

# 2023 GENOMIC SCIENCE PROGRAM PI MEETING ABSTRACTS

## April 17–19, 2023

U.S. Department of Energy Office of Science Biological and Environmental Research Program Biological Systems Science Division



Biological and Environmental Research Program

# 2023 Genomic Science Program PI Meeting

### April 17-19, 2023

## Rockville, Md.

U.S. Department of Energy Biological and Environmental Research Program Biological Systems Science Division

### **Biological and Environmental Research Program Acting Director**

Gerald Geernaert gerald.geernaert@science.doe.gov

## **Biological Systems Science Division Director**

Todd Anderson todd.anderson@science.doe.gov

## **Genomic Science Program Managers**

**Boris Wawrik** 

boris.wawrik@science.doe.gov Pablo Rabinowicz

pablo.rabinowicz@science.doe.gov
Dawn Adin

dawn.adin@science.doe.gov

Shing Kwok shing.kwok@science.doe.gov Amy Swain amy.swain@science.doe.gov

**Resham Kulkarni** resham.kulkarni@science.doe.gov

Vijay Sharma vijay.sharma@science.doe.gov Kari Perez kari.perez@science.doe.gov

Elizabeth White elizabeth.white@science.doe.gov

Kim Hixson (Lab Detailee) kim.hixson@science.doe.gov

### **About BER**

The Biological and Environmental Research (BER) program supports transformative science and scientific user facilities examining complex biological, Earth, and environmental systems for clean energy and climate innovation. BER research seeks to understand the fundamental biological, biogeochemical, and physical principles needed to predict a continuum of processes occurring across scales, from molecules and genomes at the smallest scales to environmental and Earth system change at the largest scales. This research—conducted at universities, U.S. Department of Energy national laboratories, and research institutions across the country—is contributing to a future of reliable, resilient energy sources and evidence-based climate solutions.

### **About BSSD**

The Biological Systems Science Division (BSSD) seeks to understand, predict, manipulate, and design plant and microbial systems for advances in renewable energy, insights into environmental processes, and biotechnological breakthroughs supporting the U.S. bioeconomy. To expand knowledge of biological systems, BSSD supports basic research and capabilities in foundational genomic science, systems biology, genome engineering, computational analysis, molecular imaging, and structural characterization.

This abstract book is available at https://genomicscience.energy.gov/research-summaries-genomic-science-annual-pi-meeting-april-2023/

# 2023 Genomic Science Program PI Meeting Abstracts



Biological and Environmental Research Program

# Contents

BIOENERGY	1
Bioenergy Research Centers: Center for Advanced Bioenergy and Bioproducts Innovation (CABBI)	2
Integrating Measurements and Models to Improve Projections of Ecosystem Carbon Balance in Bioenergy Agriculture.	2
Artificial Intelligence and Machine Learning for Bioenergy Research in CABBI	3
Discovering Transcriptional Regulators of Photosynthesis in Energy Sorghum to Improve Productivity	4
Increasing the Value of Bioenergy Grasses—Expressing Engineered Traits in the Right Place at the Right Time	5
Feedstocks-to-Fuels Pipeline Demonstration: End-to-End Process Synthesis	6
Overview of CABBI Conversion Theme	7
Bioenergy Research Centers: Center for Bioenergy Innovation (CBI)	8
Al-Informed Systems Biology: The Discovery of Cryptic Phenotypes and the Functional Networks that Control Them	8
Comprehensive Genome-Wide CRISPR Interference Library for High-Throughput Functional Genomic Studies in <i>Pseudomonas putida</i>	10
Deciphering the Genetic Basis of S/G Lignin Variation in Switchgrass Through Quantitative Trait Locus Mapping and Transgenic Approaches	10
Metaproteomic Insights into Mechanisms Enabling Anaerobic, Thermophilic Microbiomes to Achieve Undiminished Fractional Carbohydrate Solubilization at High Solids Loading	11
<i>Populus</i> Secondary Cell Walls: Atomistic Models and Natural Variability in Architecture Informed by Solid-State Nuclear Magnetic Resonance	13
Bioenergy Research Centers: Great Lakes Bioenergy Research Center (GLBRC)	14
High (School)-Throughput Screening of BAHD Transferases	14
Defining Transcriptomic Dynamics in Sorghum in Multiple Abiotic Stresses	15
Novosphingobium aromaticivorans for ccMA Production from Lignin Biomass	15
Structure and Molecular Mechanism of a Bacterial ADP-forming 4-Coumarate CoA Ligase	16
The Genetics of Pathogen and Microbiome Control in the Switchgrass Leaf	17
Towards Generalized Platforms for Functional Genomics in α-Proteobacteria	17
Bioenergy Research Centers: Joint BioEnergy Institute (JBEI)	18
Characterizing Lignocellulose Breakdown Mechanisms in Anaerobic Gut Fungi	18
Evolutionary Engineering is a Versatile Strain Optimization Approach for Sustainable Bioproduction	19
Expression of S-Adenosyl Methionine Hydrolase Modifies Lignin in Sorghum	19
Scenario-Based Technoeconomic and Life-Cycle Analyses of Biomass Conversion Technologies	20
Understanding the Role of Lignin in Native Architecture of Engineered Plant Cell Walls via Multi-Dimensional Solid-State Nuclear Magnetic Resonance	21
Plant Genomics: University Projects	22
Analysis of the Beneficial Associations of Sorghum with Arbuscular Mycorrhizal Fungi Studied with Genetics, Genomics, and Microbiomics	22
Defining the Influence of Environmental Stress on Bioenergy Feedstocks at Single-Cell Resolution	22

El: Pr	ucidating the Genetic Components of the Physiological and Metabolic rocesses Governed by the TORC Regulatory Module in Poplar	23
Ur	nderstanding the Effects of Populus—Mycorrhizal Associations on Plant Productivity and Resistance to Abiotic Stress	24
Pr	redicting Gene Functions in Plants with Single-Cell Genomic Data	25
Ex	xtracting Functional Traits from Large Volumes of Field Phenomics Data	25
Ha	arnessing Regulatory Variation to Elucidate Drought Resilience Mechanisms in Sorghum	27
Sy	ystems Framework to Enhance the Potential of Camelina as Oilseed Crop	27
De	evelopment of a CenH3-Based Haploid Inducer in Hexaploid Camelina sativa	28
Pla	lant Methanol Emission at the Interface of the Photosynthetic C1 Pathway, Leaf Water Status, and Growth	29
Ph	henotypic Characterization of Sorghum Nitrogen Responsive Gene Edits Using High-Throughput Phenotyping	30
Id	lentification of Regulatory Mechanisms Underlying Cell Differentiation in Sorghum Biomass	30
Si	ngle-Cell Genomics of Poplar Wood Development	31
Er	nhanced Resistance Pines for Improved Renewable Biofuel and Chemical Production	32
Im	nproving Candidate Gene Discovery by Combining Multiple Genetic Mapping Datasets	34
53	32 Genomes Reveal Natural Variation and Local Adaptation History in Pennycress	34
A <sup>·</sup> Ar	Temporal Atlas and Response to Nitrate Availability of 3D Root System rchitecture in Diverse Pennycress ( <i>Thlaspi arvense</i> L.) Accessions	35
Co Ao	omparison of Physiological and Metabolomic Responses to Drought cross Pennycress CRISPR Mutants and Natural Accessions	36
Fu	unctional Analysis of Genes Encoding Ubiquitin Proteasome System Components Affecting Poplar Wood Traits	36
W	/INTR: Winter Transcriptome Regulation in Poplar	37
Fu	unctional Characterization of Glycosyltransferases in Duckweed to Enable Predictive Biology	38
А	Leaf-Level Spectral Library to Support High-Throughput Plant Phenotyping: Predictive Accuracy and Model Transfer	38
Im	nproved Biofuel Production Through Discovery and Engineering of Terpene Metabolism in Switchgrass	38
Micr	robial Conversion: University Projects	40
Th fo	he Use of Deuterated Water as a Substrate-Agnostic and Cost-Effective Isotope Tracer or Investigating Reversibility and Thermodynamics of Reactions in Central Carbon Metabolism	40
Pr	robing Lignin Deconstruction and Catabolism in Soil Pseudomonas Species	40
O	ptogenetic Control of a Dual Yeast-Yeast Consortia for Chemical Production	41
Lij	pid Membrane Remodeling and Metabolic Response During Ethanol and Isobutanol Stress in Zymomonas mobilis	42
O	ptimizing Enzymes for Plastic Upcycling Using Machine Learning Design and High Throughput Experiments	43
М	letabolic Modeling and Genetic Engineering of Enhanced Anaerobic Microbial Ethylene Synthesis	43
Ce	ell-Free Conversion of Pyruvate to 2,3-Butanediol Using Co-Substrate Feed as pH Control Strategy	45
Ge Ca	enomic Analyses and Enzyme Characterization Provide Insights into the atabolism of Lignin-Related Aromatic Compounds in White-Rot Fungi	46
Im	nproving Bioprocess Robustness by Cellular Noise Engineering	47
Ha Sv	arnessing Robustness of Thermophilic <i>Bacillus coagulans</i> for Conversion of witchgrass Hydrolysates to Designer Bioesters at Elevated Temperatures	47

	Engineering Bacterial Microcompartments in <i>Clostridium autoethanogenum</i> to Overcome Bottlenecks in Sustainable Production of Synthetic Rubber	48
	Systems Biology to Enable Modular Metabolic Engineering of Fatty Acid Production in Cyanobacteria	49
	Microbial Conversion: Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) programs	51
	Designing Novel Enzymes for Complete Degradation of Recalcitrant Polyamides	51
	Sustainable Bioenergy: University Projects	52
	A Systems Understanding of Nitrogen-Fixation on the Aerial Roots of Sorghum	52
	Optimizing Biological Nitrogen Fixation on Sorghum by Manipulating Microbial Communities	52
	Single-Cell Omics to Examine the Cell Type-Specific Gene Regulatory Programs of Mucilage Production in Sorghum bicolor	53
	EndoPopulus: Elucidating the Molecular Mechanisms of N-Fixation by Populus Endophyte, Burkholderia vietnamiensis WPB	54
	EndoPopulus: Elucidation of the Roles of Diazotrophic Endophyte Communities in Promoting Productivity and Resilience of <i>Populus</i> Through Systems Biology Approaches	54
	EndoPopulus: Endophyte Inoculation Enhances Populus Physiological Responses to Abiotic Stress	55
	Exploring Switchgrass Genetic Diversity with Multiple Reference Genomes	56
	The Influence of Switchgrass Internode Anatomy on Biofuel Production-Relevant Traits	57
	Transcriptome and Gene Regulatory Network Analyses in Camelina Nitrogen Response and Seed Development	57
	The Root Microbiome of Camelina in the Dryland Wheat Production Areas of Eastern Washington	58
	Science Focus Area: Solvent Disruption of Biomass and Biomembranes	60
	Biofuels Disruption of Membrane Domains: An Unrecognized Mode of Solvent Stress	60
	Visualization of Solvent Disruption of Biomass and Biomembrane Structures in the Production of Advanced Biofuels and Bioproducts	61
	Science Focus Area: A Systems Biology Approach to Interactions and Resource Allocation in in Bioenergy-Relevant Microbial Communities	63
	Identifying Genomic and Metabolic Underpinnings of Algal-Bacterial Interactions via Metatranscriptomics, NanoSIMS Isotope Tracing, and Genome-Scale Metabolic Modeling	63
	The Context-Dependency of Plant-Microbial Interactions in the Bioenergy Resource Economy	64
	Science Focus Area: QPSI: Quantitative Plant Science Initiative	66
	Integrating Functional Genomics with Molecular-Level Experimentation to Understand Adaptation to Nutrient Stress in Poplar and Sorghum	66
	Functional Characterization of bHLH Transcription Factors Coordinating Abiotic Stress Response, Secondary Cell Wall Biosynthesis, and Metal Homeostasis in <i>Populus</i>	67
BI	OSYSTEMS DESIGN	68
	University Projects	69
	Accelerating Discovery of Genes Regulating Stomatal Patterning and Water Use Efficiency in C <sub>4</sub> Crops with Novel High-Throughput Methods for Mutagenesis and Phenotyping	69
	Combining GWAS of Metabolomic and Transcriptomic Datasets to Accelerate Discovery of Genes Regulating the Effect of Drought on Plant Growth and Metabolism in <i>Sorghum</i> and <i>Setaria</i>	69
	Understanding and Engineering Crown Root Development to Improve Water-Use Efficiency in Bioenergy Grasses	70

Developing Chassis for LDPE Upcycling from Microbes Native to the Gut Microbiome of Yellow Mealworms	71
Agent-Based Algal Modeling for the Rational Engineering of Chlamydomonas reinhardtii	72
Synthetic Membrane Biology of Microbial Cell Factories: Lipid Interactions that Shape the Inner Mitochondrial Membrane	73
BioPoplar: A Tunable Chassis for Diversified Bioproduct Production	74
B5: Bigger Better Brassicaceae Biofuels and Bioproducts—An Overview	74
FatPlants: A Comprehensive Information System for Lipid-Related Genes and Metabolic Pathways in Plants	75
Genome-Wide Gene Regulation by Transcriptional CRISPRa/i Tools in Nonmodel Bacteria	76
Multiomics-Driven Microbial Model Optimization	76
Pooled Microbial CRISPR Screens Using Single-Cell RNA Sequencing	77
A Pipeline for High-Throughput Genomic Recoding in Organisms Beyond E. coli	77
Construction of a Synthetic 57-Codon <i>E. coli</i> Chromosome to Achieve Resistance to All Natural Viruses, Prevent Horizontal Gene Transfer, and Enable Biocontainment	78
Domesticating Cyanobacteria Through Development of New Genome Engineering Tools and Isolation of New Bio-Prospected Model Strains	79
From Strand Design Principles to SNP Detection—Probing Oligonucleotide Hybridization at the Single Molecule Level	80
Developing Genome Engineering Technologies in Cupriavidus necator for Carbon-Negative Biomanufacturing	81
Novel Systems Approach for Rational Engineering of Robust Microbial Metabolic Pathways	82
A Cell-Free Protein Evolution Platform	83
Accelerating Carbon-Negative Biomanufacturing Through Systems-Level Biology and Genome Optimization	83
Developing, Understanding, and Harnessing Modular Carbon/Nitrogen-Fixing Tripartite Microbial Consortia for Versatile Production of Biofuel and Platform Chemicals	85
A Gapless and Phased Diploid Genome Assembly for <i>Auxenochlorella protothecoides</i> Facilitates Metabolic Modeling and Proteomics Analyses	85
Deploying Top-Down and Bottom-Up Strategies for Genetic Engineering of Auxenochlorella protothecoides for the Production of Sustainable Biofuels	87
Tools and Targets: Engineering Modified Fatty Acids and Improved Photosynthetic Resilience in <i>Auxenochlorella protothecoides</i>	88
Systems Metabolic Engineering of Novosphingobium aromaticivorans for Lignin Valorization	89
Engineering Novel Microbes for Upcycling Waste Plastic	90
Broadening Aromatic Compound Degradation in Acinetobacter baylyi Using Synthetic Metabolic Pathways	90
Engineering Synthetic Anaerobic Consortia Inspired by the Rumen for Biomass Breakdown and Conversion	92
The MoonTag Programmable Transcription Activator and Synthetic Promoters: Tools for Engineering Oil Biosynthesis in <i>Camelina</i> and Pennycress	93
Plastic Degradation by the Gut Microbiome of Yellow Mealworms	94
Developing Anaerobic Fungal Tools for Efficient Upgrading of Lignocellulosic Feedstocks	95
Encapsulin Nanocompartment Systems in Rhodococcus opacus for Compartmentalized Biosynthesis Applications	97
Microbial Guilds and Niches Enable Targeted Modifications of the Microbiome	97
Model-Guided Design of Synthetic Microbial Consortia for Next-Generation Biofuel Production	98
Reproducible Plant Growth in Fabricated Ecosystems (EcoFAB 2.0) Reveals that Nitrogen Supply Modulates Root Exudation	99

Unraveling Metabolic Interactions in a Rhizosphere Microbial Community Through Genome-Scale Modeling	
A Gene-Editing System for Large-Scale Fungal Phenotyping in a Model Wood Decomposer	101
Science Focus Area: IMAGINE BioSecurity: Integrative Modeling and Genome-Scale Engineering for Biosystems Security	103
IMAGINE BioSecurity: Genome-Scale Engineering and Modeling for Secure Biosystems Design	103
IMAGINE BioSecurity: Mesocosm-Based Methods to Evaluate Biocontainment Strategies and Impact of Industrial Microbes Upon Native Ecosystems	
Science Focus Area: From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms	105
Engineering Overlapping Genes in Bacteria	105
Engineered Overlapping Genes Paired with Directed Evolution Prolongs the Evolutionary Stability of a Genetic Circuit	106
Science Focus Area: Persistence Control of Engineered Functions in Complex Soil Microbiomes	107
Development of Synthetic Microbial Communities to Study Consortium Engraftment Dynamics and to Improve the Yield Performance of Sorghum	
Refactoring Metabolism to Control the Persistence of Genetically Engineered Microorganisms in the Environment	108
Science Focus Area: SEED: Secure Ecosystem Engineering and Design	110
Secure Ecosystem Engineering and Design (SEED) to Enable Safe Biodesign of Novel Plant-Microbe Interactions	110
Secure Ecosystem Engineering and Design (SEED) to Mitigate the Impacts of Non-Native Fungal Pathogens on Managed Ecosystems	111
Other: InCoGenTEC	112
Integrases on Demand	112
Integrases on Demand ENVIRONMENTAL MICROBIOME	112 <b>113</b>
Integrases on Demand ENVIRONMENTAL MICROBIOME. University Projects	112 
Integrases on Demand ENVIRONMENTAL MICROBIOME University Projects Conversion of Lignocellulosic Plant Biomass into Industrial Chemicals via Metabolic Engineering of Two Extreme Thermophiles, <i>Caldicellulosiruptor bescii</i> and <i>Pyrococcus furiosus</i>	112 
Integrases on Demand ENVIRONMENTAL MICROBIOME. University Projects Conversion of Lignocellulosic Plant Biomass into Industrial Chemicals via Metabolic Engineering of Two Extreme Thermophiles, <i>Caldicellulosiruptor bescii</i> and <i>Pyrococcus furiosus</i> Metabolome-Informed Proteome Imaging of Lignocellulose Decomposition by a Naturally Evolved Fungal Garden Microbial Consortium	
Integrases on Demand ENVIRONMENTAL MICROBIOME. University Projects Conversion of Lignocellulosic Plant Biomass into Industrial Chemicals via Metabolic Engineering of Two Extreme Thermophiles, <i>Caldicellulosiruptor bescii</i> and <i>Pyrococcus furiosus</i> Metabolome-Informed Proteome Imaging of Lignocellulose Decomposition by a Naturally Evolved Fungal Garden Microbial Consortium Unraveling Spatiotemporal and Physicochemical Constraints on Soil Viral Community Composition and Viral Particle Integrity	
Integrases on Demand ENVIRONMENTAL MICROBIOME. University Projects Conversion of Lignocellulosic Plant Biomass into Industrial Chemicals via Metabolic Engineering of Two Extreme Thermophiles, <i>Caldicellulosiruptor bescii</i> and <i>Pyrococcus furiosus</i> Metabolome-Informed Proteome Imaging of Lignocellulose Decomposition by a Naturally Evolved Fungal Garden Microbial Consortium Unraveling Spatiotemporal and Physicochemical Constraints on Soil Viral Community Composition and Viral Particle Integrity Metagenomic Insights into Microbial Traits Influencing Community Dynamics and Nitrogen Cycling Post-Fire	
Integrases on Demand	
Integrases on Demand ENVIRONMENTAL MICROBIOME. University Projects. Conversion of Lignocellulosic Plant Biomass into Industrial Chemicals via Metabolic Engineering of Two Extreme Thermophiles, <i>Caldicellulosiruptor bescii</i> and <i>Pyrococcus furiosus</i> Metabolome-Informed Proteome Imaging of Lignocellulose Decomposition by a Naturally Evolved Fungal Garden Microbial Consortium Unraveling Spatiotemporal and Physicochemical Constraints on Soil Viral Community Composition and Viral Particle Integrity Metagenomic Insights into Microbial Traits Influencing Community Dynamics and Nitrogen Cycling Post-Fire Predicting Post-Fire Nitrogen Cycling through Traits and Cross-Kingdom Interactions Impacts of Altered Climate on Microbial Growth and Nutrient Assimilation in an Ombitrophic Peat Bog	
Integrases on Demand	
Integrases on Demand	
Integrases on Demand	
Integrases on Demand ENVIRONMENTAL MICROBIOME. University Projects. Conversion of Lignocellulosic Plant Biomass into Industrial Chemicals via Metabolic Engineering of Two Extreme Thermophiles, <i>Caldicellulosiruptor bescii</i> and <i>Pyrococcus furiosus</i> Metabolome-Informed Proteome Imaging of Lignocellulose Decomposition by a Naturally Evolved Fungal Garden Microbial Consortium Unraveling Spatiotemporal and Physicochemical Constraints on Soil Viral Community Composition and Viral Particle Integrity Metagenomic Insights into Microbial Traits Influencing Community Dynamics and Nitrogen Cycling Post-Fire Predicting Post-Fire Nitrogen Cycling through Traits and Cross-Kingdom Interactions Impacts of Altered Climate on Microbial Growth and Nutrient Assimilation in an Ombitrophic Peat Bog The Predictive Power of Phylogeny on Growth Rates in Soil Bacterial Communities. Ultrasensitive Protein SIP to Quantify Activity and Substrate Uptake in Microbiomes with Stable Isotopes Computational Tools for Multiomic Data Standardization and Integration to Represent Whole Microbial Communities Uncovering the Microbial Networks that Degrade Plant-Derived Phenolic Compounds and Their Role in Peatland Soil Carbon Sequestration: Revisiting the 'Enzyme Latch' Hypothesis	
Integrases on Demand	

Fungal-Bacterial Interactions: Bridging Soil Niches in Regulating Carbon and Nitrogen Processes	
Using <sup>13</sup> CO <sub>2</sub> to Track Carbon Flow Mediated by Fungal-Bacterial Interactions in Grassland Soils	
Crosstalk: Interkingdom Interactions in the Mycorrhizal Hyphosphere and Ramifications for Soil Carbon Cycling	
Deciphering Virocell Metabolic Dynamics and Ecosystem Outputs Using a Novel Integration of Machine Learning and Metabolomics	
Diversity Patterns and Ecological Footprint of RNA Viruses Along a Permafrost Thaw Gradient	130
Genetic Determinants of Klebsiella Phage Infection	
From Bulk Organic Matter Profiling to Specific Metabolite Identification: Improving Metabolomics Data Analysis, Annotation, Interpretation, and Integration	
Gatekeepers of Arctic Carbon Loss: Landscape-Scale Metabolites-to-Ecosystems Profiling to Mechanistically Map Climate Feedbacks	133
Construction of a Microbial Methane Observatory Reveals Metabolic Dynamics of Freshwater Wetland Microbiomes	
Putting Microorganisms on the Map: Continental-Scale Context for Thousands of Newly Sampled Microbial Genomes from North American Wetlands	
Science Focus Area: Bacterial-Fungal Interactions and Their Role in Soil Functioning	
An Online Public Resource for Bacterial-Fungal Interaction Research	
Expanding Knowledge of Bacterial-Fungal Interactions in Environmental Microbiomes	
Science Focus Area: ENIGMA: Ecosystems and Networks Integrated with Genes and Molecular Assembl	ies138
ENIGMA Environmental Atlas: An Integrated Approach to Linking Microbial Genotype to Phenotype in a Dynamic Subsurface Ecosystem	
ENIGMA Long Read Sequencing and Assembly for Microbial Genomes: KBase Integration for Assembly and LISA Workshop	
The Use of Synthetic Communities Reveals Disturbance of Process Partitioning Among Denitrifying Microbes Leads to Increased Nitrous Oxide Emissions	
Three-Dimensional High Spatial Resolution Simulation for Groundwater Flow and Nitrogen Transport Under Rainfall Perturbations in the Subsurface of Area 3	140
Science Focus Area: m-CAFEs: Microbial Community Analysis and Functional Evaluation in Soils	142
Examine Plant-Microbe-Soil Interactions Using Fabricated Ecosystems Across Scales	142
Phage Engineering for Targeted Editing of Microbial Communities	
Science Focus Area: Phenotypic Response of the Soil Microbiome to Environmental Perturbations	
Impact of Soil Viruses on Microbial Compositions and Functions	
OMEGGA: A Computationally Efficient Omics-Guided Global Gapfilling Algorithm for Phenotype-Consistent Metabolic Network Reconstruction	
Removal of Primary Nutrient Degraders Reduces Growth and Modifies Functional Pathways of Soil Microbial Communities with Genomic Redundancy	
Science Focus Area: Plant-Microbe Interfaces	148
Plant-Microbe Interfaces: Capturing and Interpreting the Role of <i>Populus'</i> Microbiome	
Plant-Microbe Interfaces: Molecular Insights into the Mutualistic Symbiosis Between <i>Populus</i> and Plant Growth–Promoting Bacteria	

Science Focus Area: Microbes Persist: Systems Biology of the Soil Microbiome	150
Functional Succession of Growing Soil Microorganisms and Virus-Driven Mortality Following Rewetting in a California Grassland Soil	150
Microbes Persist: Towards Quantitative Theory–Based Predictions of Soil Microbial Fitness, Interaction, and Function in KBase	
The Path from Root Input to Mineral-Associated Soil Carbon is Dictated by Habitat-Specific Microbial Traits	152
Other: The Twin Ecosystems Project	
The Twin Ecosystems Project: A New Capability for Field and Laboratory Ecosystems Coupled by Sensor Networks and Autonomous Controls	153
COMPUTATIONAL BIOLOGY	154
University Projects	155
Scalable Computational Tools for Inference of Protein Annotation and Metabolic Models in Microbial Communities	155
HypoRiPPAtlas: An Atlas of Hypothetical Natural Products for Mass Spectrometry Database Search	156
Predicting and Modeling Protein-Protein Complexes at Large-Scale with Deep Learning	157
Developing a National Virtual Biosecurity for Bioenergy Crops Center	
ENABLING CAPABILITIES	159
DOE Joint Genome Institute	160
Quantify the Impact of Your Data with DOE JGI's Genome Citation Service	160
RNA Phages: Under-Estimated Players in Soil Ecosystems?	
KBase: DOE Systems Biology Knowledgebase	
KBase Science and Infrastructure Updates	
Learning and Training with KBase	
The KBase Knowledge Engine: Ecosystem Classification Prototype	164
NLP for Synthetic Biology: Providing Generalizable Literature Mining Through KBase	165
Probabilistic Annotation and Ensemble Metabolic Modeling in KBase	
Crosslinking The Department of Energy Systems Biology Knowledgebase (DOE-KBase) and the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RSCB PDB) to Support Protein Function Discovery	167
Integration of Enzyme Function Initiative Tools in the KBase Platform	168
Design and Omics Exploration of Synthetic Microbial Communities in KBase	169
NMDC: DOE National Microbiome Data Collaborative	171
Community Engagement Strategies of the National Microbiome Data Collaborative (NMDC)	171
Portals of Discovery: The NMDC Infrastructure and Products for Microbiome Research	172
INDEX	173



## **BIOENERGY RESEARCH CENTERS**

Center for Advanced Bioenergy and Bioproducts Innovation (CABBI)

## Integrating Measurements and Models to Improve Projections of Ecosystem Carbon Balance in Bioenergy Agriculture

Stephanie M. Juice<sup>1,2\*</sup> (stephanie.juice@mail.wvu.edu), Carl J. Bernacchi<sup>1,3</sup>, Bethany J. Blakely<sup>1,3</sup>, Melannie D. Hartman<sup>1,4</sup>, Caitlin E. Moore<sup>1,3</sup>, William J. Parton<sup>1,4</sup>, Bryan M. Petersen<sup>1,5</sup>, Joanna R. Ridgeway<sup>1,2</sup>, Jacob E. Studt<sup>1,5</sup>, Andy VanLoocke<sup>1,5</sup>, Wendy H. Yang<sup>1,3</sup>, **Edward R. Brzostek**<sup>1,2</sup>

<sup>1</sup>Center for Advanced Bioenergy and Bioproducts Innovation; <sup>2</sup>West Virginia University; <sup>3</sup>University of Illinois–Urbana; <sup>4</sup>Colorado State University; <sup>5</sup>Iowa State University

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

**Goals:** The goal of the Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) sustainability research is to design a sustainable bioeconomy (See Fig. 1, p. 3). A critical part of meeting this goal is to integrate empirical measurements with models to project future scenarios of bioenergy systems that can inform sustainable management choices that enhance carbon (C) sequestration and nitrogen (N) retention. Here, researchers present multiple case studies demonstrating how observational and experimental measurements have been used to enhance the predictive capabilities of ecosystem models.

Research in the CABBI sustainability theme aims to understand the interactions between candidate bioenergy feedstocks, and the environmental conditions and geographic locations under which they are grown in order to inform the understanding of key factors needed for the economically and environmentally sustainable production of bioenergy and bioproducts for fossil fuel displacement. Within this framework, advancing predictive understanding of bioenergy systems through ecological modeling is critical to address how to provide energy, economic, and ecosystem C benefits to help slow the rate of climate change. To meet this challenge, CABBI researchers couple empirical measurements to ecosystem models that simulate plant yields and nutrient dynamics in order to predict future ecosystem C stocks under different environmental and feedstock scenarios. Here, researchers present three case studies illustrating how CABBI researchers integrate data with models to improve projections of bioenergy C balance.

Researchers developed a new bioenergy ecosystem model, FUN-BioCROP that simulates plant-microbe interactions, microbial physiology, and emerging mechanisms of stable soil C creation and Model parameterization integrated both observational and experimental data. On the observational side, this included long-term measurements of total soil C pools and the form of soil C stabilization under different bioenergy crops. On the experimental side, researchers used a novel laboratory experiment to trace the fate of two bioenergy litters (*Miscanthus* and corn) to improve the parameterization of key microbial traits. When the team ran the improved model forward, researchers found divergent responses of bioenergy feedstocks to environmental change.

The predictive ability of ecosystem models is primarily constrained by data quality and quantity. Using model sensitivity analyses, researchers identified root and rhizome biomass, the response of plant C allocation to N fertilization and cycling of the litter layer as key data limitations to the predictive understanding of *Miscanthus* by the Agro-IBIS model. To address these limitations, the team designed targeted field campaigns. The team found that N fertilization alters both the magnitude and timing of belowground C allocation and that the litter layer in *Miscanthus* comprises ~1/4 of the total aboveground biomass. Both findings represent key processes and C pools that are priorities for future model revisions.

DayCent has been at the forefront of models used to predict the C and N consequences of bioenergy production. A common criticism of DayCent has been its assumption that soil decomposition follows first order decay. To address this criticism, researchers integrated the explicit microbial dynamics of FUN-BioCROP into DayCent. The microbial DayCent model better captured the seasonal profile of ecosystem respiration of *Miscanthus* and switchgrass derived from eddycovariance measurements. Moreover, it also showed an upper limit to soil C accumulation over time, whereby ongoing plant inputs enhanced microbial biomass leading to priming losses of soil C. These results have important implications for estimating soil C accumulation in the emerging bioeconomy.

Across these case studies, the integration of empirical data into models resulted in improved C balance projections that represent the most up to date state of knowledge of the factors creating persistent soil C and allowing for the most sustainable bioenergy crop production.

**Funding Information**: This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and



**Fig. 1.** CABBI sustainability research integrates measurements with models across scales to identify the combinations of feedstocks, land types, market conditions, and bioproducts that will be able to sustainably displace fossil fuels. [Courtesy University of Illinois at Urbana-Champaign for the Center for Advanced Bioenergy and Bioproducts Innovation (CABBI)]

conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

### Artificial Intelligence and Machine Learning for Bioenergy Research in CABBI

Huimin Zhao<sup>1\*</sup> (zhao5@illinois.edu), Yasuo Yoshikuni<sup>2</sup>, Jonathan Sweedler<sup>1</sup>, Andy VanLoocke<sup>3</sup>, Matt Hudson<sup>1</sup>, **Andrew D. B. Leakey**<sup>1</sup>

<sup>1</sup>University of Illinois–Urbana-Champaign; <sup>2</sup>Lawrence Berkeley National Laboratory; <sup>3</sup>Iowa State University

#### https://cabbi.bio

**Goals:** Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) aims to develop and apply a wide variety of artificial intelligence (AI)/machine learning (ML) tools for bioenergy research ranging from plant growth to biosystems design to process development to sustainability study.

Thanks to recent advances in genomics, data science, and automation, there is a growing trend in developing and applying AI/ML tools for bioenergy research. In the past few years, CABBI has made much progress in this emerging research area and is now poised to take the lead in developing it into a mature field. Researchers will highlight a few representative AI/ML projects in CABBI. These examples include: (1) leveraging AI/ML to predict end-of-season yield and phenotype for sorghum and Miscanthus from an aerial imagery time series (Leakey lab), (2) designing gRNA for CRISPR-based genetic engineering of plants (Hudson lab), (3) developing a self-driving biofoundry for biosystem design (Zhao/Sweedler labs), (4) enabling single nucleotide polymorphisms genotyping of Issatchenkia orientalis by AI/ML (Yoshikuni lab), and (5) using AI/ML and satellite data to predict Miscanthus yields in research plots and commercial fields (VanLoocke lab).

HamediRad, M., et al. 2019. "Towards a Fully Automated Algorithm-Driven Platform for Biosystems Design," *Nature Communications* **10**, 5150.

Paul, H. M., et al. 2022. "CROPSR: An Automated Platform for Complex Genome-Wide CRISPR gRNA Design and Validation," *BMC Bioinformatics* 23, 74.

Petersen, B., et al. 2021, December 13–17. Using Cross-Scale Data to Constrain an Agro-Ecosystem Model to Produce Estimates of Miscanthus Production at a Field-Scale. American Geophysical Union (AGU) Fall Meeting 2021, New Orleans, LA.

Varela, S., et al. 2022. "Deep Convolutional Neural Networks Exploit High-Spatial- and -Temporal-Resolution Aerial Imagery to Phenotype Key Traits in *Miscanthus*," *Remote Sensing* **14**(21), 5333.

Yoshikuni, Y., et al. "SNP Genotyping." In preparation.

Yu, T., et al. 2023. "*In Vitro* Continuous Protein Evolution Empowered by Machine Learning and Automation," *Cell Systems* S2405-4712(23)00115-1. DOI:10.1016/j.cels.2023.04.006.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0018420.

## Discovering Transcriptional Regulators of Photosynthesis in Energy Sorghum to Improve Productivity

Laurie Leonelli<sup>1</sup>, Elena Pelech<sup>1\*</sup> (pelech@illinois.edu), Atinder Singh<sup>1</sup>, Kithmee De Silva<sup>1</sup>, Matthew Brooks<sup>2</sup>, **Steve Long<sup>1</sup>** 

<sup>1</sup>University of Illinois at Urbana–Champaign; and <sup>2</sup>Global Change and Photosynthesis Research Unit, United States Department of Agriculture Agricultural Research Service

### https://photosynthesis.web.illinois.edu

**Goals:** This research aims to identify and investigate the transcription factors involved in the regulation of photosynthesis in energy sorghum. The major goal of this project is to model and validate gene regulatory networks that reveal the relationship between transcription factors and photosynthesis, particularly those that cause a loss of efficiency in lower leaf canopy leaves. This information will allow researchers to rank transcription factors by importance and thus, will guide future design strategies for developing energy sorghum cultivars with improved photosynthetic light-use efficiency in overall productivity.

