the microfibrils vary in size and arrangement in different wall types or at different developmental stages, but all microfibrils are comprised of a fundamental building block, namely the cellulose elementary fibril (CEF) that has a defined cross-sectional area of 6 nm^2. Based on known crystal structure of cellulose, this CEF is only consistent with an 18 - chain model. We further confirm that this CEF structure widely exists in the cell walls of other monocot and eudicot plant species, suggesting it is likely a result of the conserved cellulose biosynthesis machinery in land plants. Our results provide long-awaited direct measurement of native cellulose structure, which provide new insights into cell wall biosynthesis and plant development, as well as the rational design of cellulose-based biomaterials.



The Great Lakes Bioenergy Research Center is a U.S. Department of Energy Bioenergy Center supported by the DOE Basic Energy Research Office of Science (DE-FC02 -07ER64494).

Accurate, Sensitive, and Precise Multiplexed Proteomics using the Complement Reporter Ion Cluster

Matthew Sonnett,[1] Eyan Yeung,[1] **Martin Wühr**[1] (wuhr@princeton.edu)

[1]Lewis-Sigler Institute of Integrative Genomics & Department of Molecular Biology; Princeton University, Princeton, NJ, USA

https://scholar.princeton.edu/wuehr/home

Quantitative analysis of proteomes across multiple time points, organelles, and perturbations is essential for understanding both fundamental biology and disease states. The development of isobaric tags (e.g. TMT) have enabled the simultaneous measurement of peptide abundances across several different conditions. These multiplexed approaches are promising in principle because of advantages in throughput and measurement quality. However, in practice existing multiplexing approaches suffer from key limitations. In its simple implementation (TMT-MS2), measurements are distorted by chemical noise leading to poor measurement accuracy. The current state-of-the-art (TMT-MS3) addresses this, but requires specialized quadrupole-iontrap-Orbitrap instrumentation. The complement reporter ion approach (TMTc) produces high accuracy measurements and is compatible with many more instruments, like quadrupole-Orbitraps. However, the required deconvolution of the TMTc cluster leads to poor measurement precision. Here, we introduce TMTc+, which adds the modeling of the MS2-isolation step into the deconvolution algorithm. The resulting measurements are comparable in precision to TMT-MS3/MS2. The improved duty cycle, and lower filtering requirements make TMTc+ more sensitive than TMT-MS3 and comparable with TMT-MS2. At the same time, unlike TMT-MS2, TMTc+ is exquisitely able to distinguish signal from chemical noise even outperforming TMT-MS3. Lastly, we compare TMTc+ to quantitative label-free proteomics of total HeLa lysate and find that TMTc+ quantifies 8.0k versus 3.8k proteins in a 5-plex sample. At the same time the median coefficient of variation improves from 13% to 6%. Thus, TMTc+ advances quantitative proteomics by enabling accurate, sensitive, and precise multiplexed experiments on more commonly used instruments.

Designing Self-Assembling Protein Materials for Imaging and Energy Applications

Yuxi Liu,[1,2] Kevin Cannon[1,3]* (kacannon@ucla.edu), Joshua Laniado,[2] Scott McConnell,[3] Robert T. Clubb,[1,2,3], Shane Gonen[4], Tamir Gonen[4], James Evans[5], and **Todd O. Yeates**[1,2,3]

[1] UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA; [2]UCLA Molecular Biology Institute, Los Angeles, CA; [3]UCLA Department of Chemistry and Biochemistry, Los Angeles, CA; [4]HHMI, Janelia Research Campus, Ashburg, VA; [5]PNNL Richland, WA

http://yeateslab.mbi.ucla.edu/?page_id=58

Project Goals: The project goals are to develop and test methods for designing novel self-assembling protein materials for applications that advance Department of Energy interests in high resolution imaging methods, green chemistry and metabolic engineering. Specific current goals include the design of a modular scaffold that will enable cryo-EM methods for smaller proteins, and the design of protein cages for the attachment of multiple sequentially-acting enzymes involved in cellulose degradation.