 $C_4$  grasses such as annual energy sorghum hybrids (*Sorghum bicolor*) have great potential for both carbon sequestration and as feedstocks for biofuels and building materials. Sorghum is also exceptionally drought tolerant, which allows cultivation on land that is marginal for most food crops and as a vegetative crop, it also avoids the problems faced by grain crops during the water deficit sensitive reproductive

phase (Mullet et al. 2014). However, in contrast to most plants, sorghum belongs to a clade of C<sub>4</sub> species that has undergone a maladaptive loss of photosynthetic efficiency in self-shaded leaves within the canopy and current models predict that this loss results in a 15-20% reduction in potential productivity (Pignon et al. 2017). Specifically, most plants have evolved to dynamically tune their photosynthetic machinery by shifting the stoichiometry of proteins involved in the light reactions of photosynthesis to maintain a high maximum absolute quantum efficiency of CO<sub>2</sub> assimilation ( $\Phi CO_2$ , max) in the shade. Seminal work has shown that the lower self-shaded leaves from C4 bioenergy crops (bioenergy sorghum, Miscanthus, and maize) do not retain a high  $\Phi CO_2$ , max compared to their upper sun-exposed leaves, which is due to the change in light environment, not leaf age (Pignon et al. 2017; Collison et al. 2020). Variation in the severity of this  $\Phi CO_2$ , max loss between sorghum cultivars suggests that this maladaptive trait may be the result of difference in the expression of one or more genes (Jaikumar et al. 2021). Since transcription factors (TFs) are key regulators of gene expression in response to environmental stimuli such as changes in light intensity and quality, researchers hypothesize that key TFs cause the observed maladaptive loss of photosynthetic efficiency in energy sorghum and optimizing their expression will restore photosynthetic efficiency and alleviate suboptimal  $\Phi CO_2$ , max in the shaded canopy. Researchers further hypothesize that genes influencing  $\Phi CO_2$ , max will have expression patterns that correspond to measurable changes in photosynthetic traits and that researchers will be able to identify these genes by comparing changes in expression in response to the light environment across energy sorghum cultivars and canopy positions. Therefore, researchers will identify these key transcription factors by analyzing variations in gene expression and photosynthetic traits such as  $\Phi CO_2$ , max across light conditions and sorghum cultivars. Researchers will also use in planta validation of TF gene targets to model a gene regulatory network to describe the regulation of photosynthesis in sorghum. Identifying the cause of photosynthetic inefficiency in shaded energy sorghum canopies and engineering solutions to restore the 15-20% loss in productivity and enhance yield will improve the overall potential of this bioenergy crop to meet the growing needs for energy security.

Collison, R. F., et al. 2020. "Not Age, Underlies the Maladaptation of Maize and *Miscanthus* Photosynthesis to Self-Shading," *Frontiers in Plant Science* **11**, DOI:10.3389/fpls.2020.00783.

Jaikumar, N. S. et al. 2021. "Can Improved Canopy Light Transmission Ameliorate Loss of Photosynthetic Efficiency in the Shade? An Investigation of Natural Variation in *Sorghum Bicolor*." *Journal of Experimental Botany* **72**(13), 4965–80. DOI:10.1093/jxb/erab176.

Mullet, J., et al. 2014. "Energy Sorghum—A Genetic Model for the Design of C4 Grass Bioenergy Crops," *Journal of Experimental Botany* **65**(13), 3479–89. DOI:10.1093/jxb/eru229.

Pignon, C. P., et al. 2017. "Loss of Photosynthetic Efficiency in the Shade. An Achilles Heel for the Dense Modern Stands of Our Most Productive C4 Crops?" *Journal of Experimental Biology* **68**(2), 335–45. DOI:10.1093/jxb/erw456.

**Funding Information:** This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0023107. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science program, Office of Biological and Environmental Research, under FWP ERKP123.

## Increasing the Value of Bioenergy Grasses—Expressing Engineered Traits in the Right Place at the Right Time

Jie Fu<sup>1,2\*</sup>(jief3@illinois.edu), Brian McKinley<sup>3,4</sup>, Brandon James<sup>5</sup>, William Chrisler<sup>6</sup>, Lye Meng Markillie<sup>6</sup>, Matthew J. Gaffrey<sup>6</sup>, Hugh D. Mitchell<sup>6</sup>, Galya Orr<sup>6</sup>, Kankshita Swaminathan<sup>2,5\*</sup>, John Mullet<sup>3,4</sup>, **Amy Marshall-Colon<sup>1,2</sup>** 

<sup>1</sup>University of Illinois Urbana–Champaign; <sup>2</sup>Center for Advanced Bioenergy and Bioproducts Innovation; <sup>3</sup>Texas A&M University; <sup>4</sup>Great Lakes Bioenergy Resource Center (GLBRC); <sup>5</sup>HudsonAlpha Institute for Biotechnology; <sup>6</sup>Pacific Northwest National Laboratory

#### https://cabbi.bio/

**Goals:** The overarching goal of the Center for Advanced Bioenergy and Bioproducts Innovation (CABBI) feedstock production research is to deliver resilient, highly productive grasses that contain large amounts of lipids. Researchers have made significant advances in engineering production of oils, specialty fatty acids, and other organic compounds in vegetative tissues of these grasses, as well as increasing biomass yield, resource use efficiency, and environmental resilience. A specific challenge researchers are still working on is targeting expression of engineered traits in the right place at the right time, paving the way for CABBI crops that produce oil in stem storage tissues at the end of the season.

**Overview:** To contribute to the development of the bioeconomy, feedstocks need to be improved to be economically and environmentally sustainable for both processor and farmer. CABBI strives to do this by engineering carbon allocation to produce oils, specialty fatty acids, and other organic compounds, as well as by increasing biomass yield, resource use efficiency, and environmental resilience. Because such improvements require coordinated changes in multiple plant traits, researchers are working simultaneously on trait improvements and how best to combine and implement them to get the traits in the right place at the right time. In the high biomass C<sub>4</sub> grass feedstocks—sorghum, *Miscanthus*, and sugarcane—the right place for value-added

products is stems. Because the right time for different traits varies, researchers are performing analyses across multiple key developmental time points.

Spatial atlas of the sorghum stem: Bioenergy sorghum's 4-5m stems account for ~80% of the harvested biomass. Stems accumulate high levels of sucrose that could be used to synthesize bioproducts if information about stem celltype gene expression and regulation was available to enable engineering. To obtain this information, Laser Capture Microdissection (LCM) was used to isolate transcriptome profiles from five major cell types present in vegetative stems of Sorghum bicolor L. Moench cv. Wray. Transcriptome analysis identified genes with cell-type specific and cell-preferred expression patterns that reflect the distinct characteristics and regulatory functions of each cell type. The newly discovered cell type specific genes can be used as markers for downstream analyses, such as single cell transcriptomics. Analysis of cell-type specific gene regulatory networks (GRNs) revealed that (1) different biological functions distinguish vascular and non-vascular cell types; (2) distinct transcription factor families regulate the cell type specific expression of genes; and (3) cell type specific transcription factors have both direct and indirect modes of regulation to modulate the expression of cell type specific genes. The team used the LCM data to gain insights into stem secondary cell wall (SCW) networks. By combining the spatial resolution of the LCM-derived stem cell-type specific transcriptome with a stem developmental profile of SCW formation, researchers uncovered (1) the previously unknown spatial expression of key SCW genes across sorghum stem-cell types, (2) cell-type-specific SCW regulatory networks and network motifs, and (3) potential regulators that repress SCW formation in pith parenchyma cells. The cell-type transcriptomic dataset provides a valuable source of information about the function of sorghum stems and GRNs that will enable the engineering of bioenergy sorghum stems.

Future directions: The spatial transcriptomes provide rich information about steady-state gene expression and identified cell-type specific hub genes that likely play key regulatory roles in signaling within each cell type. However, the products of gene expression, the proteome and metabolome, better reflect macro-level phenotypes because they strengthen the link between gene expression and gene function. A continuing collaboration among CABBI, GLBRC, and Environmental Molecular Sciences Laboratory (EMSL) will expand the sorghum stem molecular atlas by incorporating spatial proteomic and metabolomic data into the existing GRNs, which will improve network predictions by more directly linking genes to stem phenotypes. Likewise, researchers will expand the network analysis to include spatial transcriptomes from two additional developmental time points (onset of anthesis and post-anthesis). These combined analyses will reveal the spatiotemporal dynamics of stem metabolic and signaling networks. This deep

understanding of these spatiotemporal dynamics is critical to being able to engineer stems and redirect bioproducts to the right place at the right time.

Connections across Feedstocks: An actionable outcome of the above analyses is the identification of specific promoter elements that drive cell type specific expression at different points in development. Such molecular tools provide a direct path to engineer sorghum stems to accumulate high-value bioproducts of interest to CABBI and its sister Bioenergy Research Centers (BRCs) and decrease the burden on conversion groups in the biofuel industry. Hence, this is a cross-BRC priority. With the close relation between sorghum, sugarcane, and Miscanthus, the team expects that these findings will provide insights into stem-specific expression across the feedstocks of interest. The team also expects many of the findings and methodologies from this study to be useful to researchers interested in engineering other grasses of interest, such as switchgrass and maize. Combining this knowledge with CABBI advances in accumulating oil in sugarcane and sorghum, improvements in water use efficiency and photosynthesis, and breeding to identify plants with higher yields and greater geographic range will generate feedstocks that can be a foundation for a strong bioeconomy.

**Funding Information:** This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-C0018420). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.

### Feedstocks-to-Fuels Pipeline Demonstration: End-to-End Process Synthesis

Shivali Banerjee<sup>1,2\*</sup> (shivalib@illinois.edu), Narendra Naik Deshavath<sup>1,2</sup>, Shraddha Maitra<sup>1,2</sup>, Somesh Mishra<sup>2</sup>, Fredy Altpeter<sup>1,3</sup>, John Shanklin<sup>1,4</sup>, Erik J. Sacks<sup>2</sup>, Bruce S. Dien<sup>1,5</sup>, Huimin Zhao<sup>2</sup>, **Vijay Singh<sup>1,2</sup>** 

<sup>1</sup>Center for Advanced Bioenergy and Bioproducts Innovation (CABBI); <sup>2</sup>University of Illinois Urbana–Champaign; <sup>3</sup>University of Florida; <sup>4</sup>Brookhaven National Laboratory; and <sup>5</sup>National Center for Agricultural Utilization Research (NCAUR), Agricultural Research Service (ARS), United States Department of Agriculture (USDA), Peoria, IL

#### https://cabbi.bio

**Goals:** One goal of the "Feedstocks-to-Fuels Pipeline" project was to demonstrate pilot- scale processing of CABBI feedstocks, namely genetically modified sugarcane (oilcane) and *Miscanthus x giganteus* for the recovery of vegetative lipids, microbial lipids, succinic acid, and anthocyanins as the main products. Overall, this work on end-to-end process synthesis tests the performance of the CABBI feedstocks, the processing methods, and the yeasts engineered to produce bioproducts at an industrially relevant scale using the Integrated Bioprocessing Research Laboratory identifies technology gaps and develops enabling technologies.

The Feedstocks-to-Fuels pipeline synthesizes advances in feedstocks, bioprocessing technology, and engineered yeasts. The deconstruction of the CABBI feedstocks is the most crucial step in recovering in planta products such as oil, waxes, sugars, and pigments. The pilot scale processing of CABBI feedstocks, such as oilcane and purple stemmed Miscanthus x giganteus, involves the development of biomass deconstruction strategies followed by their conversion into value-added products. Oilcane is produced by the metabolic engineering of sugarcane to accumulate lipids in vegetative tissues (Parajuli et al. 2020). The high biomass productivity of this transgenic bioenergy crop holds the potential to produce more oil per hectare of cultivated land than soybean (Huang et al. 2016). The vegetative lipids present in oilcane have the potential for biodiesel production while the oilcane juice, rich in sugars, can be used for the production of value-added biochemicals.

Oilcane stems were received from the University of Florida and were processed at Integrated Bioprocess Research Laboratory. About 218 kg of juice and 230 kg of wet bagasse were recovered by processing 466 kg of oilcane stems. The oilcane juice was used to produce succinic acid using a novel metabolically engineered Issatchenkia orientialis at an acidic pH of 3 and achieved a 54 g/L titer. In the downstream processing, 63.9% succinic acid was recovered from the fermentation broth with 98.5% purity through filtration, followed by decolorization and crystallization. The processed oilcane bagasse was pretreated through a continuous pilotscale hydrothermal process at 50% (w/w) solids followed by mechanical refining (HMR) for the deconstruction of the lignocellulosic network. Fed-batch enzymatic hydrolysis of pretreated bagasse was performed to achieve industrially relevant cellulosic sugar concentrations. Processing with HMR did not incur any changes to the *in situ* lipid profile in the oilcane bagasse and gives optimal recovery of lignocellulosic sugars for conversion to microbial lipids. The major fraction of vegetative lipids were thus recovered from the biomass residue following enzymatic saccharification of the pretreated bagasse (Maitra et al. 2022).

Purple-stemmed *Miscanthus* x *giganteus* is another CABBI feedstock and is known for its potential to accumulate natural colorants namely anthocyanins. The overall yield of anthocyanins recovered per unit area is significant due to the high productivity of this bioenergy crop (Banerjee et al. 2023).

Preliminary studies showed that hydrothermal pretreatment of *Miscanthus* could be used as a green approach to recover more than 90% of the total anthocyanins as an additional product stream and also enhanced the enzymatic digestibility of the biomass (Banerjee et al. 2022). For pilot-scale demonstration, 50 kg of purplestemmed *Miscanthus x giganteus*, grown at the Energy Farms at the University of Illinois, was processed through a continuous pilot-scale hydrothermal pretreatment at 50% w/w solids. The pretreated biomass was further subjected to disc milling. The pretreatment led to a recovery of 94.3% w/w of the total anthocyanins present in *Miscanthus* and also improved the enzymatic digestibility of cellulose leading to a 2.1-fold increase in the overall recovery of glucose. The cellulosic sugars thus obtained were converted into microbial lipids using an oleaginous yeast strain, which are a potential feedstock for biodiesel production.

Overall, the Feedstocks-to-Fuel pipeline successfully demonstrated the potential of two CABBI feedstocks, namely oilcane and purple-stemmed *Miscanthus*. The valorization products were vegetative and microbial lipids for biofuel production along with the production of succinic acid and anthocyanins as additional value-added product streams.

Banerjee, S., et al. 2022. "Valorization of *Miscanthus x giganteus* for Sustainable Recovery of Anthocyanins and Enhanced Production of Sugars," *Journal of Cleaner Production* **369**, 133508.

Banerjee, S., et al. 2023. "Bioenergy Crops as Alternative Feedstocks for Recovery of Anthocyanins: A Review," *Environmental Technology & Innovation* **29**, 102977.

Huang, H., et al. 2016. "Techno-Economic Analysis of Biodiesel and Ethanol Co- production from Lipid-Producing Sugarcane," *Biofuels, Bioproducts and Biorefining* **10**, 299–315.

Maitra, S., et al. 2022. "Bioprocessing, Recovery, and Mass Balance of Vegetative Lipids from Metabolically Engineered "Oilcane" Demonstrates Its Potential as an Alternative Feedstock for Drop-In Fuel Production," ACS Sustainable Chemistry & Engineering **10**, 16833–44.

Parajuli, S., et al. 2020. "Towards Oilcane: Engineering Hyperaccumulation of Triacylglycerol into Sugarcane Stems," *GCB Bioenergy* **12**, 476–490.

### **Overview of CABBI Conversion Theme**

Yihui Shen<sup>1,\*</sup> (yihuis@princeton.edu), José L. Avalos<sup>1</sup>, Bruce Dean<sup>2</sup>, David Flaherty<sup>3</sup>, Jeremy Guest<sup>3</sup>, George Huber<sup>4</sup>, Matt Hudson<sup>3</sup>, Heng Ji<sup>3</sup>, Yong-Su Jin<sup>3</sup>, Ting Lu<sup>3</sup>, Costas D. Maranas<sup>5</sup>, Sergei Maslov<sup>3</sup>, Melanie McReynolds<sup>5</sup>, Brian F. Pfleger<sup>4</sup>, John Shanklin<sup>6</sup>, Zengyi Shao<sup>7</sup>, Vijay Singh<sup>3</sup>, Saurabh Sinha<sup>3</sup>, Jonathan Sweedler<sup>3</sup>, Martin Wühr<sup>1</sup>, Yasuo Yoshikuni<sup>8</sup>, Chengxiang Zhai<sup>3</sup>, Joshua Rabinowitz<sup>1</sup>, Christopher Rao<sup>3</sup>, **Huimin Zhao<sup>3</sup>** 

<sup>1</sup>Princeton University; <sup>2</sup>United States Department of Agriculture, Agricultural Research Service (USDA ARS), Peoria, IL; <sup>3</sup>University of Illinois Urbana–Champaign; <sup>4</sup>University of Wisconsin–Madison; <sup>5</sup>The Pennsylvania State University; <sup>6</sup>Brookhaven National Laboratory; <sup>7</sup>Iowa State University; <sup>8</sup>Lawrence Berkeley National Laboratory

### https://cabbi.bio/research/conversion-theme/

**Goals:** The overarching goal of the Conversion theme is to investigate the cellular metabolism and gene regulation mechanisms of three nonmodel yeasts including *Issatchenkia orientalis, Yarrowia lipolytica* and *Rhodosporidium toruloides* and to develop new synthetic biology and systems biology tools for characterization and engineering of these yeasts for production of chemicals and fuels from renewable plant biomass.

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 (TRYs) remains a significant challenge. Key reasons include: (1) lack of understanding of how native metabolism and physiology constrains the production of non-natural compounds; (2) difficulty in identifying compounds that can be efficiently produced in living organisms and the best host, natural or engineered, for doing so; (3) the time-consuming and expensive designbuild-test-learn (DBTL) cycle for metabolic engineering; and (4) lack of known enzymes with desired activity and substrate specificity for the synthesis of target natural or non-natural compounds.

To address these scientific challenges, the Conversion theme will build on the accomplishments of the first five years of research and focus on the following main objectives: (1) Develop a self-driving biofoundry for metabolic engineering and enzyme engineering; (2) Develop artificial intelligence (AI)/machine learning (ML) algorithms for biosystem design; (3) Engineer nonmodel yeasts for cost-effective production of four main target chemicals: 3-hydroxypropanoic acid (3-HP) and citramalate in *Issatchenkia orientalis* and triacetic acid lactone (TAL) and fatty alcohols in *Yarrowia lipolytica* and *Rhodosporidium*  *toruloides.* The metabolic engineering efforts will continue to be accelerated by the tools developed, (4) Investigate the underlying biological mechanisms that lead to highlevel production of target products using systems biology and biofoundry; (5) Investigate the molecular basis for genetic instability and develop engineering strategies to create robust production organisms for large-scale fermentation; (6) Develop an end-to-end pipeline coupled with techno-economic analysis/life cycle analysis (TEA/LCA) to demonstrate the process economics for fermentative production of these target chemicals, which not only provides guidance to the metabolic engineering efforts in (3) and (5), but also integrates the Conversion theme with the Feedstock and Sustainability themes.

To achieve these objectives, researchers have assembled an interdisciplinary team with broad and complementary skills, including experts in metabolic engineering, synthetic biology, systems biology, enzyme engineering, analytical chemistry, catalysis, bioprocessing and fermentation, process economics analysis, data mining and machine learning, and plant synthetic biology.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0018420.

## **BIOENERGY RESEARCH CENTERS**

## Center for Bioenergy Innovation (CBI)

## Al-Informed Systems Biology: The Discovery of Cryptic Phenotypes and the Functional Networks that Control Them

John Lagergren<sup>1,6\*</sup> (lagergrenjh@ornl.gov), Mirko Pavicic<sup>1</sup>, Hari Chhetri<sup>1</sup>, Larry York<sup>1</sup>, Doug Hyatt<sup>2</sup>, David Kainer<sup>1</sup>, Erica Rutter<sup>3</sup>, Kevin Flores<sup>4</sup>, Jack Bailey-Bale<sup>5</sup>, Marie Klein<sup>5</sup>, Gail Taylor<sup>5</sup>, Jared Streich<sup>1</sup>, Daniel Jacobson<sup>1,6</sup>, **Gerald A. Tuskan<sup>1,6</sup>** 

<sup>1</sup>Oak Ridge National Laboratory; <sup>2</sup>University of Tennessee– Knoxville; <sup>3</sup>University of California–Merced; <sup>4</sup>North Carolina State University; <sup>5</sup>University of California–Davis; <sup>6</sup>Center for Bioenergy Innovation

### https://cbi.ornl.gov

**Goals:** The Center for Bioenergy Innovation (CBI) vision is to accelerate domestication of bioenergy-relevant, nonmodel plants and microbes to enable high-impact innovations along the bioenergy and bioproduct supply chain while focusing on sustainable aviation fuels (SAF). CBI has four overarching innovation targets: (1) Develop sustainable, process-advantaged biomass feedstocks, (2) Refine consolidated bioprocessing with cotreatment to create fermentation intermediates, (3) Advance lignin valorization for biobased products and aviation fuel feedstocks, and (4) Improve catalytic upgrading for SAF blendstocks certification.

This project enables fast and accurate automated imagebased plant phenotyping with minimal hand-annotated training data. Plant phenotyping is typically a timeconsuming and expensive endeavor, requiring large groups of researchers to meticulously measure biologically relevant plant traits, and is one of the main bottlenecks in understanding plant adaptation and the genetic architecture underlying complex traits at population scale. Here the team addresses these challenges by leveraging few-shot learning with convolutional neural networks to segment the leaf body and visible venation of 2,906 *P. trichocarpa* leaf images obtained in the CBI common garden located at University of California–Davis (see Fig. 2, p. 9). In contrast to previous methods, the approach: (1) does not require experimental or image pre-processing, (2) uses the raw RGB



Update vein region and generate new image tiles

**Fig. 2.** Region grown convolution neural network. [Reprinted by permission from Langergren, J., et al. 2023. Few-Shot Learning Enables Population-Scale Analysis of Leaf Traits in *Populus trichocarpa*," *arXiv* preprint arXiv:2301.10351. DOI:10.48550/arXiv.2301.10351]

images at full resolution, and (3) requires very few samples for training (e.g., just eight images for vein segmentation). Traits relating to leaf morphology and vein topology were extracted from the resulting segmentations using traditional image-processing tools and validated using real-world physical measurements.

To better understand the relationship among leaf phenotypes, a predictive phenomics network has been created from the leaf phenotypes with the use of iRF-LOOP, an explainable-AI-based network creation approach. Genomewide association studies (GWAS) have been performed on each leaf phenotype and network-based functional partitioning has been performed across the GWAS results to determine the shared and distinct functional interactions responsible for governing leaf traits.

In this way, the current work provides the plant community with (1) methods for fast and accurate image-based feature extraction that require minimal training data, (2) a new population-scale phenotype data set, including 68 different leaf phenotypes, (3) a new SNP dataset for 1,419 genotypes called against v4.1 of the *P. trichocarpa* genome, and (4) a unique view of the functional relationships governing these leaf phenotypes. All few-shot learning code, data, and results are publicly available. This is one of the largest single releases of new plant genotype and phenotype data [www.osti.gov/ dataexplorer/biblio/dataset/1846744].

Cliff, A., et al. 2019. "A High-Performance Computing Implementation of Iterative Random Forest for the Creation of Predictive Expression Networks," *Genes* **10**(12), 996. DOI:10.3390/ genes10120996.

Lagergren, J., et al. 2023. "Few-Shot Learning Enables Population-Scale Analysis of Leaf Traits in *Populus trichocarpa," arXiv* preprint arXiv:2301.10351. DOI:10.48550/arXiv.2301.10351.

Lagergren, J., et al. 2023. Supporting information for "Few-Shot Learning Enables Population-Scale Analysis of Leaf Traits in *Populus trichocarpa*," Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). Oak Ridge Leadership Computing Facility (OLCF). DOI:10.13139/ORNLNCCS/1908723.

**Funding Information:** Funding was provided by the Center for Bioenergy Innovation (CBI) led by Oak Ridge National Laboratory. CBI is funded as a U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science under FWP ERKP886. It also was supported by the Artificial Intelligence (AI) Initiative, an ORNL Laboratory Directed Research and Development program. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

## Comprehensive Genome-Wide CRISPR Interference Library for High-Throughput Functional Genomic Studies in *Pseudomonas putida*

Jacob Fenster<sup>1,2</sup>, Andrew Hren<sup>1,2</sup>, Margaret Spangler<sup>2,3</sup>, Bill Alexander<sup>2,4</sup>, Jerome M. Fox<sup>1</sup>, Jeffrey C. Cameron<sup>1,2</sup>, Carrie A. Eckert<sup>2,4\*</sup> (eckertca@ornl.gov), **Gerald A. Tuskan<sup>2,4</sup>** 

<sup>1</sup>University of Colorado–Boulder, CO; <sup>2</sup>Center for Bioenergy Innovation, Oak Ridge National Laboratory; <sup>3</sup>University of Tennessee–Knoxville; <sup>4</sup>Oak Ridge National Laboratory

### https://cbi.ornl.gov

**Goals:** The Center for Bioenergy Innovation (CBI) vision is to accelerate domestication of bioenergy-relevant, nonmodel plants and microbes to enable high-impact innovations along the bioenergy and bioproduct supply chain while focusing on sustainable aviation fuels (SAF). CBI has four overarching innovation targets: (1) Develop sustainable, process-advantaged biomass feedstocks, (2) Refine consolidated bioprocessing with cotreatment to create fermentation intermediates, (3) Advance lignin valorization for biobased products and aviation fuel feedstocks, and (4) Improve catalytic upgrading for SAF blendstocks certification.

The current paradigm for microbial engineering utilizes the design-build-test-learn cycle (DBTL), where production chassis are iteratively engineered based on previous findings toward the development of more robust production strains. Advanced genetic tools can accelerate this cycle, enabling the construction of strains with desirable phenotypes at a faster pace. With the advent of next-generation sequencing and CRISPR-Cas9-mediated genome editing, high-throughput genome-wide functional genomics experiments are now possible, enabling the study of hundreds of thousands of mutations in a single experiment. This work adapts existing knowledge gained using these highthroughput CRISPR-Cas9 technologies in model organisms to the promising nonmodel lignin-degrading soil bacterium Pseudomonas putida KT2440 to enable novel experimental approaches in this host for genotype-phenotype discovery for engineering efforts.

The team has optimized CRISPR-interference (CRISPRi) for *P. putida* by screening inducible promoter systems that express catalytically dead spCas9, a variant of Cas9 that can still associate with guide RNA (gRNA) and bind targeted sites in the promoter or 5' end of a gene, resulting in knockdown of gene expression. Researchers have quantified the dynamic range of repression by targeting a genomically integrated fluorescent reporter as well as key metabolic genes that compete with pathways to target products to increase titers (Fenster et al. 2022). In collaboration with the DOE Joint Genome Institute (JGI), team members have

generated genome-wide gRNA libraries to identify functional guides for gene editing and repression via CRISPRi to expand this system for genome-scale studies. For CRISPRi, researchers designed a 78,932-member library targeting each gene in the genome with 10-15 gRNAs per gene as well as 798 non-targeting gRNAs (1%) as internal controls. This library was transformed into P. putida and grown under varied growth conditions including glucose, acetate, and 10- and 50-mM p-coumaric acid, one of the major components of lignin hydrolysates that can be valorized by P. putida. Following selective growth, Illumina sequencing was performed where gRNA plasmids serve as barcodes and are used to compare pre- and post-selection to determine enrichment and drop out in the population to non-targeting gRNAs in the populations as controls. These data represent a wealth of new knowledge, currently uncovering essential genes (including a number that are of unknown function) as well as gene knockdowns that enrich or inhibit growth under each of the selective conditions. Validation of these identified targets will lead to discovery of new gene functions and optimization of gene expression in production strains to further expand engineering efforts and continue to accelerate the DBTL cycle in support of CBI research needs.

Fenster, J. A., et al. 2022. "Dynamic and Single Cell Characterization of a CRISPR-Interference Toolset in *Pseudomonas putida* KT2440 for  $\beta$ -ketoadipate Production from *p*-coumarate," *Metabolic Engineering Communications* **15**, e00204. DOI:10.1016/j.mec.2022.e00204.

**Funding Information:** Funding was provided by the Center for Bioenergy Innovation (CBI) led by Oak Ridge National Laboratory. CBI is funded as a U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science under FWP ERKP886. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

## Deciphering the Genetic Basis of S/G Lignin Variation in Switchgrass Through Quantitative Trait Locus Mapping and Transgenic Approaches

Jianxin Zhao<sup>1,4</sup>, Katrien M. Devos<sup>1,4\*</sup> (kdevos@uga.edu), Winnie Gimode<sup>1,4</sup>, Anne E. Harman-Ware<sup>2,4</sup>, Fang Chen<sup>3,4</sup>, Tom H. Pendergast IV<sup>1,4</sup>, **Gerald A. Tuskan<sup>4</sup>** 

<sup>1</sup>University of Georgia; <sup>2</sup>National Renewable Energy Laboratory <sup>3</sup>University of North Texas; <sup>4</sup>Center for Bioenergy Innovation

### https://cbi.ornl.gov

**Goals:** The Center for Bioenergy Innovation (CBI) vision is to accelerate domestication of bioenergy-relevant, nonmodel



**Fig. 3.** Colocalizing quantitative trait locus on chromosome 9N for lignin syringyl and guaiacyl ratio determined by pyMBMS and thioacidolysis. [Courtesy Tom H. Pendergast IV and Katrien M. Devos]

plants and microbes to enable high-impact innovations along the bioenergy and bioproduct supply chain while focusing on sustainable aviation fuels (SAF). CBI has four overarching innovation targets: (1) Develop sustainable, process-advantaged biomass feedstocks, (2) Refine consolidated bioprocessing with cotreatment to create fermentation intermediates, (3) Advance lignin valorization for biobased products and aviation fuel feedstocks, and (4) Improve catalytic upgrading for SAF blendstocks certification.

Switchgrass (*Panicum virgatum*) is being domesticated as a sustainable bioenergy crop due to its wide adaptability, high yield, and low agricultural inputs. Life-cycle analysis (LCA) has shown that, overall, feedstock yield is the main driver of fuel cost. When only considering the highest yielding accessions, however, biomass quality becomes an important player in biofuel yield and, hence, cost. Because the lignin syringyl and guaiacyl ratio may affect ethanol production as well as lignin monomer yields, researchers determined lignin monomeric composition using both pyrolysis molecular-beam mass spectrometry (pyMBMS) and thioacidolysis in an F2 population derived from a cross between the lowland genotype AP13 and the upland genotype VS16 (Qi et al. 2021). Quantitative trait locus (QTL) mapping for the S/G lignin ratio obtained using both methods identified colocalizing QTL on chromosome 9N (see Fig. 3, this page). The 9N QTL region harbors the genes PvBLH6, which encodes a BEL1-like homeodomain protein 6 transcription factor, and PvKNAT1, a member of the Arabidopsis TALE homeodomain transcription factor family. In Arabidopsis, BLH6 has been shown to interact with KNAT7 to affect secondary cell wall biosynthesis, including lignin content (Liu et al. 2014). Because lignin

content was affected in the *knat7* and *blh6 knat7* loss-offunction mutants but not in the *blh6* mutant, researchers transformed the AP13 (*PvKNAT1AP13*) and VS16 (*PvKNAT1VS16*) alleles of *PvKNAT1*, which differed by several nonsynonymous single nucleotide polymorphisms, into a *knat1*-null *Arabidopsis* mutant. *PvKNAT1VS16* but not *PvKNAT1AP13* rescued the phenotype of the *knat1* null mutant. These data suggest that only *PvKNAT1VS16* is functional. T3 transgenic *Arabidopsis* plants homozygous for the presence of *PvKNAT1VS16* and *PvKNAT1AP13* are being grown for assessment of the lignin S/G ratio. The study demonstrates how combining genetic mapping in switchgrass with transgenic analyses in *Arabidopsis* can help uncover critical variants in genes contributing to traits of importance to the bioeconomy.

Liu, Y., et al. 2014. "BEL1-LIKE HOMEODOMAIN6 and KNOT-TED ARABIDOPSIS THALIANA7 Interact and Regulate Secondary Cell Wall Formation via Repression of *REVOLUTA*," *Plant Cell* **26**, 4843–61. DOI:10.1105/tpc.114.128322.

Qi, P., et al. 2021. "Quantitative Trait Locus Mapping Combined with Variant and Transcriptome Analyses Identifies a Cluster of Gene Candidates Underlying the Variation in Leaf Wax Between Upland and Lowland Switchgrass Ecotypes," *Theoretical and Applied Genetics* **134**, 1957–75. DOI:10.1007/s00122-021-03798-y.

**Funding Information:** Funding was provided by the Center for Bioenergy Innovation (CBI) led by Oak Ridge National Laboratory. CBI is funded as a U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science under FWP ERKP886. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

## Metaproteomic Insights into Mechanisms Enabling Anaerobic, Thermophilic Microbiomes to Achieve Undiminished Fractional Carbohydrate Solubilization at High Solids Loading

Evert K. Holwerda<sup>1,2\*</sup> (evert.k.holwerda@dartmouth. edu), Payal Chirania<sup>1,3</sup>, Matthew R. Kubis<sup>1,2</sup>, Richard J. Giannone<sup>1,3</sup>, Yannick J. Bomble<sup>1,4</sup>, Robert L. Hettich<sup>1,3</sup>, Lee R. Lynd<sup>1,2</sup>, **Gerald A. Tuskan<sup>1,3</sup>** 

<sup>1</sup>Center for Bioenergy Innovation, Oak Ridge National Laboratory; <sup>2</sup>Dartmouth College; <sup>3</sup>Oak Ridge National Laboratory; <sup>4</sup>National Renewable Energy Laboratory

### https://cbi.ornl.gov

**Goals:** The Center for Bioenergy Innovation (CBI) vision is to accelerate domestication of bioenergy-relevant, nonmodel

plants and microbes to enable high-impact innovations along the bioenergy and bioproduct supply chain while focusing on sustainable aviation fuels (SAF). CBI has four overarching innovation targets: (1) Develop sustainable, process-advantaged biomass feedstocks, (2) Refine consolidated bioprocessing with cotreatment to create fermentation intermediates, (3) Advance lignin valorization for biobased products and aviation fuel feedstocks, and (4) Improve catalytic upgrading for SAF blendstocks certification.

Economically viable production of cellulosic biofuels requires (1) operation at high solids loadings of lignocellulosic feedstocks—on the order of 150 g/L, and (2) complete utilization of solubilized carbohydrates. Around two-thirds of the mass content of lignocellulose is carbohydrate. An efficient sugar-to-liquid-biofuel microbial metabolism can achieve an end-product yield of 0.5 g ethanol/g sugar. Not considering product titer restrictions and solids handling issues, 150 g/L solids loading would result in a maximum biofuel titer for ethanol of ~50 g/L. However, the recalcitrant character of lignocellulosic feedstocks impedes biological conversion and represents a major cost barrier. To this end, the team characterized Nature's ability to deconstruct and utilize lignocellulosic feedstocks at increasing solid loadings using defined cultures of anaerobic bacteria as well as anaerobic methanogenic microbiomes.

While the microbial community exhibits undiminished fractional carbohydrate solubilization near 0.7 at loadings ranging from 30g/L to 150g/L (Chirania et al. 2022), the defined culture shows decreasing solubilization at increasing solids loadings up to 80 g/L (Kubis et al. 2022). Note that fractional carbohydrate solubilization is defined as the portion of carbohydrate removed or solubilized per the original total amount of carbohydrate in a sample of biomass. The defined cultures reach high levels of solubilization they also leave behind small but distinct amounts of solubilized yet unutilized oligosaccharides. An ideal lignocellulose-to-biofuel process would employ characteristics of both these biocatalyst systems: the ability to (1) make a single liquid biofuel or intermediate as the metabolic end-product and (2) maintain high solubilization and utilization at high solids loadings.

To gain insight into the differences between solubilization and utilization of these two biological systems team members characterized microbiomes using metaproteomics, particularly focused on Carbohydrate Active enzymes (CAZymes) and diagnosed the defined cultures via fermentation-centered studies.

An anaerobic, thermophilic, semi-continuously fed, methanogenic microbial enrichment cultivated over an extended period (550 days), referred to as the lignocellulosefermenting microbiome, was sampled at various solids loadings at steady state. The samples were fractionated to identify key microbes and enzymes. Researchers documented changes in the abundance of CAZymes across fractions and the details of the methanogenesis pathways. Significant enrichment of auxiliary activity family 6 enzymes at higher solids suggests a role for Fenton chemistry.

Stress-response proteins accompanying these reactions are similarly upregulated at higher solids, as are  $\beta$ -glucosidases, xylosidases, carbohydrate-debranching, and pectin-acting enzymes—all of which indicate that removal of deconstruction inhibitors is important for observed undiminished solubilization.

The defined cultures reached solubilization levels as high as 75% on unpretreated feedstock in batch fermentations of corn stover and switchgrass. Cocultures consisting of cellulolytic and saccharolytic organisms showed increased solubilization and utilization over monocultures of cellulolytic bacteria. However, solubilization diminished at increasing solid loadings and this was not recovered via diagnostic experiments: the addition of a concentrated cell pellet to a running high solids fermentation did not increase solubilization compared to an equal volumetric addition of wateronly. The culture was also able to ferment additional pulses of model soluble (cellobiose) and insoluble (cellulose) substrates without issue, but this did not result in adverse or beneficial effects on the solubilization of the corn stover. Dilution with no-carbon source medium additions did partly recover the high solubilization characteristics, as did supplementation of the defined cultures with microbiome isolates or co-inocula.

Our work provides insights into the mechanisms by which natural microbiomes effectively deconstruct and utilize lignocellulose at high solids loadings and sets us on a path for engineering bacterial strains in the development of defined cultures for efficient bioconversion.

Chirania, P., et al. 2022. "Metaproteomics Reveals Enzymatic Strategies Deployed by Anaerobic Microbiomes to Maintain Lignocellulose Deconstruction at High Solids," *Nature Communications* **13**, 3870. DOI:10.1038/s41467-022-31433-x.

Kubis, M. R., et al. 2022. "Declining Carbohydrate Solubilization with Increasing Solids Loading During Fermentation of Cellulosic Feedstocks by *Clostridium thermocellum*: Documentation and Diagnostic Tests," *Biotechnology for Biofuels and Bioproducts* **15**(12). DOI:10.1186/s13068-022-02110-4.

**Funding Information:** Funding was provided by the Center for Bioenergy Innovation (CBI) led by Oak Ridge National Laboratory. CBI is funded as a U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science under FWP ERKP886. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract No. DE-AC05-00OR22725.

## *Populus* Secondary Cell Walls: Atomistic Models and Natural Variability in Architecture Informed by Solid-State Nuclear Magnetic Resonance

Bennett Addison<sup>1</sup>, Lintao Bu<sup>1</sup>, Vivek S. Bharadwaj<sup>1</sup>, Meagan Crowley<sup>1</sup>, Anne E. Harman-Ware<sup>1,2\*</sup>(anne.ware@nrel.gov), Mike Crowley<sup>1</sup>, Peter Ciesielski<sup>1</sup>, Yannick J. Bomble<sup>1</sup>, **Gerald A. Tuskan<sup>2</sup>** 

<sup>1</sup>National Renewable Energy Laboratory; <sup>2</sup>Center for Bioenergy Innovation

#### https://cbi.ornl.gov

**Goals:** The Center for Bioenergy Innovation (CBI) vision is to accelerate domestication of bioenergy-relevant, nonmodel plants and microbes to enable high-impact innovations along the bioenergy and bioproduct supply chain while focusing on sustainable aviation fuels (SAF). CBI has four overarching innovation targets: (1) Develop sustainable, process-advantaged biomass feedstocks, (2) Refine consolidated bioprocessing with cotreatment to create fermentation intermediates, (3) Advance lignin valorization for bio-based products and aviation fuel feedstocks, and (4) Improve catalytic upgrading for SAF blendstocks certification.

Interactions and higher-order architecture of the biopolymers present in plant secondary cell walls (SCW) are poorly understood. Solid-state nuclear magnetic resonance (ssNMR) measurements were used to infer refined details regarding the arrangements of the three major biopolymers in Populus cell walls: lignin, cellulose, and hemicellulose. Atomistic, macromolecular models representing multiple plant SCW arrangements were constructed in silico. Quantitative observables from molecular dynamics (MD) simulations (e.g., polymer-polymer distances, conformational analyses, quantified surface contact) were compared to ssNMR data to identify the most plausible polymeric arrangements. Results further understanding of plant SCWs and enable the development of hypotheses regarding superstructural architecture and resulting physicochemical properties of biomass cell walls. Additionally, researchers evaluated the genotypic variability of biopolymer spatial proximities in Populus SCW from a selection of natural variants exhibiting a range of compositional phenotypes. Coupling understanding of variability in SCW architectural characteristics with predictive atomistic models



**Fig. 4.** Solid-state nuclear magnetic resonance analysis of 13C-enriched poplar provides biopolymer architectural information that facilitates the construction of predictive atomistic models that will enable optimized design and deconstruction of sustainable biomass feedstocks. [Courtesy Bennett Addison, Lintao Bu, Vivek Bharadwaj Meagan Crowley, Anne E. Harman-Ware, Mike Crowley, Peter Ciesielski, Yannick J. Bomble, and Gerald A. Tuskan]

corroborated by ssNMR measurements will enable optimized design and deconstruction of sustainable biomass feedstocks. A high-level summary of the process workflow is presented in the Fig. 4, this page.

Addison, B., et al. 2020. "Selective One-Dimensional <sup>13</sup>C–<sup>13</sup>C Spin-Diffusion Solid-State Nuclear Magnetic Resonance Methods to Probe Spatial Arrangements in Biopolymers Including Plant Cell Walls, Peptides, and Spider Silk," *Journal of Physical Chemistry B* **124**(44), 9870–9883. DOI:10.1021/acs.jpcb.0c07759.

Funding Information: Funding was provided by the Center for Bioenergy Innovation (CBI) led by Oak Ridge National Laboratory. CBI is funded as a U.S. Department of Energy Bioenergy Research Centers supported by the Office of Biological and Environmental Research in the DOE Office of Science under FWP ERKP886. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This research also was supported by the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies Office (BETO), under award no. DE-AC36-08GO28308 with the National Renewable Energy Laboratory. The molecular modeling contributions to this work were performed as part of the Feedstock Conversion Interface Consortium (FCIC) with funding graciously provided by the U.S. Department of Energy Bioenergy Technologies Office.

## **BIOENERGY RESEARCH CENTERS**

## Great Lakes Bioenergy Research Center (GLBRC)

# High (School)-Throughput Screening of BAHD Transferases

**Justin F. Acheson**<sup>1</sup>, Ella Lodewyk<sup>1,2</sup>, Hailey Sieren<sup>1,2</sup>, Daniel Lee<sup>1,2</sup>, Melina Guestella<sup>1,2</sup>, Craig A. Bingman<sup>1,3</sup>, Rebecca A. Smith<sup>1,2</sup>, Steven D. Karlen<sup>1,3</sup>, John Ralph<sup>1,3</sup> Yucen Pu<sup>3,4</sup>, Zimou Sun<sup>3,4</sup>, Shawn D. Mansfield<sup>3,4</sup>, Brian G. Fox<sup>1,3</sup>

<sup>1</sup>Department of Biochemistry and <sup>2</sup>Dane County Youth Apprentice Program, <sup>3</sup>Great Lakes Bioenergy Research Center, University of Wisconsin—Madison; <sup>4</sup>University of British Columbia

Goals: BAHD acyltransferases represent a large family of enzymes typically found in plants. They use acyl-CoA donors (produced from acyl-CoA ligases) to form esters or amides with alcohol or amine acceptor molecules. The products of these reactions are incorporated into large polymers such as lignin and suberin or into small secondary metabolites including biofuel precursors, antimicrobials, antifungals, or compounds that contribute to drought resistance. The goal is to elucidate the identities and functions of these enzymes and use them in conjunction with acyl-CoA ligases, to precision engineer bioenergy crops. Pairing specific acyl-CoAs and BAHD transferases can allow fine tuning of lignin content either for simple deconstruction (Zip-lignin) or incorporation of useful aromatics that can then easily be clipped-off increasing the net value of the plant biomass (Karlen et al. 2016; de Vries et al. 2022).

BAHD acyltransferases have the ability to produce valuable molecules in bioenergy crops. The discovery and characterization of specific BAHD acyltransferases led to the creation of Zip-lignin, where introduction of ester-linked monolignols allows hydrolysis under mild conditions, avoiding harsher chemical treatments needed to remove lignin during bioenergy processing (Karlen et al. 2016). Further investigation showed that specific aromatics could be incorporated into terminal lignin positions, such as *p*-hydroxybenzoate that can easily be clipped off due to attachment via an ester linkage (de Vries et al. 2022). Thus, the ability to tune lignin composition not only allows for improved deconstruction, but also poises lignin as an attractive source of energy rich molecules. By taking advantage of consistently improving genomic data and tools the team curated lists of highpotential target genes focusing on two priority bioenergy

crops and a model plant (poplar, sorghum, Arabidopsis). Selected genes were synthesized into cell free expression vectors by the Joint Genome Institute (JGI) and were then screened by a team of high school student laboratory members using a wheat germ cell free system (Cell Free Sciences) . The expressed proteins were screened for potential activity and categorized by preferred substrates. Active enzymes catalyzing interesting reactions were then incorporated into Populus sp. to assess in vivo impacts, and cell-based expression systems such as Escherichia coli have been used to facilitate structural and biochemical characterization. The work presented here has given further understanding of the breadth of molecules this large family of enzymes can produce, and how these molecules may be useful in producing more energy efficient plants or provide engineered plant sources for fine chemicals.

Karlen, S. D., et al. 2016. "Monolignol Ferulate Conjugates are Naturally Incorporated into Plant Lignins," *Science Advances* **2**(10).

de Vries, L., et al. 2022. "pHBMT1, a BAHD-Family Monolignol Acyltransferase, Mediates Lignin Acylation in Poplar," *Plant Physiology* **188**(2).

Funding Information: This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under award no. DE-SC0020349 to B.G.F. C.A.B., R.A.S., S.D.K., S.D.M., and J.R. were also supported in part by the Great Lakes Bioenergy Research Center (GLBRC, DOE BER Office of Science DE-SC0018409). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817), GM/CA@APS has been funded by the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006, P30GM138396). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

### Defining Transcriptomic Dynamics in Sorghum in Multiple Abiotic Stresses

### Dae Kwan Ko<sup>1,2,3\*</sup> (dkko@msu.edu), **Federica** Brandizzi<sup>1,2,3</sup>, Timothy J. Donohue<sup>3,4</sup>

<sup>1</sup>MSU-DOE Plant Research Laboratory; <sup>2</sup>Michigan State University; <sup>3</sup>Great Lakes Bioenergy Research Center (GLBRC); <sup>4</sup>University of Wisconsin–Madison

### https://brandizzilab.natsci.msu.edu/

**Goals:** Increasing crops' resilience to the changing climate is critical to sustaining the bioeconomy on a large scale. Crop resilience is orchestrated by gene reprogramming events in response to environmental stresses. To better understand how gene expression changes are coordinated in sorghum, an important bioenergy crop, in response to climate stress, researchers performed a time-course transcriptome profiling under three major abiotic stress conditions followed by co-expression network analyses. These analyses provide new insights into the gene regulatory dynamics in response to stress but also fundamental resources for genetic engineering and molecular breeding.

The growth, development, and productivity of crops are challenged by abiotic stresses such as drought, heat, and salinity, whose severity is projected to increase steadily and irreversibly in the future. To survive and thrive in the environment, plants rapidly execute gene expression changes for necessary cellular and metabolic functions. Despite its significance in basic research and field applications, a comprehensive landscape of the abiotic stress-responsive transcriptome changes in bioenergy crops remains elusive to date. The use of dynamic time-course data on multiple abiotic stress-responsive transcriptomes in shoots and roots of sorghum (Sorghum bicolor L. Moench) with support of DOE Joint Genome Institute (JGI) allowed dissection of the complex stress-, tissue-, and phase-specific gene responses through a streamlined gene network modeling. The team established a series of co-expression modules of abiotic stress-responsive genes in shoots and roots, separately, where marker genes for various phytohormones are significantly enriched. The expression dynamics in the transcriptome data facilitated gene regulatory network mapping that identified potential candidate transcription factors (TFs) upstream of tissue-specific phytohormone network hub genes. The team proposes that in sorghum, the dynamic regulation of phytohormone marker genes in the abiotic stress-responsive co-expression network modules is coordinated by the master TFs in a tissue-specific manner. This knowledge can be used to boost bioenergy crop productivity and agricultural sustainability through genetic, chemical, and biotechnology engineering.

**Funding Information:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program under Award Numbers DE-SC0018409. The work conducted by the U.S. Department of Energy Joint Genome Institute (https://jgi.doe.gov/) is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

# *Novosphingobium aromaticivorans* for *cc*MA **Production from Lignin Biomass**

Avery C. Vilbert<sup>2\*</sup> (avilbert@wisc.edu), Wayne S. Kontur<sup>2</sup>, Derek M. Gille<sup>2</sup>, Daniel R. Noguera<sup>1,2</sup>, **Timothy J. Donohue<sup>1,2</sup>** 

<sup>1</sup>Great Lakes Bioenergy Research Center; and <sup>2</sup>University of Wisconsin–Madison

### https://www.glbrc.org/

**Goals:** This project will discuss *Novosphingobium aromaticivorans* as a bacterial host for production of the commodity chemical *cis,cis*-muconic acid (*cc*MA). It will provide further knowledge of the metabolism of aromatics by *N. aromaticivorans,* as well as a method to produce commodity chemicals from renewable carbon sources that are generated using green microbial engineering techniques.

Millions of tons of the commodity chemical *cis,cis*-muconic acid (ccMA) is produced annually from finite fossil fuel sources to produce nylon, polyesters, and other materials (Choi et al. 2020). The establishment of a sustainable bioeconomy hinges on the ability to use renewable sources for production of these in-demand chemicals. Lignin is an under-utilized abundant renewable resource that represents a potential carbon source for bio-based production of valuable chemicals. Due to the heterogeneity of lignin, it is challenging to extract commodity chemicals from lignin using current methods (Beckham et al. 2016). The team is working to funnel lignin streams into a single compound by leveraging the selectivity of enzymes with metabolic engineering techniques. N. aromaticivorans is an ideal candidate for lignin funneling as it is genetically tractable and has the native ability to catabolize lignin derived oligomers and metabolize multiple lignin monomers simultaneously. Previously, the lab was able to convert biomass aromatics into 2-pyrone-4,6-dicarboxylic acid from an engineered *N. aromaticivorans* strain (Perez et al. 2019). This research involves both the genetic engineering of N. aromaticivorans metabolic pathway for production of ccMA as well as the activity of enzymes along this metabolic pathway. In particular, the work highlights the

activity of putative *N. aromaticivorans* enzymes involved in decarboxylation of protocatechuic acid to catechol and oxidative ring cleavage of catechol to *cc*MA. This work directly compares the activity of the putative *N. aromaticivorans* enzymes *in vitro* and *in vivo* with other known homologous enzymes to engineer the most efficient *cc*MA production strain in *N. aromaticivorans*.

Beckham, G. T., et al. 2016. "Opportunities and Challenges in Biological Lignin Valorization," *Current Opinion in Biotechnology* **42**, 40–53.

Choi, S., et al. 2020. "Recent Advances in Microbial Production of Cis,cis-Muconic Acid," *Biomolecules* **10**, 1–14.

Perez, J. M., et al. 2019. "Funneling Aromatic Products of Chemically Depolymerized Lignin into 2-pyrone-4-6-dicarboxylic Acid with *Novosphingobium aromaticivorans*," *Green Chemistry* **21**, 1340–50.

**Funding Information:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program under Award Number DE-SC0018409.

## Structure and Molecular Mechanism of a Bacterial ADP-forming 4-Coumarate CoA Ligase

Debayan Chaudhury<sup>1\*</sup> (dchaudhury@wisc.edu), Justin F. Acheson<sup>1</sup>, Ella R. Torkelson<sup>1</sup>, Kaya A. Meyers<sup>1</sup>, Yucen Pu<sup>3</sup>, Zimou Sun<sup>3</sup>, Marco Tonelli<sup>1</sup>, Craig A. Bingman<sup>1,2</sup>, Rebecca A. Smith<sup>1,2</sup>, Steven D. Karlen<sup>1,2</sup>, John Ralph<sup>1,2</sup>, Shawn D. Mansfield<sup>3</sup>, **Brian G. Fox<sup>1,2</sup>** 

<sup>1</sup>University of Wisconsin—Madison; <sup>2</sup>Great Lakes Bioenergy Research Center; <sup>3</sup>University of British Columbia

**Goals:** Acyl-CoA ligases are enzymes that catalyze the adenosine triphosphate (ATP)-dependent conjugation of carboxylic acids to coenzyme-A (CoA) and enable the entry of these acids into metabolism as activated CoA thioesters. These CoA thioesters are important metabolic intermediates that feed carbon from the organic acids into a variety of anabolic and catabolic pathways allowing for the construction/ destruction of diverse molecular scaffolds, thereby enabling cellular function. The aim of this project is the identification and biochemical characterization of acyl-CoA ligases from plant and microbial sources with the end goal of enabling precision metabolic engineering of the important bioenergy crop *Populus trichocarpa* (black cottonwood) with a focus on lignification.

4-Coumarate CoA ligases (4CLs) are a subset of acid-thiol ligases (E.C. 6.2.1.-) that catalyze the ATP-dependent thioesterification of 4-coumarate to CoA. In plants, these

enzymes act upstream in the phenylpropanoid pathway and are therefore attractive targets for metabolic engineering as they provide control of metabolic flux over the downstream reactions that are rich in valuable and useful metabolites (Vogt 2010). 4CLs are also present in microbes where they function in aromatic degradation pathways which proceed through CoA thioester. The researchers report the structure and mechanism of a bacterial 4CL (ferA) from the ferulic degradation cluster of the lignin degrading bacterium Sphingomonas SYK-6 (Masai et al. 2002). This is the first crystal structure report of a conformation of bacterial 4-coumarate ligase. Catalysis in ferA proceeds via a reversible central metabolism-like ADP-forming mechanism through an acyl-phosphate intermediate instead of the AMP-forming mechanism widely reported for plant 4CLs, which proceeds through an acyl-adenylate intermediate. The co-crystal structure of ferA (2.46 Å, Rwork 22%, R<sub>free</sub> 25%) crystallized in the pre-ATP hydrolysis conformation bound to inorganic phosphate, feruloyl-CoA, and non-hydrolysable ATP analogue AMPPNP offers insight into the catalytic machinery and all the ligand binding sites on the ferA enzyme, and sheds light on the molecular interactions that enable enzyme specificity and catalysis. Researchers combine these structural insights with optical and NMR spectroscopic studies of the forward and reverse ligase reaction to elucidate the role of protein conformation and its intricate links to ligand binding and group transfer catalysis.

Masai, E., et al. 2002. "Cloning and Characterization of the Ferulic Acid Catabolic Genes of *Sphingomonas paucimobilis* SYK-6," *Applied and Environmental Microbiology* **68**(9), 4416–24.

Vogt, T., 2010. "Phenyl<br/>propanoid Biosynthesis," Molecular Plant ${\bf 3}(1),$  2–20.

**Funding Information:** This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under award no. DE-SC0020349 to B.G.F. C.A.B., R.A.S., S.D.K., S.D.M., and J.R. were also supported in part by the Great Lakes Bioenergy Research Center (GLBRC, DOE BER Office of Science DE-SC0018409). This study made use of the MS and NMR facilities at the GLBRC, and the National Magnetic Resonance Facility at Madison which is supported by NIH grant R24GM141526.

### The Genetics of Pathogen and Microbiome Control in the Switchgrass Leaf

Acer Van Wallendael<sup>1,2\*</sup> (vanwall1@msu.edu), Gian Benucci<sup>1,2</sup>, P. B. da Costa<sup>1,2</sup>, L. Fraser<sup>1,2</sup>, Felix Fritschi<sup>3</sup>, Thomas E. Juenger<sup>4</sup>, John T. Lovell<sup>5</sup>, Gregory Bonito<sup>1,2</sup>, **David B. Lowry<sup>1,2</sup>** 

<sup>1</sup>Michigan State University; <sup>2</sup>Great Lakes Bioenergy Research Center; <sup>3</sup>University of Missouri; and <sup>4</sup>University of Texas; <sup>5</sup>HudsonAlpha Center for biotechnology

**Goals:** This study aims to uncover genetic regions responsible for governing switchgrass responses to both pathogenic and mutualist fungi, and to understand how these responses may be coordinated.

Leaf fungal microbes can be fundamental drivers of host plant success, as they consist of pathogens that devastate crop plants as well as taxa that enhance nutrient uptake, discourage herbivory, and antagonize pathogens. In a replicated diversity panel of biofuel switchgrass, researchers quantified genetic and environmental variation in leaf fungal relationships, both for the whole microbiome and for a specific pathogen, leaf rust. While fungal colonization of the leaf varies over space and time, researchers uncovered genome-wide associations (GWAs) with several informative loci. In particular, three cysteine-rich receptor-like kinase genes (crRLKs) were linked to a genetic locus associated with microbiome structure. Since each of these genes is consistently upregulated in switchgrass genotypes typically more susceptible to fungal disease, researchers conclude that they may play a central role in the plant's response to pathogens. Response to leaf rust is polygenic and environmentally sensitive, but resistance alleles are associated with higher biomass, indicating that breeding for rust-resistant plants will benefit growth without trade-offs in the absence of rust. Switchgrass response to fungal colonists is complex and variable, but an experimental design that accounts for variation over space and time allows for greater definition on genetic loci underlying fungal interactions.

VanWallendael, 2022. "Host Genotype Controls Ecological Change in the Leaf Fungal Microbiome," *PLoS Biology* **20**(8), e3001681.

VanWallendael, 2020. "Geographic Variation in the Genetic Basis of Resistance to Leaf Rust between Locally Adapted Ecotypes of the Biofuel Crop Switchgrass (*Panicum virgatum*)," *New Phytologist* **227**(6), 1696–1708.

**Funding Information:** The plant material is based upon work supported in part by the Great Lakes Bioenergy Research Center (glbrc.org), U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409. Support for this research was provided by the National Science Foundation Long-term Ecological Research Program (DEB 1832042; https://beta.nsf.gov/funding/ opportunities/long-term-ecological-research-lter) at the Kellogg Biological Station and by Michigan State University AgBioResearch. Researchers received further funding from the U.S. Department of Energy (energy.gov) through grants DE-SC0014156 to TEJ and DE-SC0017883 to DBL and National Science PGRP Award IOS1402393 to J.T.L. The work conducted by the U.S. Department of Energy Joint Genome Institute (https://jgi.doe.gov/) is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## Towards Generalized Platforms for Functional Genomics in α-Proteobacteria

Ashley N. Hall<sup>1</sup> (aroarty@uw.edu), Kyle Kinney<sup>1</sup>, Amy B. Banta<sup>1,2</sup>, Timothy Donohue<sup>1,2</sup>, **Jason M. Peters<sup>1,2</sup>** 

<sup>1</sup>DOE Great Lakes Bioenergy Research Center; <sup>2</sup>University of Wisconsin–Madison

**Goals:** Enable rational engineering of alphaproteobacteria that can synthesize biofuels and bioproducts from lignocellulosic biomass. This project develops and validates genomic insertion sites, inducible promoters, and inducible riboswitches for bacteria such as *Novosphingobium aromaticivorans, Zymomonas mobilis,* and *Rhodobacter sphaeroides*.

Numerous alphaproteobacterial species have promising traits for converting lignocellulosic biomass to useful biofuels and bioproducts. However, many of the genes required to carry out these roles have remained enigmatic in part due to limited genetic tools developed for these organisms. Next-generation genetic tools, such as CRISPRi-seq, are capable of systematically phenotyping all genes, but they have not been broadly deployed in alphaproteobacteria. Building on success in establishing genome-scale CRIS-PRi in Z. mobilis, the team seeks to develop generalized platforms for synthetic biology and functional genomics in alphaproteobacteria. Here, researchers lay the groundwork for these platforms by optimizing site specific integration and developing synthetic, inducible promoters for N. aromaticivorans and R. sphaeroides. These enhanced genetic tools will enable basic and applied research such as delivery of CRISPRi systems to investigate gene function and expression of heterologous pathways to generate valuable bioproducts.

**Funding Information:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number DE-SC0018409.

# BIOENERGY RESEARCH CENTERS Joint BioEnergy Institute (JBEI)

## Characterizing Lignocellulose Breakdown Mechanisms in Anaerobic Gut Fungi

Shiyan Jin<sup>1\*</sup> (ShiyanJin@lbl.gov), Stephen P. Lillington<sup>1</sup>, Kai Deng<sup>2</sup>, Thomas S. Lankiewicz<sup>1</sup>, Trent R. Northen<sup>2,3</sup>, Christopher J. Petzold<sup>2,3</sup>, Steven Singer<sup>2,3</sup>, Blake A. Simmons<sup>2,3</sup>, Michelle O'Malley<sup>1</sup>, **Jay D. Keasling<sup>2,3</sup>** 

<sup>1</sup>University of California–Santa Barbara; <sup>2</sup>Joint BioEnergy Institute; <sup>3</sup>Lawrence Berkeley National Laboratory

**Goals:** The vision of Joint Bioenergy Institute (JBEI) is that bioenergy crops can be converted into economically viable, carbon-neutral, biofuels and renewable chemicals currently derived from petroleum, and many other bioproducts that cannot be efficiently produced from petroleum.

Lignocellulose is an attractive feedstock for renewable chemical manufacturing and bio-based chemical production, which reduces dependency on petroleum, but its recalcitrant structure requires multiple catalytic steps for sufficient hydrolysis, rendering current production technologies energetically and economically expensive. Anaerobic gut fungi (AGF) provide an attractive alternative strategy for biomass valorization by secreting a variety of carbohydrate-active enzymes (CAZymes) to efficiently degrade unpretreated biomass (Solomon et al. 2016). Many CAZymes are colocalized in fungal cellulosomes (enzyme complexes) that are thought to accelerate lignocellulose breakdown compared to the action of freely diffusing enzymes (Lillington et al. 2021). Anaerobic gut fungi control cellulosome compositions by regulating the production of biomass-degrading enzymes based on fungal life stage and the complexity of available substrates, as supported by advanced microscopy imaging and transcriptomic characterization (Solomon et al. 2016; Lillington et al. 2021). To further assess the differences in mechanistic properties among cellulosomes, researchers first developed an isolation technique to harvest these complexes from fungi grown on aqueous cellulose (cellobiose), fibrous cellulose (Whatman filter paper) and lignocellulose (reed canary grass), followed by protein purification via fast protein liquid chromatography. Researchers used nanostructure-initiator mass spectrometry (NIMS) probes, which quantify oligosaccharide abundance in solution, to measure cellulases and hemicellulase activity based on the hydrolysis of NIMS probes (Deng et al. 2014). Amorphous cellulose probes, crystalline cellulose probes

and hemicellulose probes were used to detect any preference in degradation among enzyme complexes. Researchers observed that cellulosomes produced when growing fungi on lignocellulose were the most active and digested all three NIMS probes at a comparable level. Additionally, cellulosomes secreted when grown on filter paper, even though they perform better in cellulose digestion than that grown on cellobiose, exhibited a strong preference toward cellulose over hemicellulose. Currently, researchers are screening a library of 200+ synthesized genes from fungal cellulosomes to unmask their functions and roles in lignocellulose degradation via a combination of proteomics and NIMS characterization.

Additionally, there is evidence from two-dimensional heteronuclear single quantum coherence nuclear magnetic resonance (2D-HSQC-NMR) spectrometry, which suggests an up to 8% reduction in  $\beta$ -aryl ether linkage, a typical lignin bond, in addition to the changes in lignin monomer composition after incubating biomass (sorghum, switchgrass, and poplar) with Neocallimastix californiae and Anaeromyces robustus, two AGF strains. The same phenomenon is observed in lignin content analysis and gel permeable chromatography. This is the first time that lignin modification is observed with high confidence under an anaerobic environment. Researchers are working on expanding the biomass library to include different types of softwood and hardwood, to develop an increased mechanistic understanding of anaerobic lignin modification in plants with different lignin compositions. Moreover, researchers are developing and applying lignin-based probes to characterize possible mechanisms of anaerobic lignin breakdown.

Deng, K., et al. 2014. "Rapid Kinetic Characterization of Glycosyl Hydrolases Based on Oxime Derivatization and Nanostructure-Initiator Mass Spectrometry (NIMS),"ACS Chemical Biology **9**(7), 1470–79.

Lillington, S. P., et al. 2021. "Cellulosome Localization Patterns Vary Across Life Stages of Anaerobic Fungi," *mBio* **12**(3), e00832-00821.

Solomon, K. V., et al. 2016. "Early-Branching Gut Fungi Possess a Large, Comprehensive Array of Biomass-Degrading Enzymes," *Science* **351**(6278), 1192–95.

**Funding Information:** 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.

## Evolutionary Engineering is a Versatile Strain Optimization Approach for Sustainable Bioproduction

Adam M. Feist<sup>1,2,3\*</sup> (afeist@ucsd.edu), Jay D. Keasling<sup>2,3,4,5</sup>

<sup>1</sup>University of California—San Diego; <sup>2</sup>Technical University of Denmark; <sup>3</sup>DOE Joint Bioenergy Institute; <sup>4</sup>Lawrence Berkeley National Laboratory; <sup>5</sup>University of California—Berkeley

### http://jbei.org

**Goals:** The vision of Joint Bioenergy Institute (JBEI) is that bioenergy crops can be converted into economically viable, carbon-neutral, biofuels and renewable chemicals currently derived from petroleum, and many other bioproducts that cannot be efficiently produced from petroleum.

Strain engineering of microbes for bioproduction is a challenge and requires multiple optimization approaches and engineering cycles to reach commercial viability. An approach that is gaining in utility in the Design-Build-Test-Learn (DBTL) cycle is adaptive laboratory evolution (ALE) as it can uniquely solve problems encountered in the stain engineering process using selection and the natural ability of microbes to adapt. A platform to effectively utilize ALE built around custom automation, process control software, and bioinformatics pipelines was developed and it has been effectively applied to engineer a range of microorganisms. Specific use cases demonstrating how ALE can be used in different implementations of the DBTL cycle will be presented (Sandberg et al. 2019). First, an implementation where ALE is used after the Build step but before the Test step to engineer *Pseudomonas putida* to utilize non-native hemicellulose monomers is presented (Lim et al. 2021). Second, an implementation where aggregated ALE generated mutations in the ALE database (ALEdb) are utilized in the Design step to introduce novel mutations for enhanced glycerol uptake is described (Phaneuf et al. 2021). The third implementation will describe how ALE can replace the Design and Build steps to generate a strain with enhanced secretion and tolerance to L-serine in one DBTL cycle to contribute to a commercially viable production strain (Mundhada et al. 2017). Finally, what is likely achievable for ALE and laboratory automation in the short term and how it can be broadly applied to solve more problems in industrial bioproduction is presented.

Lim, H. G., et al. 2021. "Generation of *Pseudomonas putida* KT2440 Strains with Efficient Utilization of Xylose and Galactose via Adaptive Laboratory Evolution," *ACS Sustainable Chemistry & Engineering* **9**(34), 11512–523.

Mundhada H., et al. 2017. "Increased Production of L-serine in *Escherichia coli* through Adaptive Laboratory Evolution," *Metabolic Engineering* **39**, 141–50. Phaneuf, P. V., et al. 2021. "*Escherichia coli* Data-Driven Strain Design Using Aggregated Adaptive Laboratory Evolution Mutational Data," *ACS Synthetic Biology* **10**(12), 3379–95. DOI:10.1021/ acssynbio.1c00337.

Sandberg, T. E., et al. 2019. "The Emergence of Adaptive Laboratory Evolution as an Efficient Tool for Biological Discovery and Industrial Biotechnology," *Metabolic Engineering* **56**, 1–16. DOI:10.1016/j.ymben.2019.08.004.

**Funding Information:** 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.

### Expression of S-Adenosyl Methionine Hydrolase Modifies Lignin in Sorghum

Aymerick Eudes<sup>1,2\*</sup> (ageudes@lbl.gov), Yang Tian<sup>1,2</sup>, Edward Baidoo<sup>1,2</sup>, Yu Gao<sup>1,2</sup>, Hemant Choudhary<sup>1,3</sup>, John M. Gladden<sup>1,3</sup>, Blake A. Simmons<sup>1,2</sup>, Jenny C. Mortimer<sup>1,2,4</sup>, Henrik Scheller<sup>1,2,5</sup>, **Jay D. Keasling<sup>1,2,5</sup>** 

<sup>1</sup>DOE Joint BioEnergy Institute; <sup>2</sup>Lawrence Berkeley National Laboratory; <sup>3</sup>Sandia National Laboratories; <sup>4</sup>University of Adelaide, Australia; <sup>5</sup>University of California–Berkeley

https://www.jbei.org/

**Goals:** The vision of Joint Bioenergy Institute (JBEI) is that bioenergy crops can be converted into economically viable, carbon-neutral, biofuels and renewable chemicals currently derived from petroleum, and many other bioproducts that cannot be efficiently produced from petroleum.