Abstract – Recent advances in designing proteins to self-assemble into specific architectures are opening up numerous exciting technology applications (1). One of these is in cryo-electron microscopy imaging. Recent advances in EM hardware and software have made it possible to image large protein and nucleic acid assemblies at near-atomic resolution detail. Notably, large and symmetric structures are the most easily studied, whereas smaller structures (e.g. below about 50 kD) are generally impossible owing to low contrast and low signal making particle identification and orientation unreliable. A long-standing goal has been to create scaffolding systems that would enable the cryo-EM imaging of smaller proteins, like those found throughout the cell, possible. An ideal scaffold would be large and symmetric and would hold many copies of a protein target to be imaged in rigid orientations in order to facilitate high resolution imaging. In this project we have demonstrated the first designed scaffold that meets those design requirements. A designed protein cage comprises the core. The core is then fused by a semi-rigid continuous alpha helical connection to a DARPin protein, which serves as a modular adaptor for selectable binding to other diverse target proteins. So far, we have imaged the scaffold itself (without target proteins attached) and have shown that the DARPin is attached rigidly enough to resolve this small 18 kD protein at resolutions from 3.5 to

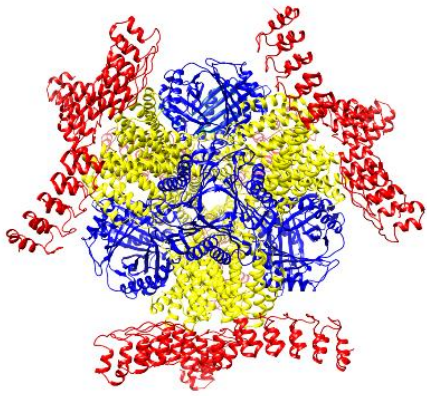


Figure 1. EM structure of a designed modular protein scaffold for cryo-EM imaging. The designed cage subunits are in blue and yellow. The DARPin adaptor domain (red) is fused to one of the two protein cage subunits by a rigid alpha helical linker. Modifying the loop sequences of the DARPin adaptor allow the rigid binding and presentation of diverse protein targets for imaging.

5 Å (2) (Fig. 1). Imaging of attached target proteins is underway. In a second set of distinct applications we are adapting designed protein cages for the display of multiple sequential enzymes. We are applying this idea to the problem of cellulose degradation. New methods based on sortase recognition sequences are being developed for the facile attachment of enzymes to designed cages (3). We have shown that a protein cage in which one component contains a C-terminal tag for recognition by the sortase enzyme SrtA can be covalently linked to the cellulase enzymes Cel8A and Cel48S. Preliminary data suggests that the cages remain intact after the modification and that both of these cellulase enzymes can be displayed on the cage surface simultaneously (Fig. 2). It is hypothesized that modifying designed protein cages in this way can increase pathway flux for the degradation of cellulose, a notoriously slow process, by forcing the enzymes that act sequentially in this pathway into close proximity. Tests of enzymatic efficiency in cellulose degradation are underway.

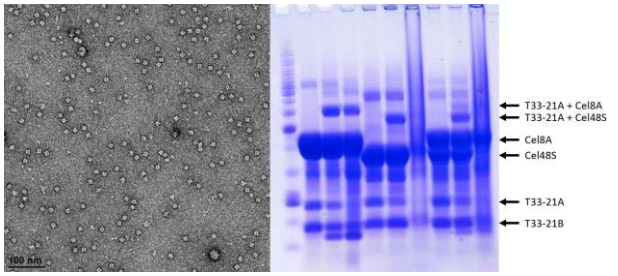


Figure 2. Left: Representative EM micrograph of T33-21, a 2-component tetrahedral cage, after attaching Cel8A via SrtA. Right: SDS-PAGE showing attachment of Cel8A and Cel48S, both separately and simultaneously, to the A component of T33-21.

References

1. Yeates TO, Liu Y, Laniado J. The design of symmetric protein nanomaterials comes of age in theory and practice. *Curr Opin Struct Biol*. 2016 Aug;**39**:134-143. doi: 10.1016/j.sbi.2016.07.003.
2. Liu Y, Gonen S, Gonen T, Yeates TO. Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. https://www.biorxiv.org/content/early/2017/11/01/212233
3. Spirig T, Weiner EM, Clubb RT. Sortase enzymes in Gram-positive bacteria. *Molecular Microbiology*, **82**: 1044-1059, 2011. doi: 10.1111/j.1365-2958.2011.07887.x

This work is funded by award E-FC02-2ER63421 from the BER program of the DOE Office of Science.

Protein Function Discovery and Sequence-Function Understanding of Stress Tolerance in the Green Lineage

Ian K. Blaby[1], Agnidipta Ghosh[2], Kevin Keck[1], Kyounghoon Kim[1], Gerry Guo[1,4], Miriam Pasquini[1], Shinjae Yoo[3], Steven C. Almo[2], Sean McSweeney[1,4], Qun Liu[1,4], and **Crysten E. Blaby-Haas*[1]**
(cblaby@bnl.gov)

[1]Biology Department, Brookhaven National Laboratory, Upton, NY
[2]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY
[3]Computational Science Center, Brookhaven National Laboratory, Upton, NY
[4]National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY

Project Goals: The Quantitative Plant Science Initiative (QPSI) seeks to provide fundamental understanding of genomics-based protein function to accelerate improvement of present and future bioenergy/bioproduct crops. The primary goals of this interdisciplinary capability are: i) to achieve a systems-level understanding of how plants across the green lineage have evolved to acclimate to suboptimal environmental conditions and ii) to provide the sequence-to-function-to-phenotype knowledge required to translate experimentally validated functions across species for enabling the design and redesign of target plant processes.