Plant biomass represents a large renewable source of fermentable sugars for the synthesis of bioproducts. These sugars are stored as cell wall polymers, mainly cellulose and hemicellulose, and are embedded with lignin, which makes their enzymatic hydrolysis challenging. One of the strategies to reduce cell wall recalcitrance is the modification of lignin content and composition. Lignin is a phenolic polymer of methylated aromatic alcohols and its synthesis in tissues developing secondary cell walls is a significant sink for the consumption of the methyl donor S-adenosylmethionine (AdoMet) due to the involvement of methyltransferases. Researchers previously demonstrated in Arabidopsis that specific expression of S-adenosylmethionine hydrolase (AdoMetase, E.C. 3.3.1.2) in stem tissues reduces AdoMet content and impacts lignin biosynthesis (Eudes et al. 2016). This engineering approach was tested in the bioenergy crop sorghum (Sorghum bicolor L.). AdoMetase was expressed in sorghum lignifying tissues using the promoter of the caffeic acid O-methyltransferase gene. Both AdoMet content

and lignin are reduced in transgenics. 2D-HSQC NMR analysis of cell walls showed relative enrichment of nonmethylated *p*-hydroxycinnamyl (H) units and a reduction of tricin (T), guaiacyl (G) and syringyl (S) lignin units in transgenics. Gel permeation chromatography revealed differences in the number average molecular weight (Mn) and weight averaged molecular weight (Mw) of lignin in AdoMetase lines. Quantification of cell wall-bound hydroxycinnamates showed a reduction of ferulate in all transgenic lines. These modifications in engineered sorghum result in a diminution of cell wall recalcitrance since higher yields of glucose and xylose were obtained after enzymatic saccharification of biomass compared to wild type plants. Considering that some transgenic lines display no important diminution of biomass yields, this engineering approach provides a valuable option for the improvement of lignocellulosic biomass feedstock.

Eudes, A., et al. 2016. "Expression of S-adenosylmethionine Hydrolase in Tissues Synthesizing Secondary Cell Walls Alters Specific Methylated Cell Wall Fractions and Improves Biomass Digestibility," *Frontiers in Bioengineering and Biotechnology* **4**(58). DOI:10.3389/fbioe.2016.00058.

**Funding Information:** This work was part of the DOE Joint BioEnergy Institute supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

### Scenario-Based Technoeconomic and Life-Cycle Analyses of Biomass Conversion Technologies

Blake A. Simmons<sup>1\*</sup> (basimmons@lbl.gov), Nawa Baral<sup>1</sup>, Aindrila Mukhopadhyay<sup>1</sup>, Steven Singer<sup>1</sup>, Taek Soon Lee<sup>1</sup>, John M. Gladden<sup>2</sup>, Alberto Rodriguez<sup>2</sup>, Hemant Choudhary<sup>2</sup>, Henrik Scheller<sup>1</sup>, Patrick Shih<sup>3</sup>, Corinne D. Scown<sup>1</sup>, **Jay D. Keasling**<sup>1,2</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory; <sup>2</sup>Sandia National Laboratories; <sup>3</sup>University of California–Berkeley

### http://www.jbei.org

**Goals:** The vision of Joint Bioenergy Institute (JBEI) is that bioenergy crops can be converted into economically viable, carbon-neutral, biofuels and renewable chemicals currently derived from petroleum, and many other bioproducts that cannot be efficiently produced from petroleum.

Cellulosic biofuels have not yet reached cost parity with conventional petroleum fuels (See Fig. 5, this page). One of the central approaches used at JBEI to identify promising biofuels and bioproducts, and the conversion technologies



**Fig. 5.** Minimum required selling price ranges for bioproducts (\$/kg) under different *in planta* accumulation amount (dry basis) in order to reach the minimum ethanol selling price (MESP) parity (\$3.61/gal) and targeted selling price of ethanol (\$2.50/gal). The inset shows the estimated bioproduct selling price of less than \$100/kg (Yang et al. 2020). [Reprinted under a Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CC BY-NC-ND) from Yang M., et al. 2020. "Accumulation of High-Value Bioproducts *In Planta* Can Improve the Economics of Advanced Biofuels," *PNAS* **117**(15) 8639–648. DOI:10.1073/pnas.2000053117]

capable of producing them, is combining technoeconomic and life-cycle analysis. At JBEI, researchers have developed scenario-based methodologies to conduct this work. Three of these scenarios are: (1) engineering bioenergy crops to generate value-added bioproducts in planta can reduce input requirements relative to microbial chassis and skip costly deconstruction and conversion steps (Yang et al. 2022; Yang et al. 2020), (2) presents detailed process configurations for the bioadvantaged sustainable aviation fuel, dimethycyclooctane (DMCO) production to estimate the minimum selling price and life-cycle greenhouse gas (GHG) footprint considering three different hydrogenation catalysts and two bioconversion pathways (Baral et al. 2021), and (3) evaluation of protic and aprotic ionic liquids for the deconstruction and conversion of mixed woody biomass feedstocks using a one-pot configuration (Achinivu et al. 2022). This work provides new insights into the tradeoffs, challenges, and opportunities present in the advanced biomass conversion technologies being developed at JBEI realized at the commercial scale.

Achinivu, E. C., et al. 2022. "In Situ Synthesis of Protic Ionic Liquids for Biomass Pretreatment," ACS Sustainable Chemistry & Engineering **10**(37), 12090–98. DOI:10.1021/acssuschemeng.2c01211.

Baral, N. R., et al. 2021. "Production Cost and Carbon Footprint of Biomass-Derived Dimethylcyclooctane as a High-Performance Jet Fuel Blendstock," *ACS Sustainable Chemistry & Engineering* **9**(35), 11872–82. DOI:10.1021/acssuschemeng.1c03772. Yang, M., et al. 2020. "Accumulation of High-Value Bioproducts in Planta can Improve the Economics of Advanced Biofuels," *Proceedings of the National Academy of Sciences* **117**(15), 8639–48. DOI:10.1073/pnas.2000053117.

Yang, M., et al. 2022. "Comparing *in Planta* Accumulation with Microbial Routes to Set Targets for a Cost-Competitive Bioeconomy," *Proceedings of the National Academy of Sciences* **119**(30), e2122309119. DOI:10.1073/pnas.2122309119.

**Funding Information:** This work was part of the DOE Joint BioEnergy Institute (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.

## Understanding the Role of Lignin in Native Architecture of Engineered Plant Cell Walls via Multi-Dimensional Solid-State Nuclear Magnetic Resonance

Yu Gao<sup>1,2\*</sup> (yugao@lbl.gov), Andrew Lipton<sup>3</sup>, Dylan Murray<sup>4</sup>, Aymerick Eudes<sup>1,2</sup>, Henrik Scheller<sup>1,2,5</sup>, Jenny C. Mortimer<sup>1,2,6</sup>, **Jay D. Keasling<sup>1,2,5</sup>** 

<sup>1</sup>DOE Joint BioEnergy Institute; <sup>2</sup>Lawrence Berkeley National Laboratory; <sup>3</sup>Environmental Molecular Sciences Laboratory; <sup>4</sup>University of California– Davis; <sup>5</sup>University of California; <sup>6</sup>University of Adelaide, Australia

### http://www.jbei.org

**Goals:** The vision of Joint Bioenergy Institute (JBEI) is that bioenergy crops can be converted into economically viable, carbon-neutral, biofuels and renewable chemicals currently derived from petroleum, and many other bioproducts that cannot be efficiently produced from petroleum.

As a major component (~30%) of the plant secondary cell wall, lignin is a promising renewable feedstock for the production of platform chemicals due to its high aromaticity. However, the heterogeneity of the lignin structure leads to biomass recalcitrance, which significantly impedes the efficiency of total biomass conversion in a biorefinery context. Researchers propose that understanding how lignin interacts with the other major cell wall components (i.e., cellulose and hemicellulose) and contributes to the 3D cell wall nanoarchitecture will provide invaluable insights into the nature of biomass recalcitrance. This, in turn, will support the successful engineering of bioenergy crops with optimized biomass, which can be fully deconstructed and valorized into biobased products. These studies employ multi-dimensional solid-state nuclear magnetic resonance (ssNMR), including one-dimensional cross-polarization and direct polarization, two-dimensional refocused Incredible Natural Abundance Double Quantum Transfer Experiment (INADEQUATE) and Proton Driven Spin Diffusion (PDSD), and spin-lattice relaxation time measurements, to monitor the native polymer arrangements in the intact secondary cell walls of engineered plants with reduced recalcitrance traits. Here researchers will present data on the arrangement of cell wall polymers in model and crop species with reduced lignin. Researchers use this information to understand factors underlying cell wall properties, and to support predictive cell wall design and biomass deconstruction.

Gao, Y., et al. 2020. "A Grass-Specific Cellulose–Xylan Interaction Dominates in Sorghum Secondary Cell Walls," *Nature Communications* **11**, 6081. DOI:10.1038/s41467-020-19837-z.

Gao, Y., et al. 2023. "Elongated Galactan Side-Chains Mediate Cellulose-Pectin Interactions in Engineered *Arabidopsis* Secondary Cell Walls," *The Plant Journal* **115**(2), 529–45. DOI:10.1111/ tpj.16242.

**Funding Information:** This work was conducted as part of the DOE Joint BioEnergy Institute (www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. Part of this work was conducted by the Environmental Molecular Sciences Laboratory (grid.436923.9), a DOE Office of Science scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at PNNL under contract DE-AC05-76RL01830.

## **PLANT GENOMICS**

## University Projects

## Analysis of the Beneficial Associations of Sorghum with Arbuscular Mycorrhizal Fungi Studied with Genetics, Genomics, and Microbiomics

Jeff Bennetzen<sup>1\*,</sup> Phil Brailey-Jones<sup>1</sup>, Tom H. Pendergast IV<sup>1</sup>, Shufan Zhang<sup>1</sup>, Beatrice Bock<sup>2</sup>, Amanda Bouffier-Landrum<sup>1</sup>, Jia Hwei Cheong<sup>1</sup>, Xiomy Pincha Davila<sup>1</sup>, Josue Fernandez<sup>1</sup>, Ching-Ting Huang<sup>1</sup>, Ben Long<sup>1</sup>, Sarah Mondibrown<sup>1</sup>, Rachel Rackers<sup>1</sup>, Sedona Spann<sup>2</sup>, Isaac Torres<sup>1</sup>, Richard Trippe<sup>2</sup>, Jonathan Arnold<sup>1</sup>, Anny Chung<sup>1</sup>, Katrien M. Devos<sup>1</sup>, Nancy Collins Johnson<sup>2</sup>

<sup>1</sup>University of Georgia; <sup>2</sup>Northern Arizona University

**Goals:** This project is designed to investigate the interactions between sorghum genes, arbuscular mycorrhizal fungi (AMF), the sorghum root-associated microbiome, and numerous environmental factors that contribute to sorghum biomass production. A key focus is on host genetics, so the team is investigating 337 Bioenergy Association Panel accessions to perform a full genome-wide association study (GWAS) analysis.

In first-year pilot studies, researchers determined that very different AMF populations were associated with sorghum roots at two field locations in Georgia and Arizona. Root types and developmental stages were also identified that were consistent for AMF abundances, infection structures, and percent colonization, thereby decreasing the number of sample extractions that will be needed in all future experiments. These first-year studies also indicated that a good deal of microbiome analysis could be performed most effectively by shotgun sequence analysis, rather than amplicon analysis.

In second-year studies of input effects, performed at a Georgia field location, the team determined efficient/ reproducible methods for sample production/collection, and generated the full set of 4044 samples (337 genotypes × three replications × four treatments) to determine the effects of phosphate and nitrogen levels on all of the investigated field/sample properties in a randomized complete block design. These measured properties included quantification and type characterization of AMF and other microbes on or within the sorghum roots, expression of eukaryotic genes on or within the sorghum roots, microscopic analysis to classify and quantitate AMF infection structures, and agronomic sorghum traits, such as plant height, plant biomass, plant tillering, plant flowering time, and plant mineral content. Most of these analyses are still in the data generation stage, but researchers did observe a trend of high nitrogen increasing tiller number, but high phosphate decreasing plant height, potentially indicating disruption of mycorrhizal associations at the Georgia field site.

The team has also made major technical advances, including creating and training an automated imaging platform to allow high-throughput classification and segmentation of AMF structures, such as arbuscules and vesicles, which in turn provided the size, density, and percent colonization of AMF structures in field-grown sorghum roots for GWAS analysis. Over 100 fungal morphotypes have been cultured from sorghum roots in Georgia, including three AMF species that are being propagated on sorghum as pure cultures. In addition, 100+ endophytic bacteria and fungi have been isolated and cultured from first-year studies in Arizona, and about half have been shown to exhibit drought tolerance by a polyethylene glycol assay. These cultured microbes will be used for planned greenhouse tests of hypotheses generated from descriptive field data from the major GWAS fields in Arizona and Georgia.

### Early Career

## Defining the Influence of Environmental Stress on Bioenergy Feedstocks at Single-Cell Resolution

Margot Bezrutczyk<sup>1</sup>, Danielle Goudeau<sup>1</sup>, Thai Dao<sup>2\*</sup>, Dusan Velickovic<sup>3</sup>, Robert Stanley<sup>3</sup>, Christopher Anderton<sup>3</sup>, Rex Malmstrom<sup>1</sup>, **Benjamin Cole**<sup>1</sup>

<sup>1</sup>DOE Joint Genome Institute; <sup>2</sup>Oregon State University; <sup>3</sup>Pacific Northwest National Laboratory

Biomass derived from plant feedstocks is a renewable and sustainable energy resource, but these resources are vulnerable to environmental stress such as water and nutrient limitations. Understanding how cells work independently and in concert to regulate plant responses to stress will be crucial to improving their performance. The Early Career Research Project awarded to this group aims to apply several cutting-edge single-cell and spatially resolved transcriptome sequencing approaches to construct a comprehensive single-cell resource for plants and to better understand the complexity behind stress response among diverse cell types. To this end, researchers have profiled thousands of individual sorghum root cells grown under normal and phosphate-limited conditions. Preliminary results have indicated several genes whose expression is potentially altered by stress in a cell type–specific way, with genes in the vasculature being particularly affected. The team is currently integrating this nascent data with additional single-cell data from other species, including maize, *Brachypodium* and switchgrass. Researchers are also characterizing environmental stress using other advanced profiling methods, including spatial transcriptomics and spatial metabolomics. The team hopes to build a multispecies model of cell type–specific stress responses that can be tested under agriculturally relevant conditions using the EcoPOD technology at Lawrence Berkeley National Laboratory, and eventually use to develop new targeted intervention strategies.

## Elucidating the Genetic Components of the Physiological and Metabolic Processes Governed by the TORC Regulatory Module in Poplar

Reuben Tayengwa<sup>1</sup>, Yiping Qi<sup>1</sup>, Edward Eisenstein<sup>1</sup>, Victor Busov<sup>2</sup>, Hairong Wei<sup>2</sup>, **Gary Coleman<sup>1\*</sup>** 

#### <sup>1</sup>University of Maryland; <sup>2</sup>Michigan Technological University

**Goals:** The goal of this research is to identify the target of rapamycin (TOR)complex-1-mediated regulatory and signaling pathways and the linkage between TOR and Rho of Plants (ROP) GTPase nutrient sensing in poplar to elucidate the functional role of genes regulating nutritional responses. Through experiments using CRISPR gene editing, genomics, biochemical, and computational approaches, the project will decipher the functional significance of poplar TOR coupled to ROP GTPase nutrient sensing and signaling. This will allow the team to elucidate function and divergence in this nutrient signaling relay allowing for identification of the regulatory and signaling networks and hubs.

Understanding the genetic basis of complex physiological and metabolic traits, including resource use efficiency and abiotic stress, is necessary for improving lignocellulose crops for a sustainable biobased economy. Poplar (*Populus* spp.) is an important and sustainable bioenergy and bioproduct plant feedstock, yet understanding of the pathways and networks governing resource use efficiency and responses to abiotic stresses is poorly developed. Nutrient sensing and signaling is a fundamental mechanism which modulates cellular activities that mediate growth, development, and biomass accrual. The protein TOR kinase is part of an evolutionally conserved central hub that integrates not only nutrient signals but also signals related to energy, hormones, biotic, and abiotic stresses through its pivotal role in regulating transcription, translation, and metabolism. In plants, TOR signaling of diverse nitrogen signals has been linked to

signal integration by ROP GTPases that modulate TOR activation and signaling outputs. Although the ROP GTPase/ TOR signaling relay has been shown to be important to integrating nitrogen signals, little is known about these signaling components in poplar. Interestingly, compared to other plants species, such as *Arabidopsis* and rice, poplar contains two TOR genes, and researchers have discovered that the two poplar TOR genes can be gene edited to vary gene dosage resulting in viable plants with discernible phenotypes. Since in other plant species TOR knockouts or nulls are embryo lethal, researchers have a unique opportunity to employ genome editing to determine if the two poplar TOR genes have functionally diverged along with establishing the role of the TOR/ROP GTPase relay in nitrogen sensing and signaling.

The overall objective of this research is to identify the TOR-mediated regulatory and signaling pathways and the linkages between TOR and ROP GTPase nutrient sensing in poplar to elucidate the functional roles of these genes in regulating nutritional responses. Through experiments that use CRISPR gene editing, genomics, biochemical, and computational approaches, the project will decipher the functional role and divergence of the two poplar TOR genes by altering gene dosage and identifying the functional linkage between ROP GTPase signal integration and TOR activity. The specific objectives of the project are to (1) characterize carbon and nitrogen mediated TOR activation in poplar cells; (2) alter gene dosage of the two poplar TOR genes to determine their functional roles and divergence in carbon and nitrogen signaling; (3) determine which members of the small GTPase ROP gene family can integrate diverse nitrogen signals to activate TOR; (4) identify how the regulatory and signaling networks and associated regulatory factors downstream of the ROP-GTPase-TOR nutrient sensing and signaling relay have diverged; and (5) validate the function of genetic factors downstream of the ROP GTPase/ TORC relay in nitrogen sensing and signaling.

The results of this project will allow the team to elucidate function and divergence in this nutrient signaling relay allowing for identification of the regulatory and signaling networks and hubs involved in nitrogen responses in poplar. Although this research is focused on nitrogen nutrition, it is likely that the results of this project will also uncover aspects of how TOR and ROP GTPases modulate other processes such as abiotic and biotic stress, and developmental processes regulating biomass yield.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0023011.

### Early Career

## Understanding the Effects of *Populus*— Mycorrhizal Associations on Plant Productivity and Resistance to Abiotic Stress

**Melissa A. Cregger1**\* (creggerma@ornl.gov), María del Rosario Ramírez-Flores<sup>1</sup>, Alyssa A. Carrell<sup>1</sup>, Spencer Roth<sup>1</sup>, David J. Weston<sup>1</sup>, Dawn M. Klingeman<sup>1</sup>, Miranda Clark<sup>1</sup>, Sara S. Jawdy<sup>1</sup>, Dana L. Carper<sup>1</sup>, Gail Taylor<sup>2</sup>, Jamie McBrien<sup>1</sup>, Leah Burdick<sup>1</sup>, Ann Wymore<sup>1</sup>, David McLennan<sup>1</sup>

#### <sup>1</sup>Oak Ridge National Laboratory; <sup>2</sup>University of California–Davis

**Goals:** The overarching goal of this project is to create sustainable, multipurpose bioeconomies whereby globally important feedstocks can be produced while simultaneously maximizing soil health and mitigating adverse impacts of climatic change. In this project, the unique ability of *Populus* species will be leveraged to associate with both ectomycorrhizal (ECM) and arbuscular mycorrhizal (AM) fungi to examine how variation in these associations alters plant productivity, abiotic stress response, and belowground soil carbon cycling.

Within the myriad of possible plant-microbe interactions occurring belowground, plant-mycorrhizal associations are widespread with the two most common mycorrhizal types being ECM and AM. Belowground plant interactions with these fungi have been shown to increase water uptake and nutrient acquisition and alter soil carbon storage. It is unclear how these two dominant mycorrhizal fungal types differ in their abilities to offer benefits to the plant and change belowground carbon and nutrient cycling. Most plants associate with one type of mycorrhizal fungi, and individual plant species are less likely to associate with both AM and ECM fungal species. Populus species, however, uniquely associate with AM and ECM simultaneously in natural settings, thus providing an ideal experimental system for examining how mycorrhizal fungal types confer benefits to their host. The team will take advantage of highthroughput plant phenotyping, greenhouse, and field experiments to characterize how variation in Populus-mycorrhizal associations alters the plant response to drought, and researchers will manipulate plant-mycorrhizal interactions to influence plant productivity, increase plant drought tolerance, and enhance soil health.

In objective one of this project, researchers identified drought tolerant and drought susceptible genotypes of *Populus trichocarpa* and *Populus deltoides* x *P. trichocarpa* hybrids. In a series of replicated greenhouse experiments, researchers grew 37 genotypes of *P. trichocarpa* and 29 unique hybrid genotypes (*P. trichocarpa* x *P. deltoides* and *P. deltoides* x *P. trichocarpa*)

in double autoclaved potting mix under well-watered and drought conditions. Before manipulating water availability, base line plant phenotypes (e.g., plant height, stomatal conductance, leaf chlorophyll, leaf protein) were measured. Next, the team initiated an acute drought on half of the plants and monitored soil volumetric water content over the course of 1 week. When plants began to wilt significantly, hyperspectral images were captured from leaf four using a Headwall camera from 900 to 2500 nm wavelength to identify early indications of drought tolerance in images. Further, researchers characterized water-use efficiency, leaf protein, leaf chlorophyll, and changes in plant height, leaf number, and above/below ground biomass. Overall, the team found significant variation in plant phenotype across genotypes and in response to acute drought. Populus genotypes that varied in drought tolerance will be used in upcoming manipulative experiments with mycorrhizal fungi. This work will be expanded to identify P. deltoides genotypes that vary in drought tolerance.

Within objective two, researchers will characterize variation in mycorrhizal community composition, colonization, and abundance across drought tolerant/susceptible *Populus* species/genotypes. In February and July of 2022, root and rhizosphere soil samples were collected from drought tolerant and susceptible *P. trichocarpa* in a genome-wide association study (GWAS) plantation in Davis, Calif. Across these genotypes, the team found that both AM and ECM fungi colonized the roots, and drought tolerant genotypes had a greater percentage of hyphae, greater number of arbuscules, and a larger hartig net compared to drought susceptible trees. Amplicon and metatranscriptomic sequencing are in progress to characterize the AM and ECM taxa across these trees. Further, culturing of these unique organisms is underway to be used in manipulative experiments.

Combined, these initial results highlight significant genetic variation in the response of *Populus* to drought when grown without microbial symbionts, and further demonstrates variation in belowground mycorrhizal communities across drought tolerant and susceptible genotypes. Plant and fungal resources resulting from these experiments will be used to evaluate how these differences drive changes in host abiotic stress tolerance and soil carbon cycling.

**Funding Information:** This work was supported by the U.S. Department of Energy Office of Science, through the Biological and Environmental Research (BER) Program's Early Career Research Program. The *P. trichocarpa* GWAS plantation in Davis, California was developed and is maintained through the Center for Bioenergy Innovation, Bioenergy Research Center funded by BER.

# Predicting Gene Functions in Plants with Single-Cell Genomic Data

Prakash Raj Timilsena<sup>1\*</sup>, Sai Deepak Gattidi<sup>1</sup>, **José R. Dinneny**<sup>2</sup>, Song Li<sup>1</sup>

#### <sup>1</sup>Virginia Polytechnic Institute and State University; <sup>2</sup>Stanford University

The rapid sequencing of genomes and transcriptomes in bioenergy crops and other plant species has outpaced the rate at which gene functions can be accurately annotated. Wet-bench validation for gene functions is very laborious and time consuming for nonmodel species. Even in a model organism like Arabidopsis thaliana, majority of the gene functions have not been validated with wet lab experiments. Traditional computational methods for assigning gene functions largely rely on sequence homology which could not account for gene expression activities in different tissue or cell types. In this project, researchers tested whether single-cell gene expression data can be used to improve the gene function annotation. The team compared bulk- and single-cell RNA seq datasets from roots and assessed the performance of seven machine learning algorithms for predicting gene functions. Researchers found that random forest works the best among these methods. The team further asked whether single-cell genomic data can provide additional information because expression data from more diverse cell populations are captured by single-cell (sc) RNA-seq as compared to bulk RNA-seq. Surprisingly, bulk RNA-seq were found to have better accuracy in predicting many gene functions as compared to scRNA-seq data. A comparison of scRNA-seq datasets from different tissues showed that leaf scRNA-seq data provides higher accuracy in predicting the chloroplast and photosynthesis related genes as compared to root scRNA-seq data. This observation suggests that the specificity of the information content in single-cell datasets from different tissues is biologically relevant. Because of the diversity of cell types captured by scRNA-seq, research found that an increasing number of Uniform Manifold Approximation and Projection clusters may help to improve the prediction accuracy for single-cell data. The future direction of this work is to incorporate stress responsive scRNA-seq data and regulatory networks (DAP-seq) information to further expand the prediction of novel gene functions in oil seed crops. Experimental validation for selected genes will be performed in the coming years.

**Funding Information:** This project is supported by DOE-BER, DE-SC0020358, and DE-SC0022985.

### Extracting Functional Traits from Large Volumes of Field Phenomics Data

Emmanuel Miguel Gonzalez<sup>1\*</sup> (emmanuelgonzalez@ arizona.edu), Ariyan Zarei<sup>1</sup>, Jeffrey Demieville<sup>1</sup>, Nimet Beyza Bozdag<sup>1</sup>, Seungho Lee<sup>1</sup>, Sarthak Bawal<sup>1</sup>, Jordan Edward Pettiford<sup>1</sup>, Yuguo Xiao<sup>2</sup>, Indrajit Kumar<sup>2</sup>, Maxwell Braud<sup>2</sup>, Eric Lyons<sup>1</sup>, Duke Pauli<sup>1</sup>, **Andrea L. Eveland**<sup>2</sup>

<sup>1</sup>University of Arizona; <sup>2</sup>Donald Danforth Plant Science Center

### https://github.com/phytooracle

https://datacommons.cyverse.org/browse/iplant/home/shared/ phytooracle

#### Goals:

- Use the University of Arizona's Field Scanner to collect high spatial and temporal resolution field phenomics data on 430 ethyl methanesulfonate (EMS)-mutagenized families in the BTx623 background under well-watered and water limited conditions.
- Develop software and machine learning (ML) models to extract fine-scale phenotypic trait data at individual plant and organ levels from field phenomics data.
- Leverage fine-scale phenotypic trait data to study genotype-phenotype associations in response to drought to facilitate discovery of genes and their functions.

Studying dynamic plant responses to environmental conditions has historically been difficult due to the low throughput and long-term cost of longitudinal data collection in the field setting (Reynolds et al. 2019). Recent technological advances have resulted in small, low-cost, and high-resolution sensors that can be used to rapidly collect phenotypic trait data at regular time intervals in field or greenhouse settings (Li et al. 2020; Sooriyapathirana et al. 2021). Today, high spatial and temporal resolution field phenomics data is being collected to extract information on dynamic plant responses to biotic and abiotic stress under real world field conditions. When phenotypic trait data from multiple sensors are combined, a multidimensional understanding of plant morphology and physiology can facilitate the discovery of genes and their functions. The University of Arizona is home to the world's largest outdoor plant phenotyping system, the Field Scanner that encompasses numerous sensors for collecting plant phenotypic trait data. The Field Scanner collects red-green-blue (RGB), photosystem II (PSII) chlorophyll fluorescence, and thermal images as well as 3D point clouds using laser scanners. The Field Scanner raw data is being processed using PhytoOracle (PO), a series of scalable, modular phenomic data processing pipelines (Gonzalez et al. 2022). The processed data generated by PO is enabling the extraction of increasingly fine-scale



**Fig. 6.** Sorghum individual plant phenotyping workflow. **(A)** Individual plants are isolated from raw field phenomics data. **(B)** Machine learning (ML) models segment soil from plant points and classify distinct leaves. **(C)** Various leaf and whole-plant phenotypes are extracted, including volume, topological data analysis (TDA) shape descriptors, and leaf descriptors. [Courtesy University of Arizona: Duke Pauli, Emmanuel Gonzalez, Jeffrey Demieville, Ariyan Zarei]

traits from various levels, from the field to plot and whole plant to individual organs. To extend phenotyping capabilities, novel machine learning (ML) algorithms are being developed to extract multidimensional phenomics datasets that can be mined for genotype-phenotype associations related to abiotic stress.

ML models that aim to segment plant point clouds can provide fine-scale phenotypic data on plant morphology at individual plant and organ levels (See Fig. 6, this page). Various traditional shape descriptors can be extracted, including height, volume, and angle. Additionally, topological data analysis (TDA), a mathematical framework for studying the underlying connections of points and their properties, can be used to study shape nuances that may not be captured by traditional shape descriptors. Persistence diagrams and Euler characteristic curves are common TDA methods, which aim to capture topological signatures that summarize shape features from which shape nuances can be studied (Amézquita et al. 2022; Amézquita 2020; Chazal and Michel 2021). A variety of traditional and TDA shape descriptors are being collected from 430 EMS-mutagenized sorghum families in the BTx623 background under well-watered and water-limited conditions across the life cycle of plants. These shape descriptors are being leveraged to identify genotypephenotype associations related to drought stress to enable gene discovery. This work will identify induced variation in drought resilient traits for enhancing the productivity of bioenergy crops under drought conditions through fine-scale phenotyping. This information will drive the breeding and

engineering of improved, climate-resilient varieties capable of maintaining productivity under limited resources.

Amézquita, E. J., et al. 2020. "The Shape of Things to Come: Topological Data Analysis and Biology, from Molecules to Organisms," *Developmental Dynamics* **249**, 816–833.

Amézquita, E. J., et al. 2022. "Measuring Hidden Phenotype: Quantifying the Shape of Barley Seeds Using the Euler Characteristic Transform," *In Silico Plants* **4**, diab033. DOI:10.1093/insilicoplants/ diab033.

Chazal, F., and B. Michel. 2021. "An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists," *Frontiers in Artificial Intelligence* **4**.

Gonzalez, E., et al. 2022. PhytoOracle Automation (Version 1.0.0). [https://github.com/phytooracle/automation]

Li, B., et al. 2020. "Phenomics-Based GWAS Analysis Reveals the Genetic Architecture for Drought Resistance in Cotton," *Plant Biotechnology Journal* **18**, 2533–44.

Reynolds, D., et al. "What is Cost-Efficient Phenotyping? Optimizing Costs for Different Scenarios," *Plant Science* **282**, 14–22.

Sooriyapathirana, S. D. S. S. et al. 2021. "Photosynthetic Phenomics of Field- and Greenhouse-Grown Amaranths vs. Sensory and Species Delimits," *Plant Phenomics* **2021**, 1–13.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy Biological and Environmental Research Award Number DE-SC0020401, the U.S. Department of Energy Advanced Research Projects Agency-Energy OPEN Award Number DE-AR0001101,
and the National Science Foundation CyVerse Award Number DBI-1743442.

## Harnessing Regulatory Variation to Elucidate Drought Resilience Mechanisms in Sorghum

Yuguo Xiao<sup>1</sup>, Indrajit Kumar<sup>1</sup>, Maxwell Braud<sup>1</sup>, Philip Ozersky<sup>1</sup>, Emmanuel Miguel Gonzalez<sup>2</sup>, Rajdeep S. Khangura<sup>3</sup>, Brian Dilkes<sup>3</sup>, Duke Pauli<sup>2</sup>, Todd C. Mockler<sup>1</sup>, **Andrea L. Eveland**<sup>1\*</sup> (aeveland@danforthcenter.org)

<sup>1</sup>Donald Danforth Plant Science Center; <sup>2</sup>University of Arizona; <sup>3</sup>Purdue University

#### Goals:

- Overall project objective: To define and functionally characterize genes and pathways related to drought stress tolerance in sorghum and the molecular mechanisms by which these factors drive phenotypic diversity.
- Establish a foundation for deep explorations of gene regulatory networks in sorghum through integrative genomics analyses.
- Enhance understanding of how genotype drives phenotype and environmental adaptation using high-resolution, field-based phenotyping of sorghum mutant collections and a novel diversity panel.
- Map and characterize genes contributing to drought responsive phenotypes in sorghum.
- Experimentally validate predictions of gene function using molecular and genetic assays and targeted gene editing.

Development of the next generation of bioenergy feedstocks will require strategies that utilize resource-limited agricultural lands, including the introduction of novel traits into crops to increase abiotic stress tolerance. This project investigates the innate drought resilience of sorghum (*Sorghum bicolor*), a bioenergy feedstock and cereal crop. Drought is a complex trait and identifying the genes underlying sorghum's innate drought tolerance and how they are regulated in the broader context of the whole plant and its environment requires advanced approaches in genetics, genomics, and phenotyping.

This project leverages a field-based phenotyping infrastructure at Maricopa, Ariz., which provides an exceptional capability for managed stress trials in a hot and arid environment through controlled irrigation. An automated field scanner system collects high-resolution phenotyping data using a variety of sensors throughout the growing season, from seedling establishment to harvest. A sorghum mutant population was phenotyped under the field scanner to compare drought-stressed and well-watered plants. Each mutant's genome has been sequenced so that sequence variants can be linked with phenotypes. Being able to assess the genotype-to-phenotype link in response to drought over the life cycle of the plant will facilitate discovery of genes and their functions. The team has also constructed a diverse panel of sorghum lines, which maximizes variation in wateruse efficiencies as well as genetics and geographic origin. Each of these lines have also been sequenced through various efforts. This population was phenotyped in a controlled environment drought response experiment where samples were collected for population-level expression analyses and will subsequently be field-phenotyped under the field scanner this summer. State-of-the-art phenotyping data analytics pipelines have been developed as part of this project and DOE-funded initiatives and are being extended to define stress-related phenotypes at multiple scales. Bulked segregant analysis-seq is used to accelerate mapping of causal loci that underlie mutants of interest. So far, researchers have identified candidate genes underlying defects in leaf senescence, plant architecture, and male fertility. Regulatory maps generated from diverse sorghum lines in response to stress are being used to nominate gene candidates and place them in the larger context of a drought response network. In addition, gene editing and transgenic methods are being used to characterize gene function.

This work will identify control points for enhancing the productivity of bioenergy crops in marginal environments through precision breeding or engineering, and thus accelerate the development of improved varieties that are high yielding with limited water resources.

**Funding Information:** This work is funded by DOE BER awards #DE-SC0023305 and #DE-SC0020401.

## Systems Framework to Enhance the Potential of *Camelina* as Oilseed Crop

Suresh Kumar Gupta, Emily Pawlowski, Jordan Brock, Fabio Gomez-Cano, Adrian Platts, Patrick Edger, **Erich Grotewold**<sup>\*</sup> (grotewol@msu.edu)

Michigan State University

https://camregbase.org/

**Goals:** The project will (1) characterize the genetic variation, gene expression and chromatin accessibility across *Camelina* varieties and growth conditions, and (2) develop the tools to understand and manipulate *Camelina* gene expression. An important goal is to identify key genes and genomic regions to target in breeding efforts to enhance productivity, while providing the research community with a number of tools to

understand and manipulate *Camelina* gene expression. An important objective of this project is to make available to the community a new set of tools and resources for *Camelina* that have been limited in the past to model plant systems.

The adoption of *Camelina sativa* as an industrial oilseed crop hinges on being able to increase its modest yield. This is in part constrained by the limited knowledge of the gene regulatory networks responsible for plant growth and responses to the environment, and by a poor understanding of the genetic diversity and gene content across Camelina accessions. To address these shortcomings, researchers have started the development of a Camelina transcription factor (TF) open reading frame (ORF) collection (TFome) that will accelerate the discovery of protein-DNA interactions. For example, Camelina TF ORFs were used for carrying out DNA affinity purification with high-throughput sequencing (DAP-seq) towards the identification of candidate fatty acid regulators (Gomez-Cano et al. 2022). Researchers have standardized conditions for embryo and seed Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) to identify accessible chromatin regions and comparing them between the three subgenomes that comprise the hexaploid genome of C. sativa. Researchers are also in the process of completing the sequencing of a new version of the Camelina variety Suneson. Finally, researchers are standardizing conditions for the growth and transformation of the diploid and tetraploid precursors of hexaploid varieties, to facilitate crop resynthesis and introduction of novel genetic diversity.

Gomez-Cano, F., et al. 2022. "Exploring *Camelina sativa* Lipid Metabolism Regulation by Combining Gene Co-Expression and DNA Affinity Purification Analyses," *The Plant Journal* **110**, 589– 606. DOI:10.1111/tpj.15682.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022987.

## Development of a CenH3-Based Haploid Inducer in Hexaploid *Camelina sativa*

**Isabelle M. Henry**<sup>\*</sup> (imhenry@ucdavis.edu), Paul Osuna-Kleist, Helen Tsai, Luca Comai

University of California–Davis

https://www.montana.edu/econproject/index.html

**Goals:** *Camelina sativa* is a promising oilseed crop that is particularly well suited for cultivation in the Northwest of the United States. The Enhancing Camelina Oilseed Production with Minimal Nitrogen Fertilization in Sustainable Cropping Systems (ECON) project is an interdisciplinary project focused on two main objectives (1) enhance nitrogen utilization efficiency and (2) boost oil yield. The long-term goals are to increase the economic profitability of Camelina cultivation, by reducing the negative impact of nitrogen fertilization and increasing productivity competitiveness with other major oilseed crops such as canola. Approaches include characterizing genetic and genomic natural variation within Camelina for the ability to absorb, translocate and assimilate nitrogen, and for recruiting beneficial rhizo-microbes to improve nitrogen acquisition. Researchers are also investigating the mechanisms underlying these differences to optimize yield potential by increasing seed size and enhancing oil synthesis. Within the ECON project, the laboratory aims at developing a haploid inducer line for Camelina, a tool that will be instrumental for accelerating the breeding of several loci of interest in a polyploid background.

Haploid induction is powerful breeding tool. Among others, it allows the rapid production of complex genotypes when constructing experimental and breeding lines. This is particularly critical when dealing with polyploid genomes such allohexaploid C. sativa. For example, selfing a parent with six heterozygous non-linked loci is expected to result in 0.024% homozygous progeny. Crossing the same parent to a haploid inducer (HI) will produce 1.5% progeny with the desired genotype. Modification of the centromere-specific histone variant CENH3 engenders haploid induction in Arabidopsis thaliana. Specifically, crosses between a haploid inducer line carrying a mutated form of CENH3 and a wild-type line results in frequent elimination of the haploid inducer chromosomes and produces offspring of different types in similar numbers: paternal haploids, aneuploids, and diploids. Haploids are formed when all the HI chromosomes are lost and only the maternal WT chromosomes are retained.

The goal is to develop a cenH3-based haploid inducer in C. sativa. C. sativa is a very close relative of A. thaliana, but the situation is complicated by the fact that the genome of Camelina harbors three functional copies of the cenH3 genes, all of which need to be modified to produce a HI. A TILLING population of Camelina var. Ames 1043 was previously developed in the laboratory and more than 300 high-reliability mutations were identified in the cenH3 genes. Through a series of crosses and selection steps, the following three mutations were combined into a single line, in the homozygous state: a nonsense mutation allele (genome 1), a missense mutation that is known to result in haploid induction in A. thaliana (genome 2), and a splicesite variant (genome 3). The resulting potential haploid inducer was crosses to WT Ames and the progeny screened for the presence of haploids. No haploid plant was recovered but >85% of the progeny lacked at least one chromosome, confirming that genome elimination is occurring in these crosses, albeit not sufficiently efficiently to eliminate all 20 chromosomes of the haploid inducer parent. Interestingly,

**Bioenergy** | Plant Genomics

While previously considered a byproduct of growth, the

chromosomes from genome 2 were preferentially lost in the aneuploid progeny, consistent with previously documented relative expression dominance of the other two subgenomes within the *Camelina* genome. Taken together, the results so far suggest that CENH3-based haploid induction is a feasible approach in *Camelina* but complicated by the fact that a large percentage of aneuploid progeny survive, presumably because of the buffering effect of the polyploid background. Researchers are in the process of combining weaker CENH3 alleles in a new potential haploid inducer line in order to further increase the loss of HI chromosomes and hopefully obtain fully haploid lines.

**Funding Information:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science program Award No. DE-SC0021369.

#### Early Career

## Plant Methanol Emission at the Interface of the Photosynthetic C1 Pathway, Leaf Water Status, and Growth

Suman Som<sup>1</sup>, Luiza Gallo<sup>1,3</sup>, Tomas Domigues<sup>2</sup>, Edward Baidoo<sup>3</sup>, Aymerick Eudes,<sup>4</sup> **Kolby J.** Jardine<sup>1\*</sup> (kjjardine@lbl.gov)

<sup>1</sup>Lawrence Berkeley National Laboratory; <sup>2</sup> Joint BioEnergy Institute; <sup>3</sup>University of São Paulo

http://cellwallesters.pbworks.com/w/page/127623629/ FrontPage

Goals: The Poplar Esterified Cell Wall Transformations and metabolic INtegration (PECTIN) project aims to study the metabolism of cell wall ester modifications and volatile intermediates, and their role in central physiological processes in the emerging biofuel species California poplar (Populus trichocarpa). A key goal of this research is to evaluate abiotic stress responses in plants with modified expression patterns of key genes involved in cell wall metabolism with altered amounts of methyl and acetyl groups present on cell walls. These genetic modifications will be evaluated for potential impacts on plant hydraulics, physiology, and stress responses. Understanding and manipulating the metabolism of cell wall modifications will not only provide important knowledge on the physiology and ecology of plants but will also allow the generation of engineered bioenergy crops such as poplar for sustainable production of biofuels and bioproducts, addressing BER's goal of developing renewable bioenergy resources.

release of methanol from methylated pectin in primary cell walls of plants dramatically alters their elasticity, a critical parameter controlling initiation and propagation of tissue morphogenesis and growth. Given the large reservoir of methylated pectin in leaves and other plant tissues associated with the primary cell wall, leaf methanol emissions are assumed to derive from this large, stored carbon reserve with no apparent direct connection to photosynthesis. Plant methanol emissions are assumed to derive from lightindependent temperature-driven growth process, and therefore assumed to have no direct metabolic connection with photosynthesis. In this study, researchers use <sup>13</sup>CO<sub>2</sub>-labeling to demonstrate that methyl esters on primary cell walls of leaves of C<sub>3</sub> plants are directly produced from photosynthetically linked C1 metabolism, not related to photorespiration within minutes of light exposure through a proposed series of intracellular and extracellular connected pathways. This occurs in parallel with methanol release from the primary cell wall during temperature-stimulated growth processes, which are constrained by midday leaf water stress. Upon illumination of individual leaves and branches, <sup>13</sup>C/<sup>12</sup>C-methanol emission ratios continuously increased during photosynthesis under an elevated <sup>13</sup>CO<sub>2</sub> atmosphere (500-1000 ppm). Dynamic branch <sup>13</sup>CO<sub>2</sub> labeling of photosynthesis lasting 2.5 days showed daily increases of <sup>13</sup>C/<sup>12</sup>C-methanol emission ratios peaking at the end of the light period with <sup>13</sup>C-methanol emissions gradually increasing at the expense of <sup>12</sup>C-methanol, reaching <sup>13</sup>C/<sup>12</sup>C-methanol up to 50%. In the dark, branch  ${}^{13}C/{}^{12}C$ -methanol emission ratios remained constant despite strong night-time <sup>13</sup>C-methanol emission dynamics that mimicked <sup>12</sup>C-methanol emissions. At midnight, when the leaf water potential recovered (-0.2 + / - 0.1 MPa) from midday values (-1.0 + / - 0.1 MPa), branch emissions of both 12C-methanol and 13C-methanol increased steadily throughout the night, despite leaf temperature and transpiration slowly decreasing. The results are consistent with a distinction between biosynthesis and incorporation of photosynthetically derived C1 carbon into leaf primary cell walls and methanol production during growth. An accelerated growth phase occurs between midnight and midday, where growth and methanol increased positively with temperature, and a decelerated growth phase between midday and midnight, where growth and methanol emissions decreased with temperature, constrained by leaf water stress. The results provide evidence for a rapid and direct connection between photosynthesis and photorespirationindependent C1 metabolism; a temperature independent hydraulic control over methanol emissions and growth rates at night; and a temperature stimulated (morning), followed by inhibited (afternoon) methanol emission and growth during the day.

The observations are consistent with a biochemical model integrating CO<sub>2</sub> fixation by the Calvin cycle with the  $C_1$  pathway involving: (1) Photosynthesis, (2) The phosphorylated serine pathway, which synthesizes the donor methyl group of methionine in the chloroplast (3) Export of methionine to the cytosol followed by activation to S-adenosylmethionine (AdoMet), which is imported into many organelles and used to transfer methyl groups to polysaccharides, nucleic acids, proteins, lipids, and secondary metabolites, (4) Methyl esterification of new pectin monomers in the Golgi with AdoMet, (5) Transport, export, and incorporation of the newly synthesized highly methyl esterified pectin into the growing primary cell wall, and (6) Methanol production during growth associated with pectin demethylation during the day and night. These observations are consistent with the emerging view of methanol emissions as a chemical signal of leaf growth primarily occurring at night and early morning, due to reduced leaf water stress. The results are also consistent with a critical role of methanol production linked to diurnal changes in primary cell wall elasticity associated with pectin demethylation and growth. Although the rise in atmospheric CO<sub>2</sub> inhibits major metabolic pathways like photorespiration and the isoprenoid pathway, the photorespirationindependent photosynthetic C1 pathway may accelerate. Thus, photosynthetic production of AdoMet may play a critical, yet poorly understood role in enhancing growth rates of plants and net primary productivity of ecosystems during terrestrial CO<sub>2</sub> fertilization.

Dewhirst, R., J. Mortimer, K. Jardine. 2020. "Do Cell Wall Esters Facilitate Forest Response to Climate?" *Trends in Plant Science* **25**(8), 729–32. DOI:10.1016/j.tplants.2020.05.01.1.

Jardine, K. J., et al. 2022. "Cell Wall Ester Modifications and Volatile Emission Signatures of Plant Response to Abiotic Stress," *Plant, Cell* & Environment **45**(12), 3429–44.

Jardine et al. In preparation. "Plant Methanol Emission at the Interface of the Photosynthetic C1 Pathway and Diurnal Patterns in Leaf Water Status and Growth."

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), Early Career Research Project (ECRP) grant no. FP00007421 and the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by contract DE-AC02-05CH11231

## Phenotypic Characterization of Sorghum Nitrogen Responsive Gene Edits Using High-Throughput Phenotyping

**Hongyu Jin**<sup>1</sup>, Alexa Nolan<sup>2</sup>, Yufeng Ge<sup>1</sup>, James C. Schnable<sup>1</sup>, Ravi V. Mural<sup>1</sup>, Thomas Clemente<sup>1</sup>, Jinliang Yang<sup>1</sup>

<sup>1</sup>University of Nebraska–Lincoln (UNL); <sup>2</sup>University of Alabama–Huntsville

Inefficient use of inorganic nitrogen (N) fertilizers can result in environmental issues and increased farming costs. To address this issue, this project aims to investigate the phenotypic and molecular effects of N treatments on a set of sorghum gene edits identified as N-responsive candidate genes through previous studies. As a pilot study, the team used UNL's LemnaTec phenotyping facility to collect imagery data for five sorghum edits (including one triple gene edit, two double gene edits, and two single gene edits) throughout different developmental stages. The team extracted numerical phenotypic values from the images and performed statistical analyses to determine the effects of N treatment on each edit. The project's results showed statistically significant phenotypic effects for several edits in response to N treatments. Based on these preliminary findings, the team will refine phenotypic characterization procedures and conduct additional phenotyping for other gene edits.

The outcomes of this project will contribute to the development of sustainable and efficient crop production methods, thus advancing agricultural practices in a more environmentally responsible manner.

## Identification of Regulatory Mechanisms Underlying Cell Differentiation in Sorghum Biomass

Wilfred Vermerris<sup>1\*</sup> (wev@ufl.edu), Christopher Dervinis<sup>1</sup>, Kelly Balmant<sup>1</sup>, Sushmita Roy<sup>2</sup>, **Matias Kirst<sup>1</sup>** 

#### <sup>1</sup>University of Florida; <sup>2</sup>University of Wisconsin

**Goals:** Researchers plan to alter the genetic regulation of the cellular developmental programs that generate the vegetative tissues of sorghum, with the aim to increase the proportion of cells that are less recalcitrant to biomass deconstruction.

Plant biomass is comprised of distinct cell types, which largely determine its physical and chemical properties, and hence, its recalcitrance to biomass processing aimed at generating fermentable sugars that microbes can convert to biofuels. For example, the walls of parenchyma cells in the stalks of maize (Zea mays L.) and sorghum (Sorghum bicolor L. Moench) can be broken down using milder pretreatment conditions and with lower cellulase loadings than the lignified cells present in the outer rind of the stalk (Zeng et al. 2012; Zeng et al. 2012; Li et al. 2018). The different cell types within the sorghum plant develop from undifferentiated meristem cells through variation in the spatiotemporal expression of regulatory genes that control structural genes. The understanding of the role of specific genes and their regulation in this developmental process is incomplete. Uncovering the function of the complete ensemble of genes involved in the differentiation and maturation of the cells that make up sorghum biomass creates the opportunity to manipulate its cellular composition, impacting its physical and chemical properties and, consequently, its value for bioenergy.

Researchers propose applying single-cell genome and transcriptome analysis of the sorghum shoot apex and stem to identify the function of genes involved in differentiating cells that determine biomass composition. Inferred cell lineage trajectories involved in the development of the cellular components of biomass will be explored to discover their regulators. The specific objectives are to (1) define the function of each gene (including specific members within gene families) with respect to the development of the main cell types that determine sorghum biomass and its cell-wall composition; (2) construct the cellular lineages that give rise to each cell type that composes biomass (from the shoot apical and vascular cambium meristem cells to cells in the stem), and identify genes and *cis*-regulatory elements that contribute to the lineage progression; (3) categorize the function of gene and associated cis-regulatory components for their relevance in the control of cellular lineages that lead to each cell type, and (4) validate multiple targets in isolation and in parallel, to confirm their role in biomass development and their potential for enhancing biomass yield and its properties.

Li, M., et al. 2018. "Physical Fractionation of Sweet Sorghum and Forage/Energy Sorghum for Optimal Processing in a Biorefinery," *Industrial Crops and Products* **124**, 607–16.

Zeng, M., et al. 2012. "Enzyme Hydrolysis and Imaging of Fractionated Corn Stalk Tissues Pretreated with Liquid Hot Water (Part 1)," *Biotechnology and Bioengineering* **109**, 390–97. DOI:10.1002/ bit.23337.

Zeng, M., et al. 2012. "Enzyme Hydrolysis and Imaging of Fractionated Corn Stalk Tissues Pretreated with Liquid Hot Water (Part 2)," *Biotechnology and Bioengineering* **109**, 398–404. DOI:10.1002/ bit.23335.

**Funding Information:** This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0023082.

### Single-Cell Genomics of Poplar Wood Development

Wendell Pereira<sup>1\*</sup>(wendellpereira@ufl.edu), Wilfred Vermerris<sup>1</sup>, Christopher Dervinis<sup>1</sup>, Kelly Balmant<sup>1</sup>, Sushmita Roy<sup>2</sup>, **Matias Kirst<sup>1</sup>** 

#### <sup>1</sup>University of Florida; <sup>2</sup>University of Wisconsin

**Goals:** (1) Uncover the cellular developmental program of woody biomass in the perennial bioenergy crop poplar by dissecting the cell lineages that originate in the vascular cambium. (2) Modify the regulatory programs that lead to the formation of the various cell types in poplar wood to achieve less recalcitrant biomass for bioenergy production.

The vascular cambium is responsible for the production of the secondary xylem, which comprises the most abundant form of biomass on Earth, wood. Despite its massive global importance, the genetic networks underlying the production of secondary xylem remain partially ambiguous. Differentiation of stem cells in the plant apex gives rise to aerial tissues and organs. These cells later differentiate to form the vascular cambium, from which secondary xylem is generated. Here the team used single-nuclei RNA sequencing (snRNA-seq) to determine cell-type specific transcriptomes of the *Populus trichocarpa* vegetative shoot apex and lignified stem to create a cell-type specific atlas of their tissues and uncover the regulators of cell lineage trajectories.

From P. trichocarpa shoot apex, researchers identified highly heterogeneous cell populations that clustered into seven broad groups represented by 18 transcriptionally distinct types (See Fig. 7, p. 32). Next, researchers established the developmental trajectories of the epidermis, leaf mesophyll, and vascular tissue. Motivated by the high similarities between Populus and Arabidopsis cell populations in the vegetative apex, researchers applied a pipeline for interspecific single-cell gene expression data integration. Researchers contrasted the developmental trajectories of primary phloem and xylem formation in both species, establishing the first comparison of vascular development between a model annual herbaceous and a woody perennial plant species. In addition to providing a cell atlas of the shoot apical meristem and its derived lineages, the results offer a valuable resource for investigating the principles underlying cell division and differentiation between herbaceous and perennial species.

In parallel to the *Populus* shoot apex analysis, researchers performed an snRNA-seq of 11,673 nuclei derived from lignified stem to profile the transcriptome and create a singlecell atlas. Cell-type specific marker genes were utilized to identify 20 transcriptionally distinct cell clusters representing nearly all cell types within the sampled tissue. Reporter



**Fig. 7.** Cell populations in the *Populus* vegetative shoot apex. **(A)** Visualization of 18 cell clusters using UMAP. **(B)** Expression pattern of characterized cell type marker genes in the *Populus* apex cell clusters. **(C)** A longitudinal section of a *Populus* apex shows the spatial location of the population of annotated cells. **(D)** Visualization of the eight cell clusters using UMAP. Colors represent the cell types. [Used with permission of The Company of Biologists Ltd. from Conde, D., et al. 2022. "Single-Nuclei Transcriptome Analysis of the Shoot Apex Vascular System Differentiation in *Populus*," *Development* **149**(21), dev200632. DOI:10.1242/dev.200632]

gene assays were carried out to confirm the cluster identity of vessel elements, fibers, ray parenchyma, cambial cells, and sub-cell type vessel-associated cells. Finally, the developmental trajectory of cambial cells and their xylem-specific derivates was carried out to identify lineages containing putative regulators related to vascular development and xylogenesis. This trajectory analysis identified putative regulators of the cell lineages that result in the formation of fibers and vessels. The functions of these regulators are being evaluated in knockout experiments, in which researchers will assess if cell lineages can be redirected toward developing specific cell types.

## Enhanced Resistance Pines for Improved Renewable Biofuel and Chemical Production

Mallory Morgan<sup>\*1</sup> (Mallory.Morgan@ufl.edu), Matthew Lane<sup>2</sup>, Salvador Gezan<sup>3</sup>, Christopher Dervinis<sup>1</sup>; Daniel Ence<sup>1</sup>, David Kainer<sup>4</sup>, Mirko Pavicic<sup>4</sup>, Manesh Shah<sup>4</sup>, Daniel Jacobson<sup>4</sup>, **Gary F. Peter**<sup>1</sup>

<sup>1</sup>University of Florida; <sup>2</sup>University of Tennessee; <sup>3</sup>VSN International, Hemel Hempstead, UK; <sup>4</sup>Oak Ridge National Laboratory

**Goals:** The goal is to genetically increase constitutive terpene defenses of loblolly and slash pine to enhance

protection against pests and pathogens and simultaneously expand terpene supplies for renewable biofuels and chemicals.

The constitutive and inducible oleoresin defense network in loblolly (Pinus taeda) and slash (Pinus elliottii var elliottii) pine provides physical and chemical resistance to insects and pathogens and the chemical composition of oleoresin can be used as a renewable source of biofuels harvested directly from live tree stems. Increasing pine terpenes is well aligned with the needs of the developing bioeconomy, as the southeastern United States hosts the world's largest biomass supply chain, annually delivering 17% of global wood products, and has the potential to expand the U.S. pine chemicals industry by increasing biofuels from pine terpenes, which is limited by relatively low average wood terpene content. The focus is to increase constitutive terpene production to enhance loblolly and slash pine resistance to pests and pathogens and to simultaneously increase biofuel feedstocks in these commercial pine species.

Pine terpenes evolved as a primary chemical and physical defense system and are a main component of a durable, quantitative defense mechanism against pests and pathogens. In previous research it was demonstrated that terpene defense traits are under genetic control and behave as quantitative traits and have used genetic engineering to validate 12 genes that can significantly increase wood terpene content. In objective one, researchers are integrating existing and new genome wide association studies (GWAS) genetic results with RNA expression, quantitative trait locus (QTL) mapping, and allele frequency information in known high oleoresin flow selections and the project's breeding populations to discover and validate loblolly and slash pine alleles/ genes that are important for resistance.

GWAS analyses of constitutive oleoresin flow, wood diterpenoid content, and resin canal number with ~83,000 biallelic single nucleotide polymorphisms (SNPs) were completed for the project's CCLONES population and constitutive oleoresin flow, mono- and diterpene content are complete and complete and resin canal number is in progress for the project's ADEPT2 population. In the ADEPT2 population, researchers simultaneously measured constitutive and induced oleoresin flow after treating clones with methyl-jasmonate (MeJA). While the goal is to increase constitutive terpene defenses, researchers use MeJA to induce defense responses to identify the genes and genetic architecture of resinosis. In the ADEPT2 population, the team found the clonal repeatability of constitutive oleoresin flow and inducible oleoresin flow to be 0.31, suggesting these traits are under moderate genetic control.

The team's estimate of clonal repeatability for constitutive oleoresin flow in the ADEPT2 population is consistent with what was previously published in the CCLONES

population, and the estimate of clonal repeatability for inducible oleoresin flow in the ADEPT2 population is the first estimate of genetic control for this trait. Importantly, researchers observed a strong genetic correlation (0.82) between induced and constitutive oleoresin flow, suggesting the genetic architecture between these traits is shared.

Researchers conducted association analyses—with constitutive and inducible oleoresin flow, wood monoterpene content and composition, and diterpenoid content obtained in the ADEPT2 population—using linear mixed models and multi-locus linear mixed models in ASRgwas and GAPIT packages using two sets of SNP markers totaling ~2.28 million biallelic SNPs.

Significant SNPs are being mapped to genes and compared with those for constitutive oleoresin flow found in the CCLONES population. In the pseudo-backcross population between one F1 slash x loblolly hybrid genotype backcrossed to slash and loblolly genotypes, researchers collected constitutive and induced oleoresin flow and needle tissue for future QTL mapping.

To identify early, mid, and late genes expressed in differentiating resin ducts, team members induced axial resin canal formation in the cambial meristem by applying methyljasmonate (MeJA), which is a known inducer of traumatic resin canal formation in the Pinaceae family. The team conducted a time course experiment where 78 RNAseq libraries were created from cambial zone tissue collected from days 0, 1 to 14, 17, and 21 after MeJA treatment. Researchers also constructed 43 RNAseq libraries from 10 transgenic pine lines from four different constructs with significantly elevated wood terpene content. They pooled all libraries from the time course and from the transgenic lines and sequenced to a 30x read depth with the NovaSeq Illumina NGS platform.

Team members mapped the reads to an improved *de novo* loblolly pine transcriptome that includes 64,671 genes composed of existing EST contigs, PacBio reads, and predicted transcripts from loblolly pine reference genome v2.01. Researchers used DESeq2 to identify thousands of significantly differentially expressed genes across the time course and in transgenic pines compared with wild type. With these differentially expressed genes researchers created a Predictive Expression Network (PEN) using iterative Random Forest Leave-One-Out Prediction to illustrate higher-order interactions between genes and to determine the gene-togene relationships that are the most highly predictive of each other. To identify and prioritize genes across the PEN that are involved in axial resin canal formation, researchers applied random walk with restart (RWR) algorithms based on a set of literature-curated seed genes that included known orthologous regulators of xylem formation and development, which are suppressed while resin canal formation

is increased. The RWR approaches allowed us to identify mechanistically associated genes that did not appear in GWAS due to a lack of statistical power or genetic variation but are still important components of resinosis. Researchers are continuing to annotate the network to identify genes whose expression supports involvement in resin canal formation and terpene synthesis.

In objective two, team members are using information from objective one to accelerate breeding for increased resistance in loblolly and slash pine through marker-assisted introgression and will develop and test genomic selection models to accelerate breeding of resistant slash pine.

## Improving Candidate Gene Discovery by Combining Multiple Genetic Mapping Datasets

Nirwan Tandukar, Fausto Rodríguez-Zapata, Jung-Ying Tzeng, **Rubén Rellán-Álvarez**\* (rrellan@ncsu.edu)

North Carolina State University

**Goals:** (1) Perform an environmental GWAS in a panel of ~2000 sorghum accessions that have already been geno-typed and georeferenced using phosphorus availability and early season cold stress as the phenotypes for the GWAS analysis.

(2) Characterize the genetic architecture of lipid content during the early stages of sorghum development using the SAP. The team will sequence the SAP accessions at 10–15X and make these data available. Researchers will perform a GWAS on lipid content under both stress conditions (low temperature and low phosphorus).

(3) Develop algorithms that incorporate all the different types of information collected (i.e., metabolite levels, GWAS candidate genes, selection signals) to improve the ability to detect signals of small effects and increase confidence in the selection of candidate genes. The algorithms and pipelines developed here will be made available to the community as R packages.

With a growing wealth of genetic datasets generated by next-generation sequencing coupled with the advent of large plant phenotyping datasets, exciting new corridors for investigations have opened for understanding complex traits due to the environment. The Genome-Wide Association Studies (GWAS) model identifies associations between single nucleotide polymorphisms (SNPs) and the phenotype. Complex biological processes involve multiple phenotypes. Researchers have previously identified lipid variation for maize adaptation in Mexican highlands, which has adapted to low phosphorus and cold. The team is now using high-dimensional *Sorghum bicolor* genetic datasets to perform environmental GWAS for various soil phosphorus phenotypes (availability, concentration, and solubility) in African region, Fst measurement in the same panel adapted to high and low phosphorus, and finally, profiling various lipid concentrations (LC/MS) in low and high phosphorus in Sorghum Association Panel (SAP) and their subsequent metabolomics GWAS. Comprehensive research on the complete genetic architecture is laborious, costly, and time extensive due to the overwhelming number of genes and their regulatory networks, different phenotypes explaining the same adaptive process, and multidimensional genomics datasets. Hence, the need for developing a robust statistical framework that can combine information from different experiments and genomics dataset in an individual p-value level that can aggregate multiple small and large effects of genes and redefine the order of emphasis of genes. For such an outcome, researchers use the Cauchy distribution to define a test statistic as a weighted sum of Cauchy transformation of individual p-values, which can be used to combine p-values across different datasets. Researchers are working towards creating such a framework through R packages that will be publicly available. Finally, researchers hope to have an accurate description of the molecular mechanisms involving phosphorus. Team members also hope to test whether lipids in sorghum play a similar adaptive role in Africa, and whether there is convergence between the sorghum and maize in the plausible molecular mechanisms for such an adaptation.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021889.

## 532 Genomes Reveal Natural Variation and Local Adaptation History in Pennycress

Xing Wu<sup>1</sup>, Sebastian E. Toro Arana<sup>1,2\*</sup> (sebatoro@stanford. edu), Lucas Czech<sup>1</sup>, Shannon Hateley<sup>1</sup>, Jason Thomas<sup>1</sup>, Ratan Chopra<sup>3</sup>, M. David Marks<sup>3</sup>, **John Sedbrook**<sup>4</sup>, Moises Exposito-Alonso<sup>1,2</sup>

<sup>1</sup>Carnegie Institution for Science, Stanford, CA; <sup>2</sup>Stanford University; <sup>3</sup>University of Minnesota; <sup>4</sup>Illinois State University

https://www.pennycressresilience.org

**Goals:** This project employs evolutionary and computational genomic approaches to identify key genetic variants that have enabled *Thlaspi arvense* L. (field pennycress; pennycress) to locally adapt and colonize all temperate regions of the world. This, combined with knowledge of metabolic and cellular networks derived from first principles, guides precise laboratory efforts to create and select high-resilience lines, both from arrays of random mutagenesis and by employing cutting-edge CRISPR genome editing techniques. This project will deliver speed-breeding methods and high-resilience mutants inspired by natural adaptations and newly formulated biological principles into a wide range of commercial pennycress varieties to precisely adapt them to the desired local environments.

Pennycress is under development as an annual winter oilseed cover crop for the 80 million acre U.S. Midwest Corn Belt and other temperate regions including the Pacific Northwest. It has demonstrated unique attributes such as extreme cold resilience, rapid spring growth, and adaptation to various environments. Identifying genomic loci contributing to its resilience and local adaptation will significantly benefit its breeding programs and shed light on other Brassicaceae bioenergy crops for climate adaptation. Here is presented an analysis of 532 high-quality whole-genome resequenced wild pennycress accessions collected from their natural Eurasia and North American range (Pennycress Genome Portal, Nunn et al. 2022, Geng et al. 2021). The team comprehensively identified 6.3 million single nucleotide polymorphisms (SNPs) using the grenepipe variant calling pipeline (Czech and Exposito-Alonso 2022). In-depth analyses of population structure and demography indicate multiple recent migrations of the North American accessions from Europe with similar extensive genetic variation structured latitudinally. Preliminary admixture results suggest that introgression may play an important role in the local adaptation of pennycress. Genome-wide scans of selection signals and climate GWAS provide candidate genomic regions responsible for local adaptation and cold tolerance. In conclusion, this study offers valuable genomic resources for pennycress breeding and helps elucidate the history of migration and local adaptation in pennycress.

Chopra, Ratan. 2020. "Pennycress – A Solution for Global Food Security, Renewable Energy and Ecosystem Benefits." www.osti. gov/award-doi-service/biblio/10.46936/10.25585/60001359.

Czech, L. and Exposito-Alonso, M. 2022. Grenepipe: A Flexible, Scalable and Reproducible Pipeline to Automate Variant Calling from Sequence Reads," *Bioinformatics* **38**, 4809–11.

Geng, Y., et al. 2021. "Genomic Analysis of Field Pennycress (*Thlaspi arvense*) Provides Insights into Mechanisms of Adaptation to High Elevation," *BMC Biology* **19**, 143.

Nunn, A., et al. 2022. "Chromosome-Level Thlaspi arvense Genome Provides New Tools for Translational Research and for a Newly Domesticated Cash Cover Crop of the Cooler Climates," *Plant Biotechnology Journal* **20**, 944–63.

**Funding Information:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021286

## A Temporal Atlas and Response to Nitrate Availability of 3D Root System Architecture in Diverse Pennycress (*Thlaspi arvense* L.) Accessions

Marcus Griffiths<sup>1\*</sup> (mgriffiths@danforthcenter.org), Alexander E. Liu<sup>1</sup>, Shayla L. Gunn<sup>1</sup>, Nida M. Mutan<sup>1</sup>, Elisa Y. Morales<sup>1</sup>, Vannessica Jawahir<sup>1</sup>, Dmitri Nusinow<sup>1</sup>, Christopher N. Topp<sup>1</sup>, **John Sedbrook**<sup>2</sup>

<sup>1</sup>Donald Danforth Plant Science Center; <sup>2</sup>Illinois State University

#### https://www.pennycressresilience.org/

**Goals:** This project employs evolutionary and computational genomic approaches to identify key genetic variants that have enabled *Thlaspi arvense* L. (field pennycress; pennycress) to locally adapt and colonize all temperate regions of the world. This, combined with knowledge of metabolic and cellular networks derived from first principles, guides precise laboratory efforts to create and select high-resilience lines, both from arrays of random mutagenesis and by employing cutting-edge CRISPR genome editing techniques. This project will deliver speed-breeding methods and high-resilience mutants inspired by natural adaptations and newly formulated biological principles into a wide range of commercial pennycress varieties to precisely adapt them to the desired local environments.

Roots have a central role in plant resource capture and are the interface between the plant and the soil affecting multiple ecosystem processes. Field pennycress (Thlaspi arvense L.) is a diploid annual cover crop species that has potential utility for reducing soil erosion and nutrient losses; and has rich oil seeds amenable as a biofuel (30% to 35% oil) or high-protein animal feed. The objective of this research was to (1) precisely characterize root system architecture and development, (2) understand adaptive responses of pennycress roots to nitrate nutrition, (3) and determine genotypic variance available in root development and nitrate plasticity. Using a root imaging and analysis pipeline 4D pennycress root system architecture was characterized under four nitrate treatments across time. Significant nitrate condition response and genotype interaction was identified for many root traits with a greater impact on lateral root traits. In trace nitrate conditions a greater lateral root count, length, interbranch density, and a steeper lateral root angle was observed compared to high nitrate conditions. Genotype by nitrate condition interaction were observed for root width, width depth ratio, mean lateral root length, and lateral root density. Further, a large format mesocosm system has been developed and used to visualize root system architecture of mature plants with neighbors. Using this system, researchers are assessing how variation

in pennycress root system architecture can affect ecosystem service and abiotic stress tolerance scaling from single plant to canopy level traits. These results illustrate root trait variance available in pennycress accessions and useful targets for breeding of improved nitrate responsive cover crops for greater productivity, resilience, and ecosystem service.

Griffiths, M., et al. 2022. "Optimisation of Root Traits to Provide Enhanced Ecosystem Services in Agricultural Systems: A Focus on Cover Crops," *Plant, Cell & Environment*. DOI:10.1111/pce.14247.

Griffiths, M., et al. 2023. "A Temporal Atlas and Response to Nitrate Availability of 3D Root System Architecture in Diverse Pennycress (*Thlaspi arvense*) Accessions," *bioRxiv*. DOI:10.1101/2023.01.14.524046.

**Funding Information:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science program grant no. DE-SC0021286.

## Comparison of Physiological and Metabolomic Responses to Drought Across Pennycress CRISPR Mutants and Natural Accessions

Jason Thomas<sup>1\*</sup> (jthomas@carnegiescience.edu), Charles Hawkins<sup>1</sup>, Liza Gautam<sup>2</sup>, Xinxin Ding<sup>3</sup>, Chaevien S. Clendinen<sup>3</sup>, Carol Kiam Assato<sup>2</sup>, John Lagergren<sup>4</sup>, Daniel Jacobson<sup>4</sup>, John Sedbrook<sup>2</sup> (jcsedbr@ ilstu.edu), Pubudu Handakumbura<sup>3</sup>, Seung Y. Rhee<sup>1</sup>

<sup>1</sup>Carnegie Institution for Science, Stanford, CA; <sup>2</sup>Illinois State University; <sup>3</sup>Pacific Northwest National Laboratory; <sup>4</sup>Oak Ridge National Laboratory

#### https://www.pennycressresilience.org

**Goals:** This project employs evolutionary and computational genomic approaches to identify key genetic variants that have enabled *Thlaspi arvense* L. (field pennycress; pennycress) to locally adapt and colonize all temperate regions of the world. This, in combination with knowledge of metabolic and cellular networks derived from first principles, is guiding precise laboratory efforts to create and select high-resilience lines, both from arrays of random mutagenesis and by employing cutting-edge CRISPR genome editing techniques. This project will deliver speed-breeding methods and high-resilience mutants inspired by natural adaptations and newly formulated biological principles to be introduced into a wide range of commercial pennycress varieties to precisely adapt them to the desired local environments.