There are multiple challenges to implementing an economically viable and sustainable bioeconomy. In particular, hurdles still exist for plant crops as renewable resources of bioenergy and bioproducts, such as development of crops that can thrive in marginal soils and maintain performance in diverse and fluctuating environments. Post-genomic resources have accelerated our ability to achieve systems-wide information of how plants respond and adapt to biotic and abiotic stresses. Capitalizing on this torrent of genome-based data, the QPSI capability is developing approaches that combine the scalability of functional genomics with protein function characterization to generate the foundational knowledge needed for the design and redesign of stress tolerance. Trans-systems functional genomics analyses that utilize 'omic-derived data, cross-kingdom comparative genomics, and protein family evolution are being used to discover novel niche-specific and lineage-wide processes as potential targets for crop improvement. Protein function predictions are being used to guide characterization via genetic, biochemical and structural biology techniques. We are also developing approaches that combine target-specific experimentation with computational platforms for the accurate propagation of experimentally grounded functional characterization across sequence space.

This work is supported by the Office of Biological and Environmental Research in the DOE Office of Science

DEVELOPMENT OF ANALYTICAL METHODS TO INVESTIGATE ANTAGONISTIC INTERACTIONS BETWEEN BACTERIA AND FUNGI

Hang Ngoc Nguyen[1], Geoffrey House[2], Patrick Chain[2], Pilar Junier[3], Jamey Young[4], Armand Dichosa[2], Jean Challacombe[2] and **Debora Frigi Rodrigues**[1]* (dfrigirodrigues@uh.edu)

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

Project Goals: Interactions between bacteria and fungi are important determinants of ecosystem function, yet little is known about these interactions or how they operate. This is a critical knowledge gap as these interactions are important in addressing multiple DOE priorities including developing renewable energy sources, understanding the possible effects of Earth system change, and understanding how these interactions may help overcome energy and environmental challenges. Here we outline a range of research questions that we are beginning to address through a new SFA in order to better understand the diversity and function of these bacterial:fungal interactions. Using a combination of bioinformatics-based data mining of existing fungal genome sequencing data and single cell isolation and cultivation techniques, we are beginning to understand the diversity of bacteria that form associations with fungi, and how these associations affect both fungal and bacterial growth.

Abstract

Fungi are cosmopolitan microorganisms with complex genetic make-up and metabolism.[1, 2] Furthermore, this group of microorganisms possesses important roles in ecology, agriculture, forestry and human health. In soil, fungi are one of the most abundant groups of microorganisms and are known to interact with different domains of life, including bacteria. Bacterial:fungal interactions in soils can be positive or negative (synergistic or antagonistic). In the present study, we developed several tools to investigate the potential antagonistic interactions of fungi with bacteria. Fungi have unique responses to stressful conditions, under antagonistic conditions between bacteria and fungi, we hypothesized that there will be certain physiological fungal responses, such as apoptotic-like cell death and production of volatile organic compounds (VOC). We developed methods to determine reactive oxygen species (ROS) and VOC produced by fungi that can trigger apoptotic-like cell death due to antagonistic interactions between bacteria and fungi. Methods to detect two different types of ROS were developed. The ROS investigated were hydrogen peroxide (H_2O_2) and super oxide radical anion ($O_2^{-\cdot}$). In addition to ROS production, fungi are also known to produce VOC. These VOC can represent a significant portion of the fungal metabolome and have been demonstrated to provide information in real-time about metabolic changes that occur in the fungal cell under normal or stressful conditions.[3] We have developed and optimized VOC analysis using solid phase micro-extraction (SPME) coupled to gas chromatography and mass spectrometry (GC/MS). To further investigate the effects of VOC and ROS on fungal survival, we have also developed methods to determine morphological changes associated with apoptosis. Fungi are known to go through early apoptosis and late apoptosis. During early apoptosis there is chromatin condensation. In this study, we used the Hoechst 33258 dye that labels nuclear material, followed by microscopic observation to determine early apoptosis.

Late apoptosis was also investigated by observing hyphal death. For late apoptosis, we used the Evan Blue method, which allowed us to determine plasma integrity of fungal hyphae under normal and stressful conditions. These methods will be essential for the project to gain a better understanding of the antagonistic relationships between bacteria and fungi in soil ecosystems.