Field pennycress is an overwintering bioenergy cover crop that is rapidly being domesticated. It produces oilseeds that can be converted into various products from cooking oil to renewable diesel and sustainable aviation fuel (SAF). By replacing fossil fuels, pennycress directly combats climate change. However, abiotic stresses brought by climate change threatens stable production of many crops including pennycress. Pennycress's drought response has not been studied in much detail. Here researchers explored the drought responses of wild-type, gene-edited, and natural accessions by physiological assays, metabolic profiling, and data integration via a genome-scale metabolic pathway database for pennycress. Team members first measured phenotypes relevant to drought stress in the reference line Spring32-10 seedlings and plants including yield, biomass, stomatal conductance, and photosynthetic efficiency over various drought conditions. Researchers subjected plants to drought and control conditions and then collected above- and belowground tissues for metabolomic analyses via liquid chromatography coupled mass spectrometry, which allows detection of changes in plant metabolism during drought stress. These data will be analyzed in the context of the PennycressCyc database the team created for the pennycress community. Additionally, team members generated pennycress single, double, and triple mutants using CRISPR-Cas9 mutagenesis targeting 10 genes important for drought responses in other species. These mutants, along with a subset of 800 pennycress worldwide natural accessions predicted to have varied levels of drought resilience based on the team's climatype predictions, were subjected to drought conditions and phenotyped to identify relative differences in drought responses. Taken together, this work helps decipher how pennycress uniquely responds to drought stress and is identifying natural and induced genetic changes that could improve pennycress drought resilience.

**Funding Information:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program grant no. DE-SC0021286.

## Functional Analysis of Genes Encoding Ubiquitin Proteasome System Components Affecting Poplar Wood Traits

Daniela Rodriguez-Zaccaro<sup>1,2</sup>\* (dzaccaro@ucdavis.edu), Andrew Groover<sup>1,2</sup>, Justin Walley<sup>3</sup>, **Nitzan Shabek**<sup>1</sup>

<sup>1</sup>University of California–Davis; <sup>2</sup>U.S. Forest Service, Pacific Southwest Research Station, Davis, CA; <sup>3</sup>Iowa State University

**Goals:** Wood vessel trait candidate genes coding for E3 ubiquitin ligase enzymes will be functionally characterized and examined through CRISPR-cas9 genome editing, TurboID proximity labeling, and drought and ABA treatments. Specifically, the data generated in this project will be further used to study the transcriptome, proteome, interactome, and ubiquitinome in poplar wood forming tissues.

Wood is the water-conducting tissue of tree stems. Like most angiosperm trees, poplar wood contains water-conducting vessel elements whose anatomical properties affect water transport and growth rates as well as susceptibility to cavitation and hydraulic failure during drought. Despite their key role in determining the hydraulic physiology of trees, the genetic regulation of vessel element morphological traits is poorly understood. In a preliminary study, a dosage- based genome-wide screen found significant associations (or dosage QTL regions) between wood vessel traits and specific regions of the genome. Poplar wood forming tissues were then sampled to conduct a gene coexpression network analysis. Height-corrected vessel frequency was significantly correlated to a group of co-expressed genes that code for E3 ubiquitin ligase components of the ubiquitin proteasome system. The team found that some of these genes are located within a chromosome 9 dosage QTL region identified in its previous screen, suggesting that these could affect vessel trait variation in a dosage-dependent manner. From these genes, the team selected vessel trait-related candidates for further characterization, including E3 ubiquitin ligases makorin (MKRN), SKP1-interacting protein 2 (SKIP2), and Big Brother (BB). Based on these preliminary findings, future aims involve the novel functional characterization of key components of ubiquitin-proteasome regulation in poplar wood forming tissue. To meet this goal, researchers will generate CRISPR-cas9 mutants for poplar ubiquitin E3 ligase candidate genes to determine changes in wood phenotype, gene expression, protein abundance and ubiquitinomes. The team will also use a TurboID proximity labeling strategy in poplar to identify the interacting partners for candidate proteins and characterize their specific structure and function. Similarly, the transcriptome, proteome, interactome, and ubiquitinome of trees that were grown under drought or treated with ABA will be determined. Ultimately, these strategies will shed light on the role of the ubiquitin proteasome system in wood formation, vessel trait variation, and tree responses to the environment.

**Funding Information:** This project is supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research, Genomic Science Program grant no. DE-SC0023158. F.D.R.Z. was supported in part by an appointment to the Science Education Programs at National Institutes of Health (NIH), administered by ORAU through the U.S. Department of Energy Oak Ridge Institute for Science and Education.

## WINTR: Winter Transcriptome Regulation in Poplar

**Chung-Jui Tsai**<sup>1\*</sup> (cjtsai@uga.edu), Wellington Muchero<sup>2</sup>, Priya Ranjan<sup>2</sup>

<sup>1</sup>University of Georgia; <sup>2</sup>Oak Ridge National Laboratory

**Goals:** This project aims to advance understanding of how molecular control of the winter latent state is linked to perennial woody biomass productivity.

Woody biomass growth of trees comprises a significant contribution to the supply of renewable feedstocks for biofuels and biomaterials in the emerging bioeconomy. Genetic diversity in the molecular and physiological underpinnings of woody biomass growth continues to be explored for its potential adoption in tree improvement. Most of what is known about these underpinnings has come from the study of development and expansion growth during the summer or indoors. However, the woody bole of a field-grown tree is physiologically active year-round. In temperate deciduous tree species, physiological and metabolic adjustments are essential to confer winter protection in the wood-forming tissues. Mechanisms for the avoidance or tolerance of freeze-related intracellular or extracellular desiccation and for the protection and maintenance of plasma membrane and cell wall become critical.

Researchers focus on two complementary Populus experimental systems, P. trichocarpa with rich population genomic resources and the fast-growing hybrid P. tremula × P. alba INRA 717-1B4 (717) with proven transformation and genome editing efficiencies. A multipronged approach integrating stem RNA-Seq, genome-wide association studies (GWAS), expression quantitative trait loci (eQTL) mapping, high-precision CRISPR genome editing, and gene network modeling will be used to investigate transcriptome regulation in woody stem tissues during the winter. Of particular interest are genome duplicates that exhibit either winter-biased expression or divergent seasonal expression in both species. GWAS and eQTL predictions will be experimentally tested by CRISPR editing of coding or cis regulatory sequences for investigating their functional links to seasonal growth transitions and woody biomass accrual. Confirmed mutants will be field tested for seasonal growth transitions, transcript and metabolite profiling, and histological analysis. The transcriptomics data will feed back into regulatory network construction to improve inference of winter processes controlled by seasonal biased genes or their regulators. A key deliverable will be the contribution of winter stem transcriptomes to existing expression data obtained primarily from actively expanding tissues. The data will be integrated within the DOE Systems Biology KBase to promote further research efforts. Understanding how seasonal growth dynamics impact woody biomass productivity will offer new targets for bioenergy crop improvement.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0023166. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

## Functional Characterization of Glycosyltransferases in Duckweed to Enable Predictive Biology

Pradeep Kumar Prabhakar<sup>1</sup>, Digantkumar Gopaldas Chapla<sup>1</sup>, Charles Joseph Corulli<sup>1</sup>, Brayden Paul Smith<sup>1</sup>, Samantha Hennen<sup>2</sup>, Morgan J. Willis<sup>1</sup>, Tasleem Javaid<sup>1</sup>, Samantha J. Ziegler<sup>2</sup>, Vivek S. Bharadwaj<sup>2</sup>, Kelley W. Moremen<sup>1</sup>, Maria J. Pena<sup>1</sup>, Yannick J. Bomble<sup>2</sup>, **Breeanna R. Urbanowicz<sup>1\*</sup>** 

<sup>1</sup>University of Georgia; <sup>2</sup>National Renewable Energy Laboratory

Glycosyltransferases (GTs) catalyze the formation of glycosidic linkages to produce complex carbohydrates. This project involves the use of a multidisciplinary, high-throughput biochemical and computational biology approach to study carbohydrate metabolic processes in duckweed, a promising energy crop. The role of enzymatic microenvironments is being assessed through a combined proteomic and computational biology approach. This combined data will be used to populate a deep-learning framework to predict plant GT function.

Functional validation achieved through this research will be used to assign gene function and study plant processes at the systems level to efficiently link genome sequence with gene function in a feedstock agnostic manner.

**Funding Information:** This research is supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program, Genomic Science program grant no. DE-SC0023223.

## A Leaf-Level Spectral Library to Support High-Throughput Plant Phenotyping: Predictive Accuracy and Model Transfer

Nuwan K. Wijewardane<sup>1</sup>, Huichun Zhang<sup>2</sup>, Jinliang Yang<sup>1</sup>, James C. Schnable<sup>1</sup>, Daniel P. Schachtman<sup>1</sup>, Yufeng Ge<sup>1</sup>

<sup>1</sup>Mississippi State University; <sup>2</sup>Nanjing Forestry University, China

Leaf-level hyperspectral reflectance data has become an effective tool for high-throughput phenotyping of plant leaf traits due to its rapid, low-cost, multisensing, and nondestructive nature. However, model calibration is often expensive regarding the number of samples, time, and labor; and models show poor transferability among different datasets. Building large spectral datasets across multiple species enables accurate model calibration and improves model transferability. The team pursued three specific objectives: (1) assemble a large library of leaf hyperspectral data (n=2,460) constructed from maize and sorghum, (2) evaluate two machine-learning approaches to estimate nine leaf properties, and (3) investigate the utility of this spectral library for predicting external datasets (n=445) including maize, sorghum, soybean, and *Camelina* using extraweighted spiking. Partial least squares regression models exhibited higher predictive performance than deep neural network models.

Models calibrated solely using the spectral library showed poor performance when applied to external datasets (R2<0.3 for N, P, and Ca with *Camelina* samples). However, model transferability improved significantly when extraweighted spiking with a small dataset (n=20) was employed (R2>0.6 for N, P, and Ca with *Camelina* samples), indicating that it can improve the effectiveness and utility of spectral libraries in high-throughput phenotyping contexts.

#### Early Career

## Improved Biofuel Production Through Discovery and Engineering of Terpene Metabolism in Switchgrass

Kira Tiedge<sup>1,2\*</sup> (kjtiedge@ucdavis.edu), Gabby Wyatt<sup>1\*</sup>, **Philipp Zerbe**<sup>1</sup>

<sup>1</sup>University of California–Davis; <sup>2</sup>University of Groningen

Goals: Of the myriad specialized metabolites that plants form to adapt to environmental challenges, terpenoids form the largest group. In many major crops, unique terpenoid blends serve as key stress defenses that directly impact plant fitness and yield. In addition, select terpenes are used for biofuel manufacture. Thus, engineering of terpenoid metabolism can provide a versatile resource for advancing biofuel feedstock production but requires a system-wide knowledge of the diverse biosynthetic machinery and defensive potential of often species-specific terpenoid blends. This project merges genome-wide enzyme discovery with comparative omics and protein structural studies to define the biosynthesis and stress-defensive functions of switchgrass (Panicum virgatum) terpenoid metabolism. These insights would be combined with the development of genome editing tools to design plants with desirable terpene blends for improved biofuel production on marginal lands.

Diterpenoids form a diverse class of metabolites with critical functions in plant development, defense, and ecological adaptation. Major monocot crops, such as maize (*Zea mays*) and rice (Oryza sativa), deploy diverse blends of specialized diterpenoids as core components of biotic and abiotic stress resilience. This project reports the genome-wide discovery and functional characterization of the stress-related diterpenoid-metabolic network in switchgrass (P. virgatum). Mining of the allotetraploid switchgrass genome identified expansive diterpene synthase (diTPS) and cytochrome P450 monooxygenase (P450) enzyme families critical for the chemical diversity of bioactive diterpenoids. Tissue-specific transcriptome and metabolite analyses of drought-resistant (Alamo) and drought-susceptible (Cavein-Rock) genotypes showed an earlier onset of transcriptomic changes and significantly more differentially expressed genes in response to drought in Cave-in-Rock. Diterpenoidbiosynthetic genes showed drought-inducible expression in Alamo roots, contrasting largely unaltered triterpenoid and phenylpropanoid pathways. In addition, metabolomic analyses identified common and genotype-specific terpenoids. Consistent with transcriptomic alterations, several root diterpenoids showed significant drought-induced accumulation. Structural analysis of drought-responsive root diterpenoids verified these metabolites as oxygenated furanoditerpenoids that are perhaps unique to switchgrass. Together, these findings support a role of diterpenoids in switchgrass drought stress tolerance and provide resources

for understanding the molecular mechanisms underlying switchgrass environmental resilience.

Muchlinski, A., et al. 2021. "Cytochrome P450-Catalyzed Biosynthesis of Furanoditerpenoids in the Bioenergy Crop Switchgrass (*Panicum virgatum* L.)," *The Plant Journal* **108**, 1053–68.

Pelot, K. A., et al. 2018. "Functional Diversity of Diterpene Synthases in the Biofuel Crop Switchgrass," *Plant Physiology* **178**, 54–71.

Tiedge, K., A. Muchlinski, and P. Zerbe. 2020. "Genomics-Enabled Analysis of Specialized Metabolism in Bioenergy Crops: Current Progress and Challenges," *Synthetic Biology* **5**, ysaa005.

Tiedge, K., et al. 2022a. "Comparative Transcriptomics and Metabolomics Reveal Specialized Metabolite Drought Stress Responses in Switchgrass (*Panicum virgatum*)," *New Phytologist* **236**, 1393–408.

Tiedge, K., et al. 2022b. "Foxtail Mosaic Virus-Induced Gene Silencing (VIGS) in Switchgrass (*Panicum virgatum* L.)," *Plant Methods* **18**, 71.

**Funding Information:** Financial support for this work was provided by the U.S. Department of Energy (DOE) Early Career Research Program (DE-SC0019178, to PZ), the German Research Foundation (DFG) Research Fellowship (TI 1075/1-1, to KT), and the DOE Joint Genome Institute (JGI) DNA Synthesis Science Program (grant #2568, to PZ). The gene synthesis work conducted by the U.S. Department of Energy JGI, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

## MICROBIAL CONVERSION University Projects

#### Early Career

## The Use of Deuterated Water as a Substrate-Agnostic and Cost-Effective Isotope Tracer for Investigating Reversibility and Thermodynamics of Reactions in Central Carbon Metabolism

Melanie M. Callaghan<sup>1,2</sup>, Eashant Thusoo<sup>1,2</sup>, Bishal D. Sharma<sup>2,4</sup>, David M. Stevenson<sup>1,2</sup>, Daniel G. Olson<sup>2,4</sup>, Lee R. Lynd<sup>2,4</sup>, **Daniel Amador-Noguez<sup>1,2,3\*</sup>** (amadornoguez@ wisc.edu)

<sup>1</sup>University of Wisconsin–Madison; <sup>2</sup>Center for Bioenergy Innovation (CBI); <sup>3</sup>Great Lakes Bioenergy Research Center; <sup>4</sup>Dartmouth College

https://danlab.bact.wisc.edu/

**Goals:** Integrate advanced mass spectrometry, computational modeling, and metabolic engineering to develop an experimental-computational approach for the *in vivo* genomescale determination of Gibbs free energies ( $\Delta G$ ) in metabolic networks suitable for high-throughput thermodynamic profiling of engineered organisms and emerging model systems.

Successful manipulation of microbial systems for biotechnology applications requires a quantitative understanding of their metabolism. Stable isotope tracers (e.g., <sup>13</sup>C, <sup>15</sup>N, <sup>18</sup>O, and <sup>2</sup>H tracers), in combination with mass-spectrometry– based metabolomics, have become a widely used tool for the quantitative analysis of metabolism. Steady-state and dynamic isotope tracer experiments can provide information on metabolic network structure, metabolic fluxes, and thermodynamics of metabolic reactions and pathways.

The use of isotope tracers to estimate  $\Delta G$  of reactions in central carbon metabolism constitutes a recent development. Rather than relying on measurements of product and reactant concentrations, this approach estimates  $\Delta G$  from forward ( $J^+$ ) and backward ( $J^-$ ) reaction fluxes via the relation  $\Delta G = -RT \ln(J^+/J^-)$ . The measurement of forward-to-backward  $J^+/J^-$  ratios can be performed using <sup>2</sup>H-labeled and <sup>13</sup>C-labeled tracers and relies on the generation of distinctive metabolite labeling patterns by reversible reactions within a pathway. Both the measurement of metabolic fluxes and the estimation of *in vivo*  $\Delta G$  of reactions have been previously accomplished by placing the heavy isotopes—most commonly <sup>13</sup>C, <sup>15</sup>N, or <sup>2</sup>H—into nutrient substrates such as <sup>13</sup>C or <sup>2</sup>H-labeled sugars, <sup>2</sup>H-labeled fatty acids, <sup>15</sup>N-labeled amino acids, <sup>15</sup>N-labeled ammonia, <sup>13</sup>C-labeled CO<sub>2</sub> or formic acid, and many others.

For some applications, the use of isotopically labeled nutrient substrates may not be feasible due to availability and/ or high cost. One salient example of this is metabolic flux analysis in industrially relevant cellulolytic microbes, such as Clostridium thermocellum, that metabolize complex substrates (i.e., lignocellulosic biomass). In this work, the use of deuterated water (also named as "heavy water," <sup>2</sup>H<sub>2</sub>O, or  $D_2O$ ) is explored as a non-nutrient, substrate-agnostic, cost-effective isotope tracer for investigating reversibility of reactions in central carbon metabolism. Researchers reasoned that the use of  ${}^{2}\text{H}_{2}\text{O}$  as a tracer (i.e., by growing bacteria in culture media containing a defined amount of  ${}^{2}H_{2}O$ can provide information on reversibility of dehydration/ hydration reactions, isomerization reactions, aldol reactions, and transamination reactions that result in the incorporation of protons from water into C-H bonds, and thus allow characterization of pathway thermodynamics. The team reports the successful use of deuterated water to investigate the reversibility of glycolytic reactions on three bacterial species of industrial interest: the model bacterium Escherichia coli, the cellulolytic and ethanologenic bacterium C. thermocellum, and the ethanologenic bacterium Zymomonas mobilis, each harboring distinct versions of glycolysis. This work will aid in the construction of accurate metabolic models that incorporate thermodynamic constraints and guide fast rational engineering of microbial networks.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program, Early Career Research Program under Award Number DE-SC0018998. This work was also supported by the Center for Bioenergy Innovation (CBI), from the U.S. Department of Energy Bioenergy Research Centers supported by the BER Program in the DOE Office of Science.

## Probing Lignin Deconstruction and Catabolism in Soil *Pseudomonas* Species

Beth DiBiase<sup>1</sup>, Nanquing Zhou<sup>1</sup>, Rebecca A. Wilkes<sup>2</sup>, Jacob Waldbauer<sup>3</sup>, Gregg T. Beckham<sup>2</sup>, Allison Z. Werner<sup>2</sup>, Neha P. Kamat<sup>1</sup>, **Ludmilla Aristilde<sup>1\*</sup>** (ludmilla.aristilde@ northwestern.edu)

<sup>1</sup>Northwestern University; <sup>2</sup>National Renewable Energy Laboratory; and <sup>3</sup>University of Chicago

**Goals:** The overall goal of this project is to elucidate the metabolic reaction networks within outer membrane vesicles (OMVs) secreted from soil *Pseudomonas* species. In particular, this project aims to evaluate how OMVs catabolize

lignin-derived aromatics in *Pseudomonas* strains, and in turn to maximize aromatic catabolic activity via engineered or synthetic systems. The results from this work will enhance understanding of carbon cycling by soil bacteria and have implications in the use of engineered pseudomonads for lignin valorization to value-added compounds to support the bioeconomy.

Valorization of lignin is an important component of a sustainable bioeconomy. Gram-negative soil *Pseudomonas* strains, which natively catabolize lignin-derived aromatics (LDAs), are commonly engineered for the conversion of LDAs to value-added compounds. It was recently shown that *Pseudomonas putida* secretes OMVs enriched with enzymes that catalyze LDA turnover (Salvachúa et al. 2020). However, the metabolic reaction networks of pseudomonad OMVs remain uncharacterized.

Towards characterizing the regulatory controls and bottlenecks of OMV-localized fluxes from LDAs, the team first characterized a method for OMV isolation and enumeration from Pseudomonas cultivations and applied this to compare OMV secretion rates across growth conditions and stages. For OMV isolation, affinity-based kits were compared to ultracentrifugation (UC). Nanoparticle tracking analysis (NTA) was used both to count and measure the size of OMVs. Affinity-based isolation had improved throughput and lower processing time, but lowered particle yields. The isolation method did not have an effect on OMV size distribution, and OMV secretion was determined at different growth stages during growth on hydroxycinnamic acid, p-coumarate, or nutrient-rich medium. Proteomics analysis is underway to evaluate whether isolation strategies result in selective enrichment of certain OMV populations. The objective is to identify optimal sampling strategy for P. putida cultivations on p-coumarate.

Regarding OMV function, researchers hypothesized that the differential abundance of enzymes packaged into OMVs from P. putida fed on different LDAs will give rise to OMVs with different reaction networks. To test this hypothesis, the OMV metabolic functionality is being characterized for a variety of LDAs using OMVs produced by cultivation in lignin-rich media prepared with alkaline pretreated lignin liquor. After addition of an LDA and cofactors (i.e., ATP and NAD(P)H) were added to purified OMVs, preliminary metabolic profiling was conducted. Metabolic intermediates such as 4-hydroxybenzoate and protocatechuate were identified, demonstrating OMVs are actively catabolizing the LDA substrate. High-resolution kinetic profiling and kinetic <sup>13</sup>C-labeling experiments with several LDAs will be performed next to both obtain direct evidence of and to quantify the metabolic functionality of OMV-localized reaction networks.

Lastly, quantification of metabolic functionalities in the OMVs may reveal metabolic bottlenecks within the pathways for LDA catabolism. Overcoming these bottlenecks would be of interest in engineering improved biocatalysts for lignin valorization. However, genetic tools for OMV biogenesis and cargo packaging are not currently available in P. putida. To this end, a library of genetic mutants has been screened to identify mutations that induce vesiculation but do not significantly impact growth on LDAs. Current work is focused on developing a SpyCatcher-SpyTag system for targeting protein cargo into OMVs. Protocols for proteomics analysis of OMV preparations from wild-type and mutant strains are under refinement and will enable assessment of aberrant protein cargo sorting in hypervesiculating mutants. The project aims to enable OMV deployment as a standardized synthetic biology tool in pseudomonads.

Salvachúa, D., et al. 2020. "Outer Membrane Vesicles Catabolize Lignin-Derived Aromatic Compounds in *Pseudomonas putida* KT2440," *Proceedings of the National Academy of Sciences* **117**, 9302–10.

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

## Optogenetic Control of a Dual Yeast-Yeast Consortia for Chemical Production

Sergio A. Garcia Echauri\* (sechauri@princeton.edu), Saurabh Malani\* (smalani@princeton.edu), **José L. Avalos** 

Princeton University

Goals: The goal of this project is to develop optogenetic tools and applications-the use of light-responsive proteins to modulate biological processes-for the control of microbial consortia for biofuel and chemical production. Researchers have developed novel optogenetic circuits to control growth rates in several strains of yeast and bacteria; this allows the team to not only stabilize microbial consortia with light, but also optimize their population ratios for chemical production. The team will develop light-controlled co-culture fermentations and use mathematical models and feedback controls to advance basic understanding of these biological systems and optimize them for growth rate and chemical production. These technologies constitute a new paradigm for the engineering and control of microbial consortia, which could help to realize their promise for biofuel and chemical production.

Microbial co-culture fermentations can improve the production of chemicals and biofuels over single-strain fermentations; optimizing and segregating production modules among the consortia members can lower the metabolic burden from overexpression of metabolic enzymes (Zhou et al. 2015). Dynamically tuning the consortia composition is integral in optimizing multistep production processes, in which growth and production phases are uncoupled and therefore different for each consortia member. Optogenetics has enhanced the ability to seamlessly control gene expression with the input of light. Light as a gene inducer has many advantages when compared to traditionally used chemical inducers: it's inexpensive, can be applied and removed instantly, is highly tunable, is active in different media compositions, and has minimal cellular side-effects. The lab has developed optogenetic tools to control gene expression in yeast with blue light (Lalwani et al. 2021). Using this system, researchers engineered two Saccharomyces cerevisiae strains with opposite growth phenotypes, one requiring blue light to grow and stops growing in darkness, while the other requires the absence of blue light (darkness) to grow and does not grow under blue light. Using these strains, researchers demonstrate the control of synthetic yeast-yeast consortia to achieve desired set-points of cell densities in batch and continuous culture conditions.

Lalwani, M. A., et al. 2021. "The *Neurospora crassa* Inducible Q System Enables Simultaneous Optogenetic Amplification and Inversion in *Saccharomyces cerevisiae* for Bidirectional Control of Gene Expression," *ACS Synthetic Biology* **10**, 2060–75.

Zhou, K., et al. 2015. "Distributing a Metabolic Pathway Among a Microbial Consortium Enhances Production of Natural Products," *Nature Biotechnology* **33**, 377–83.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, Award Number DE-SC0022155.

#### Early Career

## Lipid Membrane Remodeling and Metabolic Response During Ethanol and Isobutanol Stress in *Zymomonas mobilis*

Julio Rivera Vazquez<sup>1,2\*</sup> (riveravazque@wisc.edu), Edna Trujillo<sup>1,2</sup>, Jonathan Williams<sup>1,2</sup>, Fukang She<sup>2</sup>, Melanie M. Callaghan<sup>1,2</sup>, Joshua J. Coon<sup>1,2</sup>, **Daniel Amador-Noguez<sup>1,2</sup>, Timothy J. Donohue<sup>1,2</sup>** 

<sup>1</sup>Great Lakes Bioenergy Research Center; <sup>2</sup>University of Wisconsin–Madison

#### https://danlab.bact.wisc.edu/

**Goals:** *Zymomonas mobilis*, an ethanologenic gram-negative bacterium, is currently being bioengineered to produce isobutanol. However, it has been observed that exposure to isobutanol elicits detrimental physiological changes, including a reduction in growth rate and glucose consumption. This project aims to systematically investigate the physiological response of *Z. mobilis* to isobutanol with a particular

emphasis on changes in lipid membrane composition and proteome allocation.

Despite being a proficient ethanol producer, Z. mobilis experiences growth inhibition at high ethanol titers and is highly sensitive to isobutanol. It is known that bacteria can modulate lipid membrane composition to increase their tolerance to environmental stressors. In this study, researchers used liquid chromatography-mass spectrometry (MS)/ MS-based lipidomics to measure changes in lipid membrane composition that occur when Z. mobilis is exposed to increasing concentrations of ethanol and isobutanol. Exposure to ethanol and isobutanol resulted in significant but distinct changes to the lipid and fatty acid composition. Affected lipid classes included cardiolipins, phosphatidylcholines, and phosphatidylethanolamines. The fatty acid composition was also significantly affected. Most notably, a substantial increase in C19 cyclopropane fatty acid content was observed when cells were grown at high ethanol concentrations, suggesting that the changes comprise a defense mechanism in response to solvent stress. Previous evidence showed that cyclopropane-ringed fatty acids modify membrane fluidity and act as a barrier to prevent detrimental molecules from entering the cell. To test the hypothesis that  $C_{19}$  cyclopropane fatty acids and derived lipids contribute to solvent resistance in Z. mobilis, researchers engineered a strain that overexpressed the cyclopropane fatty acyl synthase (CFA synthase) protein (ZMO1033) responsible for transforming unsaturated fatty acids into cyclopropane fatty acids. Analysis of the lipid membrane composition of the CFA synthase overexpressing strain showed a significant increase in C19 cyclopropane fatty acid content for all lipid classes. This increase correlated with significantly improved growth rates in the presence of high ethanol and isobutanol concentrations. These data demonstrate the importance of cyclopropane fatty acids to solvent stress resistance and the effects of isobutanol on protein activity in Z. mobilis. This data will allow engineering of strains that are more resistant to high ethanol and isobutanol concentrations.

**Funding Information:** This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program under Award Number DE-SC0018409.

## Optimizing Enzymes for Plastic Upcycling Using Machine Learning Design and High-Throughput Experiments

Benjamin Fram<sup>1,2\*</sup> (fram@g.harvard.edu), Elizabeth Bell<sup>3</sup>, Gregg T. Beckham<sup>3</sup>, Richard Brizendine<sup>4</sup>, Karla Ilic Durdic<sup>4</sup>, Perry Ellis<sup>4</sup>, Japheth Gado<sup>3</sup>, Erika Erickson<sup>3</sup>, Samuel Lim<sup>2</sup>, Maria Llases<sup>3</sup>, Debora Marks<sup>2</sup>, Natasha Murphy<sup>3</sup>, Brenna Norton-Baker<sup>3</sup>, Nathan Rollins<sup>2</sup>, Chris Sander<sup>1,2</sup>, Nicole Thadani<sup>2</sup>, David Weitz<sup>4</sup>, Allison Z. Werner<sup>3</sup>, Wentao Xu<sup>4</sup>, Xinge Zhang<sup>4</sup>, **Nicholas P. Gauthier<sup>1,2</sup>** 

<sup>1</sup>Dana-Farber Cancer Institute; <sup>2</sup>Harvard Medical School; <sup>3</sup>National Renewable Energy Laboratory; <sup>4</sup>Harvard University

**Goals:** This project aims to create new and optimized polyethylene terephthalate (PET)–depolymerizing enzymes (PETases) useful for industrial application.

- Design novel PETases that are significantly different (25-65+ mutations) from known PET-depolymerizing enzymes and contain unique properties useful for performant enzymatic PET recycling and upcycling. Introducing many simultaneous mutations, while maintaining function, will enable researchers to more efficiently search for altered properties that depend on primary amino acid sequence.
- Optimize previously described PETases by testing millions of mutagenized variants using directed evolution. Starting with existing functional PETases and exploring small changes in many distinct sequences using a novel ultra-HTP functional assay, researchers will optimize enzymes with improved properties by varying experimental conditions.
- Characterize performance metrics of new and optimized PETases in detail including solvent tolerance, stability, catalytic rate, and substrate promiscuity.

Plastic use is ubiquitous in the modern world, and PET is one of the most abundantly produced plastics (and the most highly produced polyester), with ~65 million metric tons manufactured annually. To the consumer, PET is likely most recognizable as the plastic used to make beverage bottles. Like many plastics, traditional mechanical or chemical means of PET deconstruction and upcycling are costly and inefficient.

Recently, biological enzymes capable of breaking down PET into its basic building blocks (terephthalic acid and ethylene glycol) have garnered significant attention as an attractive means of dealing with the plastic problem. These enzymes are currently undergoing pilot studies for implementation in enzyme-based recycling. However, there are significant limitations to current enzymes, including the need to perform costly pre-processing of the plastic waste before the enzymes are able to work. Further optimization of these enzymes is necessary to make the process profitable and thereby incentivize commercialization of this biology-based green recycling technology.

This project aims to apply recent advances in artificial intelligence and machine learning to design new versions of enzymes capable of breaking down PET. Based on preliminary experiments using this evolution-informed computational design strategy, researchers believe it is possible to create a highly diverse set of enzymes that have exceptional properties useful for industrial recycling. Testing these enzymes is typically labor intensive but using a new robotically enabled platform researchers will be able to experimentally characterize key enzymatic properties of thousands of these designed enzymes.

In addition to applying machine learning approaches to design new enzymes, the team has developed a novel method that, by experimentally testing millions of small changes to enzyme structure, enables optimization of existing enzymes that are known to break down plastic. The key to this approach is to encapsulate individual variations of each enzyme in single droplets together with plastic nanoparticles creating a mini reaction, and then select those droplets, which successfully break down PET to isolate the winning enzyme variants.

Ultimately, the result of these studies will be the discovery of highly optimized enzymes capable of breaking down PET plastics in an industrial recycling setting, enabling a powerful and green solution to the plastic problem.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022024.

## Metabolic Modeling and Genetic Engineering of Enhanced Anaerobic Microbial Ethylene Synthesis

Sarah, J. Young<sup>1</sup>, Adrienne B. Narrowe<sup>3</sup>, Kamal M. Deep<sup>1</sup>, Elizabeth Morgan<sup>1</sup>, Rebecca Daly<sup>3</sup>, Mikayla A. Borton<sup>3</sup>, Ethan King<sup>2</sup>, Kelly C. Wrighton<sup>3</sup>, William R. Cannon<sup>2</sup>, **Justin A. North<sup>1\*</sup>** (north.62@osu.edu)

<sup>1</sup>The Ohio State University; <sup>2</sup>Pacific Northwest National Laboratory; <sup>3</sup>Colorado State University

**Goals:** To develop robust and optimized anaerobic ethylene pathways in photosynthetic and lignocellulosic bacteria for high-yield conversion of renewable  $CO_2$  and lignocellulose into bioethylene. This will be accomplished by:



**Fig. 8.** Anaerobic ethylene cycle for microbial synthesis of ethylene. Optimal enzymes for reactions indicated by (\*) are identified by gene synthesis and screening. The ethylene pathway plus central carbon metabolism and DNA/RNA/protein synthesis is being assessed by physics-based modeling to identify metabolic bottlenecks of burdens. [Courtesy The Ohio State University]

(1) First, bioinformatically mining and experimentally screening methylthio-alkane reductase homologs, SAM hydrolase homologs, and alcohol dehydrogenase homologs from cultivated and uncultivated organisms for functional enzymes that enhance ethylene yields.

(2) Next, constructing and employing predictive systemslevel models of ethylene production. Using a physicsbased *R. rubrum* model, researchers predict enzymes that participate in competing or supporting pathways and are thus targets for selection studies to increase ethylene yields.

(3) Finally, researchers metabolically engineer bacteria for enhanced, sustained ethylene production from  $CO_2$  and lignocellulose. Researchers assemble the best-performing genes under control of optimized active transcription elements on a modular DNA fragment in a combinatorial manner with guidance from predictive models.

Previously, researchers have detailed a pathway in phototrophic bacterium, *Rhodospirillum rubrum* that produces ethylene in the absence of oxygen from methionine and ATP (Fig. 8, this page; North et al. 2020). Traditional ethylene production involves energy intensive cracking of petroleum fossil fuels to meet the 300 million metric ton annual demand. Thus, a sustainable microbial platform for the renewable production of ethylene is urgently needed. The goal of this project is to optimize this anaerobic ethylene production pathway, as outlined above. Enzyme Screens: Experimentally verified amino acid sequences for each reaction were queried against Uniref, KEGG, GenBank, and JGI-IMG databases using an e-value cutoff of 1e-10. Sequences were also searched using BLAST (e-value cutoff of 1e-10) in a curated set of assembled genome databases spanning soils, rivers, and the human gut. Combined sequences from all databases were dereplicated using CD-HIT to remove identical sequences. Methylthio-alkane reductase genes were synthesized by the JGI DNA synthesis program, and SAM hydrolases and alcohol dehydrogenases synthesis in process. Insertion of the methylthio-alkane reductase genes into an R. rubrum native methylthio-reductase deletion strain revealed sequences from closely related alphaprotebacteria like Rhodoblastus sphagnicola increase ethylene synthesis from 2-methylthioethanol up to twofold. Furthermore, these screens revealed sequences from clostridia, bacilli, negativicutes, and fibrobacter species that are also methylthioalkane reductases, expanding the understanding of the diversity of organisms that can synthesis hydrocarbons from volatile organic sulfur compounds. In parallel to these enzyme screens, researchers have further enumerated the number of substrates and reactions catalyzed, revealing that propane, propylene, and butane can also be synthesized.

**Physics-based Modeling:** In order to ensure that the model of cell metabolism is complete enough to capture the important dynamics for ethylene production, researchers

iteratively built and tested the model on four photoheterotrophic growth conditions: growth on fumarate, malate, acetate and ethanol. Testing and refining the model on a variety of growth conditions helps to ensure that the principles discovered regarding the metabolic processes are general and not artifacts of an incomplete model. Researchers have now tested the model with each growth condition and a range of redox states and compared the results to various experimental studies. Summarizing the results, the model predicts that flux through phosphoenolpyruvate carboxy kinase (PEP-CK) during growth on malate and fumarate is thermodynamically feasible to synthesize phosphoenol-pyruvate in lower glycolysis, as observed experimentally (McCully and McKinlay 2016). However, the model indicates that the reductive TCA (rTCA) cycle is the thermodynamically favored mechanism for pyruvate synthesis to support gluconeogenesis, amino acid, and de novo nucleotide synthesis. In contrast, during growth on acetate and ethanol, the thermodynamically preferred route to synthesize pyruvate utilizes the ethylmalonyl-CoA pathway for acetyl-CoA synthesis followed by reductive carboxylation to pyruvate by pyruvate ferredoxin oxidoreductase. To predict ethylene production, the computational model has been analyzed with metabolic control analysis (concentration control coefficients) and with computational protein knock-down (reduced pathway activity) and overexpression (increased pathway activity) to predict which enzymes need to be up or down regulated to enhance ethylene production. Initial results have had very good agreement with parallel experimental work.

Metabolic Engineering: In Rhodospirillum rubrum at least 6 ethylene pathway enzymes are under transcriptional regulation of SalR, a member of LysR family transcription factors. SalR and thus ethylene synthesis activity is sensitive to available exogenous sulfur, such that ethylene synthesis is turned off at sulfur concentrations  $\geq 200~\mu M.$  Instead of replacing the promoter for each gene, researchers are taking a more efficient approach by engineering SalR expression and activity to activate ethylene synthesis. When SalR is constitutively expressed the engineered strain continuously produces ethylene, even in presence of  $\geq$  500 uM of exogeneous sulfate and methionine. However, exogenous cysteine still inhibits ethylene similar to when SalR is expressed from its native promoter suggesting cysteine is a SalR effector molecule. Through predictive structural modeling (Alpha Fold) researchers have identified residues for substitution studies.

McCully A. L., and J. B. McKinlay 2016. "Disrupting Calvin Cycle Phosphoribulokinase Activity in *Rhodopseudomonas palustris* Increases the  $H_2$  Yield and Specific Production Rate Proportionately," *International Journal of Hydrogen Energy* 41(7), 4143–9.

North, J. A., et al. 2020. "A Nitrogenase-like Enzyme System Catalyzes Methionine, Ethylene, and Methane Biogenesis," *Science* 369(6507), 1094–8. DOI:10.1126/science.abb6310. **Funding Information:** This project is supported by the U.S. Department of Energy's Office of Biological and Environmental Research under contracts DE-SC0022091 and DE-SC0019338 (Ohio State University) and 78266 (PNNL).

## Cell-Free Conversion of Pyruvate to 2,3-Butanediol Using Co-Substrate Feed as pH Control Strategy

Bilal Jilani<sup>1</sup>, Wheaton L. Schroeder<sup>2</sup>, Costas D. Maranas<sup>2,3</sup>, **Daniel G. Olson<sup>1,3</sup>** 

<sup>1</sup>Dartmouth College; <sup>2</sup>The Pennsylvania State University; <sup>3</sup>Center for Bioenergy Innovation

Cell-free systems are promising tools for production of chemicals and to improve understanding of biochemical pathways. Researchers are interested in developing well-characterized modules whose behavior can be mathematically predicted, allowing them to be combined into larger systems either as production modules or indicator reactions. The conversion of pyruvate to 2,3-butanediol is a good model system due to interest in 2,3-butanediol as a commodity chemical, thermodynamic favorability of all reaction steps, and prior demonstration of high titer production both *in vitro* and *in vivo*.

In the present work, the team demonstrate development of a high-performance pyruvate to 2,3- butanediol conversion system. Researchers start by characterization of individual enzymes in the pathway and demonstrate that none of them are subject to significant inhibition by substrates or products. Researchers then demonstrate conversion of 1063.1 (±19.0) mM (~93.7 g/L) acetoin to 1017.3 (±2.0) mM  $(\sim 91.7 \text{ g/L})$  2,3-butanediol, which represents 95.7% of the theoretical maximum yield, high titer and yield using a 2-enzyme system consisting of butanediol dehydrogenase and formate dehydrogenase. Team members subsequently extended the system to allow conversion of pyruvate to 2,3-butanediol using a 4-enzyme system. Researchers were able to convert 2045.5 (±30.7) mM (~225.0 g/L) pyruvate to 929.1 (±20.4) mM (~83.7 g/L) 2,3-butanediol, which represents 90.8% of the theoretical maximum yield. Achieving high titer production required careful attention to proton recycling. Further increases to product titer were limited by experimental limitations (e.g., substrate solubility, foaming due to gas formation) rather than intrinsic limitations of the enzymatic pathway.

The team subsequently developed mechanistic kinetic models for each enzyme individually and showed that these models (1) can be combined to predict the behavior of the 4-enzyme system, or (2) can be used to predict targeted

modifications to minimize enzyme concentration (while maintaining the overall conversion rate) or to minimize concentration of a particular metabolic intermediate.

#### Early Career

## Genomic Analyses and Enzyme Characterization Provide Insights into the Catabolism of Lignin-Related Aromatic Compounds in White-Rot Fungi

Teeratas Kijpornyongpan,<sup>1</sup> Eugene Kuatsjah,<sup>1</sup> Alexa Schwartz,<sup>1</sup> Allison Yaguchi,<sup>1</sup> Sean Woodworth,<sup>1</sup> Michael Zahn,<sup>2</sup> **Davinia Salvachúa<sup>1\*</sup>** (davinia.salvachua@nrel.gov)

<sup>1</sup>National Renewable Energy Laboratory; <sup>2</sup>University of Portsmouth, United Kingdom

**Goals:** The overall goal of this project is to test the hypothesis that white-rot fungi (WRF) can simultaneously depolymerize lignin extracellularly and catabolize depolymerization products intracellularly as carbon and energy sources. The results from this project will lead to improved understanding of lignin utilization by WRF and enable identification of promising fungal strains and enzymes for lignin catabolism and valorization. To date, researchers have confirmed the utilization of lignin-related compounds as carbon sources by two WRF species and have initiated pathway elucidation via systems biology approaches. To continue this effort, the team is using enzymology approaches and comparative genomic and phylogenetic analyses to validate and broaden the knowledge on the catabolism of lignin-related compounds by WRF.

Plant-derived biomass is the most abundant biogenic carbon source on Earth. Despite this abundance of lignocellulose, only a small clade of organisms known as WRF can efficiently break down both the polysaccharide and lignin components of plant cell walls. This unique ability imparts a key role for WRF in global carbon cycling. To date, research on WRF has almost universally and intensely focused on their extracellular enzymes that depolymerize plant polymers, whereas knowledge of their intracellular metabolism remains underexplored (Kijpornyongpan et al. 2022). This project aims to elucidate intracellular pathways in WRF with a particular focus on aromatic catabolic pathways. Recently, the team confirmed the utilization of two lignin-related compounds (e.g., 4-hydroxybenzoic acid and vanillic acid) as carbon sources in two species of WRF (*Trametes versicolor* and *Gelatoporia subvermispora*). Additionally, the team proposed a catabolic route via the hydroxyquinol pathway from 4-hydroxybenzoic acid to central carbon metabolism, which was informed by differential

transcriptomic, proteomic, and metabolomic analyses (del Cerro et al. 2021). This pathway has been considerably less studied than the b-ketoadipate pathway, which is well known in aromatic catabolic bacteria to convert 4-hydroxybenzoic acid to the tricarboxylic acid cycle but which includes different catabolic intermediates. Based on this discovery, researchers were motivated to conduct mechanistic pathway validation and to seek better understanding of the distribution and prevalence of these pathways in WRF and the fungal kingdom generally.

For the enzyme validation work, team members focused on three main proposed steps in the catabolism of 4-hydroxybenzoic acid: oxidative decarboxylation, hydroxylation, and ring-cleavage (del Cerro et al. 2021). To date, five enzymes have been validated via *in vitro* approaches, and two additional enzymes have also been validated *in vivo* in a bacterial host. The former five enzymes have been purified, and enzymology analyses have been conducted to understand substrate and cofactor preference as well as inhibitory mechanisms in the presence of various lignin-related compounds. Structural biology efforts are also underway.

In parallel to its enzyme validation efforts, the team sought to understand the distribution of these catabolic activities in WRF and other microbes. For that purpose, researchers performed a large-scale comparative genomic and phylogenetic study across the bacterial and fungal kingdoms. Team members selected protein domains related to specific catabolic activities and sampled 255 bacterial genomes and 317 fungal genomes. Researchers have shown that some of these enzymes are highly conserved in certain fungal lineages and that substrate specificity of aromatic ring-cleavage enzymes has expanded during fungal evolution. In addition, the abilities to depolymerize lignin and catabolize ligninrelated aromatic compounds seem to be independent. Lastly, these analyses have also revealed a series of unique enzymes from WRF that may broaden the spatial location of some catabolic reactions with aromatic compounds. These new enzymes are undergoing biochemical characterization. Overall, these studies will provide a deeper understanding of carbon turnover during wood decay and discover enzymes and pathways that could be exploited to convert the undervalued biopolymer lignin into value-added compounds.

del Cerro, C., et al. 2021. "Intracellular Pathways for Lignin Catabolism in White-Rot Fungi," *Proceedings of the National Academy of Sciences* **118**, e2017381118.

Kijpornyongpan, T., et al. 2022. "Systems Biology–Guided Understanding of White-Rot Fungi for Biotechnological Applications: A Review," *iScience* **25**, 104640.

**Funding Information:** This research is supported by the U.S. Department of Energy (DOE) Office of Science, Office of Biological and Environmental Research under the Early Career Award Program. This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36- 08GO28308.

## Improving Bioprocess Robustness by Cellular Noise Engineering

Georgios Daletos<sup>1\*</sup> (daletosg@mit.edu), Irum Perveen,<sup>1</sup> Andreas Vasdekis,<sup>2</sup> **Gregory Stephanopoulos** 

<sup>1</sup>Massachusetts Institute of Technology <sup>2</sup>University of Idaho–Moscow

**Goals:** The overall goal of this project is to enhance the robustness of biofuel production in adverse and fluctuating environments, such as media containing toxic hydrolysates, by introducing cellular noise engineering as a means of improving the production. The approach involves the identification of factors in the transcription process that increase cellular noise and the deployment of such factors to generate cells with increased noise. Researchers use modeling and a single-cell analysis workflow to engineer *Yarrowia lipolytica* variants that can tolerate, grow, and efficiently synthesize biofuel precursors under steady state, albeit stressful, conditions. Overall, the team anticipates that strains with optimal levels of cellular noise will also exhibit robustness that maintains production under time-varying stresses.

Robustness represents a system-level trait that allows cell populations to maintain function under adverse and fluctuating environments. When observed at the cellular or subcellular level, an isogenic cell population exhibits increased cell-to-cell variability, or noise, even under steady-state conditions. In this context, isogenic cells undergo division of labor with some expressing the pathways that enable them to continue functioning in the new environment. This concept guides this project in developing workflows for introducing and manipulating cellular noise to enhance cellular tolerance to environmental stressors. The focus has been placed on the construction of Yarrowia lipolytica strains with the double-phenotype of tolerance and high lipid productivity. In the team's first steps on cellular noise engineering, researchers refined gene editing toolboxes that can deterministically vary the level of cellular noise in protein expression levels. In this context, the team followed an approach previously demonstrated in Saccharomyces cerevisiae for designing a synthetic promoter library in which increasing numbers of transcription factor binding sites led to enhanced noise levels (Sharon et al. 2014). Similarly, the team introduced three to five tandem upstream activating sequences to the erythritol-inducible pEYK1 promoter. The synthetic hybrid promoters were placed upstream of a red fluorescent protein, fused into plasmids, and stably integrated into the genome of Y. lipolytica. In the induction

experiments, the team independently varied the concentration of erythritol and glucose to determine the relationship between expression level and noise. Each transformant bearing the erythritol-inducible pEYK1 promoter fused to the upstream activating sequences was screened separately by flow cytometry. The results were categorized into expression and noise levels and compared to those of parental strains. As a next step, researchers introduced key genes that play a significant role in viability at varying inhibitor levels. To this end, rational design was applied to develop a cellulosic oil *Y. lipolytica* strain that is tolerant to the primary lignocellulosic inhibitor furfural. To enable tolerance to furfural, researchers constructed Y. lipolytica overexpressing an evolved reductase enzyme (GRE2<sup>evol</sup>), which was directly obtained from prior work with S. cerevisiae (Lam et al. 2021), or an endogenous aldehyde dehydrogenase that converts furfural to the less toxic furoic acid. The team finally evaluated front-runner *Y. lipolytica* strains under both stressful and non-stressful conditions to quantify the effects of noise and expression levels on furfural tolerance.

Lam, F. H., et al. 2021. "Engineered Yeast Tolerance Enables Efficient Production from Toxified Lignocellulosic Feedstocks," *Science Advances*, **7**(26), eabf7613.

Sharon, E., et al. 2014. "Probing the Effect of Promoters on Noise in Gene Expression Using Thousands of Designed Sequences," *Genome Research* **24**(10), 1698–1706.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0022016.

## Harnessing Robustness of Thermophilic Bacillus coagulans for Conversion of Switchgrass Hydrolysates to Designer Bioesters at Elevated Temperatures

Seunghyun Ryu<sup>1\*</sup>, Jackson Edwards<sup>2</sup>, David Dooley<sup>1</sup>, Khanh Ha<sup>1</sup>, Richard J. Giannone<sup>3</sup>, Bruce S. Dien<sup>2</sup> (bruce.dien@usda.gov), **Cong T. Trinh<sup>1</sup>** (ctrinh@utk.edu)

<sup>1</sup>University of Tennessee; <sup>2</sup>National Center for Agricultural Utilization Research, USDA-ARS, Peoria, IL; <sup>3</sup>Oak Ridge National Laboratory

**Goals:** To fundamentally understand and redirect metabolism and regulation of thermophilic *Bacillus coagulans* for the efficient conversion of undetoxified lignocellulosic biomass hydrolysates into designer bioesters.

*B. coagulans* is a gram-positive thermophilic bacterium that is capable of growing at elevated temperatures, utilizing biomass hydrolysates, and producing lactate at high levels. Researchers aim to understand and harness its robustness for conversion of biomass hydrolysates to designer bioesters (e.g., acetate esters, lactate esters) that have broad use as solvents, flavors, fragrances, and advanced biofuels. Through comprehensive screening of diverse undomesticated B. coagulans strains isolated from different environmental niches against a wide range of temperatures (30° to 60°C), either single or mixed C5/C6/C12 sugars, and lactate concentrations (0 to 60 g/L), researchers found that most of the strains could grow in all of the environments tested with different degrees of robustness. Some candidates could utilize all sugars with minimal exhibition of diauxic growth, grow optimally at 55°C, and tolerate high concentrations of lactate up to at least 40 g/L, which serve as reference strains for elucidating cellular robustness and metabolic engineering. Genome sequencing and proteomics of the representative strain UT-1 showed that it has a 10% larger genome than those reported in literature and exhibited metabolic capability to utilize sugars in biomass hydrolysates and produce lactate. To rewire the metabolism of these novel undomesticated B. coagulans strains to overproduce designer bioesters, researchers have been developing genetic tools (e.g., antibiotic selection, plasmid compatibility, transformation, and gene expression) to manipulate them. The team has identified and designed thermostable enzymes to build exchangeable ester production modules compatible with B. coagulans for biosynthesis of acetate and lactate esters. Overall, B. coagulans is a promising microbial manufacturing platform that will be advanced by a fundamental understanding of its robustness and the ability to harness it for production of designer bioesters from lignocellulosic hydrolysates.

Lee, J., et al. 2021. "Probing Specificities of Alcohol Acyltransferases for Designer Ester Biosynthesis with a High-Throughput Microbial Screening Platform," *Biotechnology and Bioengineering*, 1-13.

Lee, J. and C.T. Trinh. 2021. "Controlling Selectivity of Modular Microbial Biosynthesis of Butyryl-CoA-Derived Designer Esters," *Metabolic Engineering* **69**, 262–74.

Lee, J. and C.T. Trinh. 2020. "Towards Renewable Flavors, Fragrances, and Beyond," *Current Opinion in Biotechnology* **61**, 168–80.

Lee, J. and C.T. Trinh. 2019. "Microbial Biosynthesis of Lactate Esters," *Biotechnology for Biofuels and Bioproducts* **12**(1), 226.

Seo, H., et al. 2021. "Repurposing Chloramphenicol Acetyltransferase for a Robust and Efficient Designer Ester Biosynthesis Platform," *Metabolic Engineering* **66**, 179–90.

**Funding Information:** This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award No. DE-SC0022226.

## Engineering Bacterial Microcompartments in *Clostridium autoethanogenum* to Overcome Bottlenecks in Sustainable Production of Synthetic Rubber

**Brett Palmero**<sup>1\*</sup>, Rebecca Ostien<sup>2</sup>, Nolan Kennedy<sup>1</sup>, Joanna Cogan<sup>2</sup>, Heidi Schindel<sup>2</sup>, Carolyn Mills<sup>1</sup>, Elizabeth Johnson<sup>1</sup>, Alexander P. Mueller<sup>2</sup>, Fungmin (Eric) Liew<sup>2</sup>, Michael Köpke<sup>2</sup>, **Danielle Tullman-Ercek**<sup>1</sup>

<sup>1</sup>Northwestern University; <sup>2</sup>LanzaTech

https://dtelab.northwestern.edu/research/#nanobioreactors

**Goals:** To investigate bacterial microcompartments in *Clostridium autoethanogenum* and engineer them to compartmentalize synthetic metabolic pathways.

One promising route to sustainable bioproduction of fuels and chemicals is the engineering of organisms such as acetogens to efficiently convert abundant and low-cost gases containing carbon monoxide or carbon dioxide and hydrogen to desirable, value-added products at high efficiency and low cost. This approach not only provides an avenue for repurposing greenhouse gases (GHG), but also minimizes the use of harsh chemicals and hazardous byproducts common in petroleum-based processes. However, many biochemicals are not yet produced biologically due to roadblocks in the cellular biosynthesis process. These roadblocks can include intermediate toxicity, redox imbalances, and loss of product to off-pathway reactions. In nature, these issues are often alleviated using spatial organization strategies, such as sequestration in organelles. In bacteria, such organization often occurs in protein-based organelles known as bacterial microcompartments (MCPs).

The team will investigate the native regulation, assembly, and function of MCPs in the industrially relevant nonmodel host *C. autoethanogenum*. In the *C. autoethanogenum* genome, two unique gene clusters have been identified as putative MCP operons. These putative operons contain sequences encoding possible hexamers, trimers, pentamers, and enzyme encapsulation sequences. The team tested potential inducers of these operons and found that some of these small molecules were consumed by *C. autoethanogenum*. RNAseq data show that these same small molecules transcriptionally activate the MCP operons. MCP formation in these conditions was corroborated by electron microscopy of *C. autoethanogenum*, which shows distinctive polyhedral shapes within the cells, indicative of MCP formation.

Beyond understanding the native function of these putative MCP operons, the engineering goal is to sequester key biosynthesis enzymes from two distinct metabolic pathways



**Fig. 9.** Acetyl-CoA is the hub metabolite connecting photosynthesis and fatty acid biosynthesis (FAB). The team's strategy uses systems biology and omics tools to assess flux control in upstream and downstream modules. Systems analysis of high and low FFA-producing strains will enable detection of pathway bottlenecks in each module. Optimization of the upstream module will maximize acetyl-CoA availability ("push"). Optimization of the downstream module will enhance conversion of acetyl-CoA to FFA products ("pull"). [Courtesy Jamey Young, Vanderbilt University]

into MCPs to make compounds involved in rubber production. Specifically, researchers aim to showcase the power of enzyme encapsulation in an MCP for reducing toxicity and product losses to side reactions for these pathways. Towards enabling heterologous enzyme encapsulation in these new MCP systems, 16 *C. autoethanogenum* reporter strains were generated with different putative encapsulation peptides fused to sfGFP. Fluorescence microscopy shows that 11 of these 16 sfGFP-encapsulation peptide fusions exhibit punctate fluorescence upon MCP induction indicating successful encapsulation of the fluorescent reporter within MCPs. These results will pave the way for encapsulating biosynthesis enzymes for rubber production in future years and enable the cost-efficient production of chemicals currently derived from petroleum.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0022180.

## Systems Biology to Enable Modular Metabolic Engineering of Fatty Acid Production in Cyanobacteria

Baltazar Zuniga<sup>1\*</sup> (baltazar.e.zuniga@vanderbilt.edu), Joshua P. Abraham<sup>1</sup>, Margarita Orlova<sup>2</sup>, Hawkins S. Shepard<sup>2</sup>, Jody C. May<sup>2</sup>, Yao Xu<sup>2</sup>, Maneesh Lingwan<sup>3</sup>, Juthamas Jaroensuk<sup>3</sup>, Somnath Koley<sup>3</sup>, Bo Wang<sup>1</sup>, Carl H. Johnson<sup>1, 2</sup>, John A. McLean<sup>2</sup>, Doug K. Allen<sup>3, 4</sup>, Brian F. Pfleger<sup>1</sup>, **Jamey D. Young<sup>1, 2</sup>** 

<sup>1</sup>University of Wisconsin–Madison; <sup>2</sup>Vanderbilt University; <sup>3</sup>Donald Danforth Plant Science Center; <sup>4</sup>USDA-ARS St. Louis, MO

**Goals:** The overall objective of this project is to use systems biology to identify metabolic control points and bottlenecks that regulate flux to free fatty acids (FFAs) in cyanobacteria. The central hypothesis is that cyanobacterial lipid metabolism can be modularized into pathways upstream and downstream of the nodal metabolite acetyl-CoA, which can be separately studied and optimized to enhance overall FFA production. The team plans to test its central hypothesis and accomplish the overall objective of this project by pursuing the following specific aims:

- Identify (upstream) metabolic control points regulating acetyl- CoA precursor availability. The working hypothesis is that engineering glycolytic pathways in PCC 7002 will reveal rate-controlling steps that can be manipulated to maximize acetyl-CoA availability.
- (2) Assess flux bottlenecks in the (downstream) fatty acid biosynthesis pathway. The working hypothesis is that multiomics analyses of thioesterase-expressing strains will elucidate regulatory nodes that control FFA production and overall lipid metabolism in PCC 7002.

Cyanobacteria are attractive hosts for biomanufacturing because of their ability to rapidly fix CO<sub>2</sub>, grow in nutrient-poor environments, and produce renewable chemicals directly from photosynthesis. Unlike triacylglycerol production in green algae, producing free fatty acids (FFAs) using genetically engineered cyanobacteria results in the secretion of the product into the culture medium for efficient recovery. However, an incomplete understanding of the regulation of cyanobacterial lipid metabolism limits the ability to engineer high-titer FFA-producing strains rationally. The overall objective of this project is to use systems biology to identify metabolic control points and bottlenecks that regulate flux to FFAs in the fast-growing, halotolerant Synechococcus sp. strain PCC 7002 via the modular optimization of metabolic pathways that are upstream and downstream of the nodal metabolite acetyl-CoA in a push-pull metabolic engineering strategy. Recent work has centered around the previously identified bottleneck of cyanobacterial FFA synthesis, FabH. Unexpectedly, overexpressing the kinetically superior *E. coli* FabH inhibits FFA production in a C8-producing PCC 7002 strain. Team members are using a suite of systems biology approaches, including <sup>13</sup>C flux analysis, metabolomics, lipidomics, and proteomics to investigate this and other distinctive FFA production phenotypes. These data will allow us to identify and correct metabolic bottlenecks limiting FFA biosynthesis, (i.e., pull), and optimize the carbon flux directed towards FFA synthesis, (i.e., push). This approach will provide a deeper understanding of how fatty acid flux is regulated upstream and downstream of acetyl-CoA, enabling integrated push-pull metabolic engineering strategies to produce lipid products directly from photosynthetic CO<sub>2</sub> fixation in cyanobacteria.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0022207.

## **MICROBIAL CONVERSION**

Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) programs

## Designing Novel Enzymes for Complete Degradation of Recalcitrant Polyamides

Amandeep Sangha<sup>1\*</sup> (amandeep.sangha@arzeda. com), Elise Span<sup>1\*</sup> (elise.span@arzeda.com), James Leinas<sup>1</sup>, Jacob Bale<sup>1</sup>, Christopher Voigt<sup>2</sup>, **Alexandre Zanghellini<sup>1\*</sup>** (alexandre.zanghellini@arzeda.com)

<sup>1</sup>Arzeda Corporation; <sup>2</sup>Massachusetts Institute of Technology

**Goals:** The project objectives are to (1) design enzymes capable of complete depolymerization of nylon 6 and nylon 66 and (2) engineer bacterial strains able to metabolize the degradation products to higher-value sustainable materials.

As of 2015, a total of 6.3 billion tons of plastic waste had been generated globally. It is estimated that only 9% of this total has been recycled, while 12% has been incinerated to recover energy values, and the remainder has entered landfills. New technologies are needed to address this ever-growing problem. An alternative approach, harnessing the power of biology to not just depolymerize plastics back to their monomer precursors but convert them into higher-value products, offers stronger economic incentives and in turn would be expected to drive more rapid and widespread adoption. Toward that end, this work focuses on combining cutting-edge computational protein design and synthetic biology to address the challenge of complete biodegradation and upcycling of the recalcitrant polymers nylon 6 and nylon 66.

Although natural enzymes have been shown to be capable of degrading amorphous portions of polyamides such as nylon 6 and nylon 66, complete enzymatic degradation has not been demonstrated. Researchers hypothesize this to be due in large part to a lack of natural enzymes able to efficiently catalyze degradation of the crystalline portion of the polymer. Researchers are computationally designing enzymes to alleviate this limitation by introducing and optimizing polyamide hydrolysis activity in scaffolds with open active sites. In conjunction, researchers are screening and engineering bacterial strains able to metabolize nylon 6 and nylon 66 degradation byproducts directly into central metabolism. Such platform strains can be used to produce a wide variety of fermentation products from central metabolites. Integration of nylon 6 and nylon 66 depolymerizing enzymes into these engineered hosts will provide a novel, elegant, and cost-effective consolidated fermentation process for nylon upcycling to higher-value sustainable materials.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0022077.

## SUSTAINABLE BIOENERGY

## University Projects

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

Saddie Vela<sup>1</sup>, Emily Wolf<sup>1</sup>, Jennifer Wilker<sup>2</sup>, Megan Kelly<sup>1</sup>, Saptarshi Pyne<sup>2</sup>, Sushmita Roy<sup>2</sup>, **Jean-Michel Ané<sup>2</sup>**, **Wilfred Vermerris<sup>1\*</sup>** (wev@ufl.edu)

<sup>1</sup>University of Florida; <sup>2</sup>University of Wisconsin–Madison

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

**Goals:** This project aims to understand the molecular and cellular networks controlling biological nitrogen fixation in sorghum aerial roots using a combination of genetics, synthetic bacterial communities, and systems biology.

Biological nitrogen fixation (BNF) by crop plants via microbial symbiosis is an effective approach to lowering the economic and environmental costs of crop production by decreasing fertilizer dependence. BNF is commonly associated with legumes, but cereals have been reported to be able to support nitrogen-fixing bacteria in the mucilage of their aerial roots (adventitious nodal roots), such as indigenous landraces of maize (Zea mays L.) in Oaxaca, Mexico (Van Deynze et al. 2018). Sorghum (Sorghum bicolor L. Moench) is an attractive bioenergy crop due to its ability to produce high biomass yields with minimal inputs and to withstand biotic and abiotic stresses. Some sorghum accessions can host nitrogen-fixing bacteria in the mucilage produced by their aerial roots, and this project is investigating the mechanisms enabling the symbiotic interactions with nitrogen-fixing microbes. Since this interaction relies on the presence of aerial roots, researchers have performed a genome-wide association study (GWAS) of two panels of genetically diverse sorghum genotypes, the sorghum minicore (Upadhyaya et al. 2009), a collection of landraces, and the sorghum association panel (SAP; Casa et al. 2008), a collection of sorghum genotypes representing all major cultivated races and important U.S. breeding lines and their progenitors. Traits of interest include the number of nodes producing aerial roots, the total number of aerial roots, aerial root length, and aerial root diameter. Since the traits supporting efficient BNF are hypothesized to be under genetic control but influenced by the environment, the team has also analyzed the effect of the nitrogen fertilization level and location (Florida vs. Wisconsin) on these traits. The proportion of genotypes forming aerial roots was substantially greater in the minicore than in the SAP, suggesting the

presence of aerial roots has been under negative selection in modern breeding programs. The GWAS resulted in several candidate loci associated with the number of nodes producing aerial roots consistently in both locations. The genetic analysis of segregating breeding populations derived from crosses between commercial bioenergy sorghums and landraces that form aerial roots will be used to validate these loci. In addition, backcross populations with inbred line RTx430 will be the basis for future transgenic validation experiments with an improved sorghum leaf-whorl transformation system (Silva et al. 2020).

Casa, A. M., et al. 2008. "Community Resources and Strategies for Association Mapping in Sorghum," *Crop Science* **48**, 30–40.

Silva, T., et al. 2020. "Use of *Sorghum bicolor* Leaf Whorl Explants to Expedite Regeneration and Increase Transformation Throughput," *Plant Cell, Tissue and Organ Culture* **141**, 242–55.

Upadhyaya, H. D., et al. 2009. "Developing a Mini Core Collection of Sorghum for Diversified Utilization of Germplasm," *Crop Science* **49**, 1769–80.

Van Deynze, A., et al. 2018. "Nitrogen Fixation in a Landrace of Maize Is Supported by a Mucilage-Associated Diazotrophic Microbiota," *PLoS Biology* **16**, e2006352.

**Funding Information:** The authors gratefully acknowledge funding from the U.S. Department of Energy Biological and Environmental Research (BER) Program grant no. DE-SC0021052.

## Optimizing Biological Nitrogen Fixation on Sorghum by Manipulating Microbial Communities

Claire Palmer<sup>1\*</sup> (claire.palmer@wisc.edu), Paulo Ivan Fernandes-Júnior<sup>1,2</sup>, April MacIntyre<sup>1,3</sup>, Sushmita Roy<sup>1</sup>, Wilfred Vermerris<sup>4</sup>, Ophelia Venturelli<sup>1</sup>, **Jean-Michel Ané<sup>1</sup>** 

<sup>1</sup>University of Wisconsin–Madison; <sup>2</sup>Brazilian Agricultural Research Corporation (Embrapa); <sup>3</sup>Valent BioSciences, Inc.; <sup>4</sup>University of Florida

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

**Goals:** As part of a multidisciplinary collaboration, this project aims to address key sustainability challenges facing the cultivation of sweet sorghum, an important biofuel crop. In particular, this project aims to decrease reliance on synthetic nitrogen fertilizers by improving biological nitrogen fixation, focusing on sorghum aerial roots as the main sites for diazotrophic activity. In tandem with collaborators investigating this trait from the plant perspective, researchers evaluate the bacterial communities associated with sorghum responsible for biological nitrogen fixation. The team plans to (1) isolate and characterize diverse bacterial strains from aerial root mucilage, (2) determine the bacterial interspecies interactions impacting biological nitrogen fixation, and

(3) develop and test synthetic communities with robust biological nitrogen fixation.

Cultivation of the key biofuel crop, sorghum, relies heavily on using natural gas-intensive and environmentally damaging nitrogen fertilizers (Rütting, Aronsson, and Delin 2018). As an alternative, biological nitrogen fixation (BNF) through the activity of plant-associated diazotrophic bacteria has the potential to improve crop production sustainability and reduce environmental damage (Pankievicz et al. 2019). Demonstrating the potential of BNF in cereals, some indigenous corn landraces from Central America can obtain 29%-82% of their nitrogen from BNF in the low oxygen and sugar-rich mucilage secreted by aerial roots (Deynze et al. 2018). Several sorghum accessions also produce aerial roots and mucilage, but the diazotrophic activity of sorghum aerial root-associated communities has not been investigated. Following these observations, the team hypothesized that aerial root mucilage provides the ideal environment in which to improve BNF in sorghum.

The project uses a synthetic microbial community approach to investigate the bacterial interspecies interactions impacting BNF. Researchers drew upon a previous study that identified a seven-member community that stably and reproducibly assembled on corn roots and added five additional diazotrophic strains of interest, developing a 12-member nitrogen-fixing community which is referred to as PComm1 (Niu et al. 2017). Community BNF and composition in low and high species richness subcommunities of PComm1 is evaluated using acetylene reduction assays followed by 16S sequencing. All possible 1-, 2-, 11-, and 12-member communities are investigated. Preliminary results suggest that most interspecies interactions negatively impact nitrogen fixation, with less nitrogenase activity observed in communities with more members. This data set will be used to build computational models to further elucidate interspecies interactions impacting community growth and fixation, as well as design communities with enhanced BNF.

In addition to investigating community BNF in a simplified model system, researchers are also working to isolate and characterize bacteria from aerial root mucilage. Previously, the team isolated a collection of ~90 individual strains with a range of plant growth–promoting traits, such as auxin production, siderophore production, phosphate solubilization, and 1-aminocyclopropane-1-carboxylate (ACC) degradation. However, less than 10% of these isolates have nitrogen-fixing capabilities. Therefore, focus is currently on expanding the diversity of nitrogen-fixing strains in the project's collection. A semisolid nitrogen-free medium approach was used to obtain new bacterial strains and increase the genetic variability of the mucilage-borne strain collection. Researchers obtained more than 320 new bacterial strains and used DNA fingerprinting to narrow down the collection to ~200 strains, excluding redundant profiles. In the next steps, the team will identify these strains and determine their diazotrophic ability, aiming to select new strains for future community manipulation and genetic engineering.

Deynze, A. V., et al. 2018. "Nitrogen Fixation in a Landrace of Maize Is Supported by a Mucilage-Associated Diazotrophic Microbiota," *PLOS Biology* **16**, e2006352.

Niu, B., et al. 2017. "Simplified and Representative Bacterial Community of Maize Roots," *Proceedings of the National Academy of Sciences* **114**, E2450–59.

Pankievicz, V. C. S., et al. 2019. "Are We There Yet? The Long Walk Towards the Development of Efficient Symbiotic Associations Between Nitrogen-Fixing Bacteria and Non-Leguminous Crops," *BMC Biology* **17**, 99.