References

[1] L. de Aguirre, S.F. Hurst, J.S. Choi, J.H. Shin, H.P. Hinrikson, C.J. Morrison, Rapid Differentiation of Aspergillus Species from Other Medically Important Opportunistic Molds and Yeasts by PCR-Enzyme Immunoassay, Journal of Clinical Microbiology, 42 (2004) 3495-3504.
[2] R.E. McKinney, Environmental pollution control microbiology: a fifty-year perspective, CRC Press, 2004.
[3] L.V. Roze, A. Chanda, M. Laivenieks, R.M. Beaudry, K.A. Artymovich, A.V. Koptina, D.W. Awad, D. Valeeva, A.D. Jones, J.E. Linz, Volatile profiling reveals intracellular metabolic changes in Aspergillus parasiticus: veA regulates branched chain amino acid and ethanol metabolism, BMC Biochemistry, 11 (2010) 33.

This study was supported by a U.S. Department of Energy Biological and Environmental Research Science Focus Area grant (grant no. DE-AC52-06NA25396).

m-CAFÉs Development of novel CRISPR-Cas technologies for precise manipulation of microbial networks

Adam M. Deutschbauer*[1](AMDeutschbauer@LBL.Gov), Peter F. Andeer[1], Jens Heller[1], Peter Kim[2], Trenton Owens[1], Joelle Schlapfer[1], Anup K. Singh[2], Trent R. Northen[1], Jill Banfield[3], Eoin L. Brodie[1], Jennifer A. Doudna[3], Mary K. Firestone[3], Peter S. Nico[1], Rodolphe Barrangou[4] and **N. Louise Glass[1]**

[1]Lawrence Berkeley National Laboratory, Berkeley; [2]Sandia National Laboratory, Livermore; [3]University of California, Berkeley; [4]North Carolina State University, Raleigh

Project Goals: To derive mechanistic understandings of plant-microbe-soil interactions using reproducible, simplified ecosystems

The m-CAFÉs program is a collaborative, coordinated and integrated, mission-driven program to interrogate the function of soil microbiomes with critical implications for carbon cycling and sequestration, nutrient availability and plant productivity in natural and managed ecosystems. As part of this effort, we aim to interrogate the importance and function of individual microorganisms in synthetic and natural microbiomes within realistic, fabricated ecosystems (EcoFABs). However, the technologies to enable this precise manipulation of single species within a complex community environment are currently lacking. Therefore, a major goal of the m-CAFÉs program is to pioneer the development of CRISPR-Cas soil microbiome editing to uncover the functions of cultivated and uncultivated microorganisms. This powerful new platform will provide the fundamental genetic basis for the formation and functional importance of microbial metabolic interaction networks in soil. Here, we present some of our preliminary studies as well as a roadmap for the development of this approach including strategies for delivery and precise microbial targeting and ablation.

Funding statement.

This material by m-CAFÉs 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

DIVA Services: PCR, Full DNA Construction, and MiSeq Validation

Garima Goyal[1,2,3], Nurgul Kaplan[1,2,3], Jennifer L. Chiniquy[2,3], Hector A. Plahar[2,3], Annabel Large[1,3], Lisa Simirenko[3,4], Samuel Deutsch[3,4], and **Nathan J. Hillson**[1,2,3,4]* (njhillson@lbl.gov)

[1]Technology Division, DOE Joint BioEnergy Institute, Emeryville, CA; [2]DOE Agile BioFoundry, Emeryville, CA; [3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [4]DNA Synthesis Science Program, DOE Joint Genome Institute, Walnut Creek, CA.

https://diva.jbei.org

Project Goals:

DNA construction is vital to a broad range of biological research, yet it is predominantly an inefficient diversion from researchers' core research goals and expertise. We have developed a Design, Implementation, and Validation Automation (DIVA) platform to liberate researchers from building and validating DNA, enabling them to instead focus on designing and carrying out their experiments of interest. The DIVA web interface enables users to track the progress of the construction and validation of their DNA designs online. As part of DIVA, a PCR service has been established, wherein researchers specify their designs using design tools in DIVA, and the DIVA team provides researchers gel purified, DpnI-digested PCR products that are ready for DNA assembly and transformation. Recently, DIVA services have been extended to full DNA construction (including DNA assembly, transformation, cloning, and sequence validation steps). A MiSeq Validation service has also been recently added to DIVA, providing full plasmid sequencing coverage at a fraction of the per sample price of single-read Sanger sequencing.

This work was part of the DOE Joint BioEnergy Institute (http:// www.jbei.org) and part of the DOE Joint Genome Institute (http://jgi.doe.gov) 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.