Rütting, T., H. Aronsson, and S. Delin. 2018. "Efficient Use of Nitrogen in Agriculture," *Nutrient Cycling in Agroecosystems* **110**, 1–5.

**Funding Information:** The authors gratefully acknowledge funding from the U.S. Department of Energy Biological and Environmental Research (BER) Program grant no. DE-SC0021052.

## Single-Cell Omics to Examine the Cell Type-Specific Gene Regulatory Programs of Mucilage Production in *Sorghum bicolor*

Rafael E. Venado<sup>1</sup>, Saptarshi Pyne<sup>1</sup>, Matias Kirst<sup>2</sup>, Sushmita Roy<sup>1</sup>, **Jean-Michel Ané<sup>1\*</sup>** 

#### <sup>1</sup>University of Wisconsin–Madison; <sup>2</sup>University of Florida

Specific accessions of sorghum (Sorghum bicolor) develop aerial roots and produce carbohydrate-rich mucilage. The presence of mucilage suggests potential nitrogen-fixing microbial associations that contribute to nitrogen nutrition, as reported in a maize (Zea mays) landrace. However, the specific cell types and their underlying gene regulatory programs that produce mucilage are poorly understood. To understand the molecular basis of mucilage production, researchers developed a protocol to isolate nuclei of the cells collected from aerial roots that produced mucilage. The isolated nuclei were then sequenced using the 10X Genomics chromium technology. The preliminary single-nucleus dataset of circa 7,000 cells detected cell types in the aerial root based on the expressions of known marker genes. In parallel, the team has been developing novel analytical pipelines to identify cell types, cellular lineage structure, and celltype-specific gene regulatory networks (GRNs). Using the team's approach to published single-cell RNA-seq datasets, researchers were able to recapitulate putative GRN components and regulators for the cell fate specification. Applying these approaches to the sorghum single nucleus datasets should advance understanding of mucilage production in

the aerial root. This opens the possibility of identifying and validating candidate genes for breeding this important trait.

## EndoPopulus: Elucidating the Molecular Mechanisms of N-Fixation by *Populus* Endophyte, *Burkholderia vietnamiensis* WPB

Jayde Aufrecht<sup>2\*</sup>, Andrew W. Sher<sup>1</sup>, Daisy Herrera<sup>2</sup>, Amy E. Zimmerman<sup>2</sup>, Young-Mo Kim<sup>2</sup>, Nathalie Munoz<sup>2</sup>, Vimal Balasubramanian<sup>2</sup>, Dusan Velickovic<sup>2</sup>, Tanya Winkler<sup>2</sup>, Emma Gomez-Rivas<sup>1</sup>, Amir H. Ahkami<sup>2</sup>, **Sharon L. Doty<sup>1</sup>** (sldoty@uw.edu)

<sup>1</sup>University of Washington; <sup>2</sup>Pacific Northwest National Laboratory

**Goals:** The overall goal of this project is to investigate the roles and molecular mechanisms of endophytes in supporting productivity and fitness of *Populus*. Using systems biology approaches at both lab and field scales, the project will identify the molecular and physiological impacts of the bio-inoculants on the host plant in responding to nutrient and water limitation and determine if bio-inoculants not only increase plant nutrient stores but also "prime" plants for tolerance and resilience to abiotic stresses. Goals include identifying the molecular mechanisms of enhanced plant production and fitness by diazotrophic endophytes. The team will then integrate the plant physiology data with the molecular plant-microbe interactions data to develop a systems-level understanding of the genetic and molecular basis for diazotrophic endophytic mutualism in *Populus*.

Biological nitrogen fixation (BNF) by microbial diazotrophs can significantly contribute to nitrogen (N) availability and uptake in non-nodulating plant species, like Populus spp. There is currently a knowledge gap surrounding the molecular mechanisms underlying plant-diazotroph interactions and the spatial and temporal variations in microbial expression of genes involved in nitrogen fixation. In this work, the team seeks to identify which nitrogenous biomolecules diazotrophs are producing, how BNF is regulated in an axenic culture, and finally how Populus trichocarpa regulates N fixation during co-culture with diazotrophic endophytes. Through a <sup>15</sup>N<sub>2</sub> time course enrichment study, researchers identified key nitrogenous metabolites and proteins that are synthesized by diazotroph Burkholderia vietnamiensis (WPB). Using a fluorescent transcriptional reporter in the nitrogen fixing gene, *nifH*, researchers found that *nifH* is not uniformly expressed across genetically identical colonies of WPB. This result led to conducting a follow-on targeted metabolomics study in colony sections with and without *nifH* expression to identify which of the key nitrogeneous metabolites are produced in each scenario. Although WPB does not require the host plant to fix N, it

was hypothesized that the plant can regulate N fixation via metabolite exchange with the diazotroph under environmental changes including water-limiting conditions. Using liquid chromatography-mass spectrometry, researchers have identified 39 compounds in P. trichocarpa root exudates that are differentially abundant in a drought vs. well-watered condition. Currently, the team is testing the influence of these root exudates on the differential gene expression of WPB (including *nifH* gene) in axenic culture. Additionally, researchers have cultured WPB with the host plant directly in the RhizoChip, a synthetic soil habitat, which enabled direct imaging of the expression of microbial nifH within root epidermal cells. The team found that *nifH* expression is heterogeneous within root tissues, depends on the presence of soluble N compounds, and is localized to the root elongation zone where the WPB forms a unique physical interaction with the root cells. Finally, to understand the spatial distribution of metabolites exchanged by the plant, matrix-assisted laser desorption ionization-mass spectrometry imaging (MALDI-MSI) was used to image the distribution of various hormones and metabolites in root cross sections in plants subjected to drought vs. a well-watered condition. This experiment was repeated with plants that were inoculated with a community of endophytes to determine how the presence of endophytes alters the internal molecular environment of the root and how abiotic stressors like drought affect these interactions. These comprehensive experiments merge multiomics, chemical, and optical imaging data from axenic microbial, plant, and co-cultures to identify the key molecular mechanisms regulating beneficial plant-endophyte interactions.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0021137.

## EndoPopulus: Elucidation of the Roles of Diazotrophic Endophyte Communities in Promoting Productivity and Resilience of *Populus* Through Systems Biology Approaches

**Sharon L. Doty**<sup>1\*</sup> (sldoty@uw.edu), Andrew W. Sher<sup>1</sup>, Robert Tournay<sup>1</sup>, Darshi Banan<sup>1</sup>, Jayde Aufrecht<sup>2</sup>, Sun Woo Chung<sup>1</sup>, Jun Hyuk Jeon<sup>1</sup>, Matt Hendrickson<sup>1</sup>, Sriram Parasurama<sup>1</sup>, Morgan Raimondo<sup>1</sup>, Emma Gomez-Rivas<sup>1</sup>, Amir H. Ahkami<sup>2</sup>, Adam M. Deutschbauer<sup>3</sup>, Soo-Hyung Kim<sup>1</sup>

<sup>1</sup>University of Washington; <sup>2</sup>Pacific Northwest National Laboratory; <sup>3</sup>Lawrence Berkeley National Laboratory

**Goals:** The overall project goal is to move toward an understanding of the holobiont, how plants and the microbial community within them interact in ways that promote the productivity of the whole. Integration of plant physiology data with the molecular plant-microbe interactions (multiomics) data from greenhouse and field experiments will allow development of a systems-level understanding of the genetic and molecular basis for diazotrophic endophytic mutualism in *Populus*. This deeper level of understanding of the plant responses will guide construction of microbial communities in order to optimize the impacts of bioinoculants for environmental sustainability of bioenergy crops.

Poplar trees are important feedstocks for bioenergy and ecosystem services, but more efficient and resilient growth is essential for sustainability. The microbiome of wild poplar is a rich resource for plant growth promotion with some of the contributing micro-organisms able to provide nitrogen (N) and bioavailable phosphorus. In addition, these microorganisms may promote plant tolerance of other environmental stresses as well, such as drought. The first objective of this project is to unravel the molecular mechanisms of nitrogen fixation in an optimized constructed community of endophytes isolated from wild poplar. Initially studying a single aerobic diazotrophic strain with highly dynamic nitrogen-fixation, researchers characterized N-fixation using a nitrogenase gene promoter fusion to green fluorescent protein, fluorescence-activated cell sorting, poplar rhizo-chip assays, and a time series of <sup>15</sup>N-guided molecular analyses. Nitrogenase gene expression patterns both in vitro and in planta suggested environmental signals are at play, and the resulting signature of nitrogenous compounds was identified using <sup>15</sup>N-metabolomics.

A synergistic effect of specific non-diazotrophic strains with diazotrophs was revealed in vitro, moving towards an understanding of inter-species cooperation in a constructed community. Identification of the potential crosstalk molecules between strains will be a next step. A constructed community of eight endophyte strains was optimized with complementary symbiotic traits including N-fixation and synergistic interactions, phosphate solubilization, and hormone production. Genome-scale transposon mutagenesis of two of the diazotrophs of the constructed community is complete, allowing for a series of experiments to be conducted to evaluate the genetic requirements for N-fixation by these aerobic strains. Directed mutagenesis of specific genes is underway. To determine the molecular and physiological impacts of N-fixing endophytes on the host plant, the team is conducting both field and greenhouse level studies. In April 2022, a field site at the Roza Research Station in Prosser, Washington, was planted with poplar trees inoculated with the constructed community. Trees are being monitored and samples collected for analysis of the impacts of the endophytes versus controls at multiple levels including plant physiological, genomic, and metabolomics levels. Meanwhile, a series of more controlled greenhouse level

experiments have been performed, with more currently in progress.

Full genomic sequencing and de novo assemblies were completed for the strains making up the constructed community. The annotated draft-assemblies were submitted to the Type (Strain) Genome Server (TYGS) for whole genome-based taxonomic classification, with four strains (Azospirillum sp. SherDot2, Sphingobium sp. WW5, Herbiconiux sp. 11R-B1, and Rhizobium sp. PTD1) classified as potentially novel species and four strains identified to the species level (Rahnella aceris R10, R. aceris WP5, Azotobacter beijerinckii SherDot1, and Rhodotorula graminis WP1). None of the strains were predicted to be human pathogens by PathogenFinder (v1.1), and *in silico* analyses were completed cataloging the genetic features associated with plant growth-promoting (PGP) traits for each of the strains. Strain-specific-primer (SSP) sets have been designed and are currently being screened for the ability to successfully target unique DNA sequences in each of the strains. These will be used in digital droplet PCR assays to verify colonization and localization by plant compartment and to estimate the relative abundance of the strains within and between treatment groups. The SSPs will be used in conjunction with metabarcoding studies targeting the 16S rRNA genes to investigate changes in the community structure of the plant microbiome from the field and greenhouse experiments under abiotic stresses, providing information toward the objective of identifying the mechanisms of plant impacts on the microbial community.

By studying both the impacts of a constructed microbial community on the host plant as well as the impacts of the plant on the microbiome, the team hopes to move toward an understanding of the holobiont, how plants and the microbial community within them interact in ways that promote the productivity of the whole.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0021137.

## EndoPopulus: Endophyte Inoculation Enhances *Populus* Physiological Responses to Abiotic Stress

Darshi Banan<sup>1\*</sup> (dbanan@uw.edu), Jayde Aufrecht<sup>2</sup>, Sun Woo Chung<sup>1</sup>, Jun Hyuk Jeon<sup>1</sup>, Matt Hendrickson<sup>1</sup>, Sriram Parasurama<sup>1</sup>, Andrew W. Sher<sup>1</sup>, Robert Tournay<sup>1</sup>, Amir H. Ahkami<sup>2</sup>, Soo-Hyung Kim<sup>1</sup>, **Sharon L. Doty<sup>1</sup>** 

<sup>1</sup>University of Washington; <sup>2</sup>Pacific Northwest National Laboratory

**Goals:** The overall goal of the EndoPopulus Project is to understand how, at a molecular level, the micro-organisms

within the poplar tree microbiome can affect the host plant health and stress tolerance. The project will use a plant physiological approach to determine the impacts of endophyte inoculation on host plant performance under normal, nutrient-limited, and water-limited conditions. Data from field and greenhouse trials will be used to develop a process-based plant physiological model to generate testable hypotheses for further greenhouse examination of the microbial mechanisms responsible for altered plant productivity. Finally, physiological results will be integrated with microbiology, metabolomic, and transcriptomic data to generate an improved systems-level understanding of the plant-endophyte system scaling from the molecular to the canopy level.

Sustainable biofuel feedstock production is a key target for securing energy supply under climate change while minimizing environmental impacts. Inoculation with endophytes, mutualist microbes living inside plants, is a potential strategy to achieve this in forestry applications by improving host resource use efficiency, productivity, and stress tolerance. Endophytes isolated from trees in family Salicaceae have been shown to provide benefits such as fixing atmospheric nitrogen and synthesizing phytohormones in various in vitro and in planta experiments. Predicting how these benefits operate under production contexts requires a process-based understanding of plant, endophyte, and environmental interactions. This research investigates the physiological mechanisms by which inoculation with Salicaceae endophytes improves Populus performance under nutrient-limited or water-limited conditions.

Leaf epidermal morphology, photosynthesis, and wholeplant architecture and biomass were measured on native and hybrid poplars inoculated with Salicaceae endophytes in a series of greenhouse and field experiments. Experimental data were used to parameterize a coupled leaf gas exchange model to estimate the contribution of endophyte inoculation on biophysical and biochemical aspects of photosynthesis. Endophyte associated reductions in stomatal conductance varied with time of day and were most pronounced under higher light intensities. Likewise, improvements in photosynthetic water-use efficiency were greatest in inoculated plants under drought while inoculation reduced stomatal guard cell size regardless of water availability. Additionally, late season photosynthetic capacity  $(V_{cmax} and J_{max})$  was greater for inoculated plants under both greenhouse and field conditions. Under nitrogenlimited conditions, inoculated plants were taller and had root systems with greater total root length, branch number, and branching frequency compared to non-inoculated controls. These results suggest that the greatest benefits of endophyte inoculation on host productivity and resource use efficiency may be realized under stress conditions while other aspects of physiology, particularly stomatal

morphology and photosynthesis, responded across a range of environments. A combination of microbiology, transcriptomic, and metabolomic approaches will be used to connect changes in plant physiology to endophyte diversity, abundance, and activity. Finally, process-based modeling will be used to scale these leaf- and plant-scale changes to the canopy scale to predict tree performance in biofuel feedstock production applications.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0021137.

## Exploring Switchgrass Genetic Diversity with Multiple Reference Genomes

Jeremy Schmutz<sup>1,2\*</sup> (jschmutz@hudsonalpha.org), John T. Lovell<sup>1,2</sup>, Jerry Jenkins<sup>1</sup>, Shengqiang Shu<sup>2</sup>, David Goodstein<sup>2</sup>, Jane Grimwood<sup>1</sup>, **Thomas E. Juenger<sup>3</sup>** 

<sup>1</sup>HudsonAlpha Institute for Biotechnology; <sup>2</sup>Joint Genome Institute; <sup>3</sup>University of Texas—Austin

#### https://sites.cns.utexas.edu/juenger\_lab/switchgrass

Overall, researchers are striving to improve bioenergy feedstock production by understanding the genetic basis of plant-environment interactions. This goal includes testing for climate adaptation, modeling beneficial and stressful biotic interactions, and exploring the mechanisms of abiotic stress responses. During their work (Lowry et al. 2019; Lovell et al. 2021; Napier et al. 2022), researchers discovered a massive amount of physiological and molecular variation in switchgrass. While this diversity is the raw material that allows breeders to improve feedstock production, making use of this variation is very challenging-the immense DNA differences between some switchgrass genotypes means that traditional methods to explore genetic diversity simply do not work. Under the work presented here, researchers are developing multiple genome resources that span this diversity to provide the foundation for molecular characterization of switchgrass biomass production, stress responses and biotic interactions.

A single haploid reference genome gives breeders the resources to connect alleles to traits; a significant step towards accelerating crop improvement. However, breeding programs often leverage highly diverged germplasm, which contain large-scale variants that are not readily identified by a single reference genome. For example, in switchgrass, the fast-growing southern lowland AP13 genotype (which serves as the reference genome; Lovell et al. 2021) is approximately 1 million years diverged from the cold-tolerant northern upland gene pool. To assist breeding and gene discovery efforts, researchers have developed 16 total reference genome haplotypes from eight outbred heterozygous genotypes that span the genetic diversity of switchgrass, including a new annotated, haplotype resolved AP13. These fully *de novo* chromosome-scale genomes include three northern uplands, three southern lowlands, and two spanning the latitudinal range of the newly discovered coastal ecotype. Here, researchers present the progress on these genomes and an analysis of structural variation, including the presence of a putative minor chromosome that appears to segregate in coastal switchgrass populations. These variants can serve as *a priori* targets for ongoing molecular breeding efforts to make switchgrass a more economically and ecologically sustainable biofuel feedstock.

Lovell, J. T., et al. 2021. "Genomic Mechanisms of Climate Adaptation in Polyploid Bioenergy Switchgrass," *Nature* **590**, 438– 44. DOI:10.1038/s41586-020-03127-1.

Lowry, D. B., et al. 2019. "QTL x Environment Interactions Underlie Adaptive Divergence in Switchgrass Across a Large Latitudinal Gradient," *Proceedings of the National Academy of Sciences* **116**(26).

Napier, J., et al. 2022. "A Generalist-Specialist Tradeoff Between Switchgrass Cytotypes Impacts Climate Adaptation and the Geographic Range," *Proceedings of the National Academy of Sciences* **119**(15), e2118879119. DOI:10.1073/pnas.2118879119.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021126. The work conducted by the 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-AC02-05CH1123.

## The Influence of Switchgrass Internode Anatomy on Biofuel Production-Relevant Traits

David J. Thomas<sup>1,2</sup>, Jason Bonnette<sup>3</sup>, Rahele Panahabadi<sup>2</sup>, Millicent Sanciangco<sup>4</sup>, David B. Lowry<sup>4,5</sup>, Heather R. McCarthy<sup>1</sup>, **Thomas E. Juenger<sup>3</sup>**, Laura E. Bartley<sup>2\*</sup>

<sup>1</sup>University of Oklahoma; <sup>2</sup>Washington State University; <sup>3</sup>University of Texas–Austin; <sup>4</sup>Great Lakes Bioenergy Research Center; <sup>5</sup>Michigan State University

**Goals:** This project aims to understand the environmental and genetic influences on switchgrass composition towards increasing sustainability of switchgrass production for biorefining by developing generalist and specialist plant ideotypes that maximize biomass yield and composition, stress tolerance, and carbon sequestration capacity.

Switchgrass (*Panicum virgatum* L.) is a perennial warmseason Tallgrass and a promising feedstock for production of lignocellulosic biofuels. Local environmental conditions produce phenotypic variation across the geographical range of switchgrass. Researchers hypothesized that internode anatomy plasticity, representing evolutionarily driven local adaptation, modulates traits important for efficient biofuel production, including biomass yield (i.e., height), hydraulic conductivity, and biomass digestibility. Researchers analyzed internode anatomy of lowest above ground internodes in clones of upland (VS16, DAC) and lowland (AP13, WBC) switchgrass genotypes at three common garden sites in: South Texas, central Missouri, and central Michigan. Lowlands are larger in many traits including height, average xylem diameter, and internode annulus radius. A few traits such as cortical cell wall thickness (CCWT) and sclerenchyma radial thickness deviate from this pattern, lack isometry with height or internode diameter, and rank differently among genotypes across sites. Tillers with larger total leaf area, height, outer diameter, annulus radius, and average xylem diameter had larger maximum stem specific hydraulic conductance  $(K_S)$ . Additionally, sheath radial thickness positively correlates with  $K_S$  (SCC = 0.94) while CCWT and rind % annulus radius negatively correlate with  $K_S$  (SCC = -0.43 and -0.47, respectively). Comparing different digestion treatment parameters revealed that CCWT negatively correlates with biomass digestibility (R = -0.52) while no significant relationships were found with lignin or cellulose content. Thus, the significant anatomical plasticity in biofuel-production relevant traits present among genotypes across environments provides support towards optimizing internode anatomy that favors cell wall deconstruction efficiency for lignocellulosic biorefining.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant nos. DE-SC0014156 and DE-SC0021126. Work at the Great Lakes Bioenergy Research Center is supported under award DE-SC0018409.

## Transcriptome and Gene Regulatory Network Analyses in *Camelina* Nitrogen Response and Seed Development

Fernando Henrique Correr\* (fernando.correr@montana. edu), Racheal Upton, **Chaofu Lu**, Jennifer Lachowiec

Montana State University

https://www.montana.edu/econproject/

**Goals:** *Camelina* is a *Brassica* oilseed crop that has great potential to become a sustainable source of bioenergy in the U. S. However, the low nitrogen use efficiency and the low seed and oil yield compared to other major oilseed crops hinder this potential. The goal of this project is to decipher the genetic and physiological mechanisms that determine

the nitrogen use efficiency and oilseed yield during the most critical processes of the *Camelina* life cycle: (1) how *Camelina*, in partnership with soil microbes, maximizes its ability to absorb and assimilate nitrogen into vegetative biomass; and (2) upon the transition to reproductive growth, how nitrogen is efficiently remobilized from senescing tissues (leaves and silicles) into sinks (seeds) to optimize yield potential by increasing seed size and enhancing oil synthesis.

*Camelina* (*Camelina sativa* (L.) Crantz) is an oilseed with potential as a crop for second generation biofuel production. To achieve sustainable production of Camelina oil, the energy-consuming inputs need to be optimized, especially reducing the amount of nitrogen in the production system. Researchers need to obtain a systems-level understanding of genetic and physiological mechanisms that may be used to enhance the nitrogen use efficiency (NUE) and to improve agronomic and seed traits in Camelina. The team grew three C. sativa accessions-Suneson, Cam 70 and Cam 116-under two nitrogen conditions of 0.55 mM and 5.5 mM of nitrate in the nutrient solution, and assessed the transcriptomes of six organs-flower, 10 days after fertilization (DAF) seed, pod, leaf, stem, and root. The largest number of differentially expressed genes (DEGs) is due to the genotype effect for almost all tissues; but differences attributed to nitrogen input are remarkable in flower, pod, and stem. The team identified functional terms enriched with DEGs due to nitrogen, which involved mostly in abiotic responses and changes in photosynthesis-associated processes. In seeds, researchers also identified significant differences in genes of the phenylpropanoid pathway, biosynthesis of flavonoids, and maintenance of seed dormancy.

There is a lack of knowledge in processes that regulate the development of Camelina seeds and affect their viability. The previous study showed that the miR167A overexpression results in lower levels of a-linolenic acid, but also larger seeds and delayed seed maturation in Camelina (Na et al, 2019). Researchers therefore aimed to identify genes responsible for this altered seed development through a co-expression weighted network analysis. Researchers examined published gene expression profiles of the Camelina wildtype (cv. Suneson) and miR167OE, a transgenic line overexpressing miR167A, at 8, 10 and 12 days after flowering (DAF). One group of co-expressed genes increased in expression from 8 to 10-12 DAF in miR167OE; however, their expression levels at 10-12 DAF was similar to those in Suneson at 8 DAF. In this group, the team found significant enrichment of genes in auxin response, seed oil body biogenesis, seed maturation and seed germination. Genes with the opposite expression pattern included those with functions in the flavonoid biosynthesis pathway including BANYLUS and multiple members of the TRANSPAR-ENT TESTA family. They are potentially involved in the determination of seed coat color but also can influence

other seed traits. In addition to genes found in the enrichment analysis, the team also identified other candidates involved in seed germination: DELAY OF GERMINATION 1 (DOG1), ABA-HYPERSENSITIVE GERMINATION 1 (AHG1) and FIE2. The expression profiles of those genes showed profound differences at 8 DAF in miR167OE compared to all the other samples. Based on the patterns of expression in these genes, researchers hypothesized that earlier germination occurs in miR167OE compared to Suneson despite delayed seed maturation. This was confirmed by a germination assay to evaluate the germination rates of both genotypes for two weeks, and the transgenic genotype indeed germinated faster. These outcomes provide evidence that besides changing the oil profile, miR167A overexpression also interferes with seed development and maturation, resulting in faster germination rates in miR167OE.

Na, G., et al. 2019. "Enhancing Micro RNA 167A Expression in Seed Decreases the *a*-Linolenic Acid Content and Increases Seed Size in *Camelina sativa*," *The Plant Journal* **98**(2), 346–58.

**Funding Information:** This research is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science program Award No. DE-SC0021369.

## The Root Microbiome of *Camelina* in the Dryland Wheat Production Areas of Eastern Washington

Hao Peng<sup>1\*</sup> (hao\_peng@wsu.edu), Elle M. Barnes<sup>2</sup>, Chuntao Yin<sup>3</sup>, Daniel Schlatter<sup>4</sup>, Cody Willmore<sup>5</sup>, **Timothy Paulitz<sup>1,5</sup>**, Susannah G. Tringe<sup>2,6</sup>, Chaofu Lu<sup>7</sup>

<sup>1</sup>Washington State University; <sup>2</sup>DOE Joint Genome Institute; <sup>3</sup>U.S. Department of Agriculture Agricultural Research Service (USDA-ARS), Brookings, SD; <sup>4</sup>USDA-ARS, St. Paul, MN; <sup>5</sup>USDA-ARS, Pullman, WA; <sup>6</sup>Lawrence Berkeley National Laboratory; <sup>7</sup>Montana State University

**Goals**: Identify microbial diversity of the soil, rhizosphere, and root of the oilseed crop, *Camelina sativa*; Determine the effects of cropping zone and precipitation on microbial composition; Characterize the core bacterial and fungal microbiome of *Camelina*.

*Camelina sativa* (L.) is a broadleaf member of the Brassicaceae family. A short season crop (85 to 100 days), *Camelina* is fairly disease- and pest resistant, drought- and frost tolerant, and can be grown under low-input conditions (Gao et al. 2018; Zanetti et al. 2017; Matteo et al. 2020; Neupane et al. 2022; Séguin-Swartz et al. 2009). An important part of the success of this plant under low-input conditions may be attributed to its rhizosphere and endosphere microbiome, but to date there have been no publications on the microbiome of *Camelina*. In this study, soil was collected from 33 sites across a 200 km gradient in the dryland wheat production area in eastern Washington. This included four main cropping zones based on precipitation: wheat/summer fallow (< 300 mm/yr), intermediate (300 to 450 mm/yr), annual cropping zone (450 to 600 mm/yr) and south (400 to 500 mm/yr). *Camelina* was planted into the soils in the greenhouse. After six weeks, plants were harvested and DNA was extracted from root, rhizosphere and bulk soil. Bacterial 16S rRNA and fungal internal transcribed spacer (ITS) amplicons were amplified and sequenced with Illumina MiSeq (Gohl et al. 2016). Researchers then identified the core Camelina microbiome using an abundance-occupancy model (Shade and Handelsman 2012; Shade and Stopnisek 2019). They found that microbiome diversity decreased from the soil to the endosphere and, for the soil microbiome, increased with average precipitation. Plant compartment, zone, previous crop, and site all significantly affected composition of the Camelina microbiome, with the largest differences seen between the annual cropping zone and wheat/summer fallow zone. Several Actinobacteriota and Alphaproteobacteria (e.g., Actinoplanes, Aeromicrobium, Mycobacterium, Rhizobium, Caulobacter, and Sphingomonas) and two fungi (e.g., *Pseudogymnoascus* and one unidentified ASV) were differentially abundant in the plant-associated core microbiome. Fitting the abundance-occupancy model to the Sloan neutral model of community assembly provided additional evidence that these genera appear to be deterministically selected by the plant, suggesting that these core genera may possess traits that would make them good candidates for efficient plant colonization.

Gao, L., et al. 2018. "Photosynthesis and Growth of *Camelina* and Canola in Response to Water Deficit and Applied Nitrogen," *Crop Science* **58**(1), 393–401. DOI:10.2135/cropsci2017.07.0406.

Gohl, D.M., et al. 2016. "An Optimized Protocol for High-Throughput Amplicon-Based Microbiome Profiling," Nature Protocol Exchange. DOI:10.1038/protex.2016.030.

Matteo, R., et al. 2020. "Camelina (*Camelina sativa* L. Crantz) Under Low-Input Management Systems in Northern Italy: Yields, Chemical Characterization and Environmental Sustainability," *Italian Journal of Agronomy* **15**(2), 132–43. DOI:10.4081/ija.2020.1519.

Neupane, D., et al. 2022. "Realizing the Potential of *Camelina sativa* as a Bioenergy Crop for a Changing Global Climate," *Plants* **6**, 772. DOI:10.3390/plants11060772

Séguin-Swartz, G., et al. 2009. "Diseases of *Camelina sativa* (False Flax)," *Canadian Journal of Plant Pathology* **4**, 375–86. DOI:10.1080/07060660909507612.

Shade, A. and Handelsman, J. 2012. "Beyond the Venn Diagram: The Hunt for a Core Microbiome," *Environmental Microbiology* 1, 4–12. DOI:10.1111/j.1462-2920.2011.02585.x.

Shade, A. and Stopnisek, N. 2019. "Abundance-Occupancy Distributions to Prioritize Plant Core Microbiome Membership," *Current Opinion in Microbiology* **49**, 50–58. DOI:10.1016/j.mib.2019.09.008.

Zanetti F., et al. 2017. "Agronomic Performance and Seed Quality Attributes of Camelina (*Camelina sativa* L. Crantz) in Multi-Environment Trials Across Europe and Canada," *Industrial Crops and Products* **107**, 602–08. DOI:10.1016/j.indcrop.2017.06.022.

**Funding Information:** This research was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Genomic Science Program grant no. DE-SC0021369, and Contract No. DE-AC02-05CH11231 to the U.S. Department of Energy Joint Genome Institute (http://jgi.doe.gov/), a DOE Office of Science User Facility.

## **SCIENCE FOCUS AREA**

# Solvent Disruption of Biomass and Biomembranes

## Biofuels Disruption of Membrane Domains: An Unrecognized Mode of Solvent Stress

Jonathan D. Nickels<sup>1</sup>, Luoxi Tan<sup>1</sup>, Micholas Dean Smith<sup>2,3</sup>, Haden L. Scott<sup>3</sup>, Sai Venkatesh Pingali<sup>3</sup>, Hugh M. O'Neill<sup>3</sup>, Jennifer Morrell-Falvey<sup>3</sup>, John Katsaras<sup>3</sup>, Jeremy C. Smith<sup>2,3</sup>, James G. Elkins<sup>3\*</sup> (elkinsjg@ornl.gov), **Brian H. Davison<sup>3</sup>** 

<sup>1</sup>University of Cincinnati; <sup>2</sup>University of Tennessee–Knoxville; <sup>3</sup>Oak Ridge National Laboratory (ORNL)

#### https://sfa-biofuels.ornl.gov/

Goals: The Solvent Disruption of Biomass and Biomembranes Science Focus Area (SFA) provides fundamental knowledge about how solvents alter the structures of plant cell walls and microbial membranes. The project's overarching hypothesis is that knowledge of partitioning or binding of the solvent from the bulk phase to biomass or biomembranes can help predict maximal or minimal disruption. Solvents disrupt biological structures comprising amphiphilic molecules and polymers (e.g., membranes and biomass). Determining common biophysical principles of solvent disruption will lead to new understandings of how solvents affect the relevant structures. This information will help determine the ultimate microbial limits in tolerating specific solvents, as well as the eventual design of cosolvents best suited for pretreatment. The SFA will integrate the power of world-class neutron scattering capabilities and leadership-class supercomputing facilities available at ORNL. These capabilities are complemented by expertise in biodeuteration and biomembranes at ORNL, plant cell wall chemistry at the University of Tennessee, and interpreting small-angle neutron scattering (SANS) data at the University of Cincinnati.

A sustainable bioeconomy will undoubtably rely on the efficient production of lignocellulosic biofuels that can be combusted directly in automobile engines or catalytically upgraded to long-chain hydrocarbons for use as diesel and aviation fuels. Typical target biofuels include ethanol and n-/isobutanol which act as amphiphilic co-solvents in the aqueous environment of fermentation. Amphiphilic alcohols are well known to have chaotropic effects on biological molecules which lends to their inherent toxicity to the microbial biocatalysts used for fermentation. While these

toxic effects can be broad, targeting all biological macromolecules, the cellular membrane is recognized as especially vulnerable to disruption due to the partitioning of amphiphilic co-solvents into the lipid bilayer. Functionally, this leads to membrane thinning, destabilization, loss of membrane potential, and eventually, cell death.

While the impact on the transverse structure of the membrane from high co-solvent titers is known, the effects on lateral biomembrane structure have not been well-studied. Lateral membrane structure is described as differences in lipid composition and membrane physical properties across the plane of the membrane. This can be understood in analogy to an in-plane phase separation which is dictated by the presence of high and low melting point lipid species as well as sterols (or their microbial analogs). Depending on the ratio of these molecules, phase separation occurs in the membrane, creating regions of different local composition and physical properties. In the biological context, these structures are colloquially known as membrane microdomains or lipid rafts. Rafts play an important role in many cellular functions due to their role as platforms to segregate and organize membrane proteins; this co-localization is critical to oligomerization of membrane proteins, and by extension, optimal cellular function.

In this project, researchers pursue the hypothesis that amphiphilic co-solvents, such as ethanol and n-/isobutanol, alter or disrupt functional membrane microdomains, leading to an unrecognized mode of co-solvent toxicity and cellular stress at non-lethal co-solvent concentrations. The team examines this hypothesis in a range of model and in vivo lipid membrane systems including phaseseparating membrane mimics in the form of unilamellar vesicles, microbial membrane extracts, and engineered microbial systems that allow tunable membrane compositions. Domain structure and behavior is examined with a variety of non-destructive techniques including microscopy and SANS with and without the solvent ethanol (See Fig. 10, p. 61). Computational approaches including molecular dynamics simulations (MD) are also leveraged to provide molecular detail and experimentally inaccessible understanding of membrane behavior in the presence of co-solvents.

Tan et al. (2023) has made a significant step forward in validating the SFA's hypothesis by (1) demonstrating a direct disruption of a model lipid raft due to the presence of an amphiphilic co-solvent (ethanol) and (2) elucidating the physical mechanism by which the disruption of the lipid domains occurred. The team shows that unequal partitioning of ethanol between the co-existing phases leads to an increase hydrophobic mismatch of the thickness of these phases and a corresponding increase in the domain line tension. This is a driver to minimize the domain interface



**Fig. 10.** A membrane composed of two lipids (blue and red) segregates into thinner and thicker domains (left). When ethanol (green) is added it differentially partitions, thins the membrane and disrupts the lateral structure (right). Research shows precisely this type of disruption in model membranes and elucidates the mechanism by which it is occurring. This represents an unrecognized mode of solvent-induced stress in biofuel production. These findings establish new targets for intervention to improve fermentation yields. [Reprinted with permission from Tan et al., 2023. "Amphiphilic Co-Solvents Modulate Structure of Membrane Domains," *ACS Sustainable Chemistry & Engineering* **11**(4), 1598–1609. DOI:10.1021/acssuschemeng.2c06876]

to domain area ratio. This represents the physical basis for a novel mode of co-solvent induced cell stress due to domain disruption. Continuing work will further test the hypothesis using more complex sample types, including live cells, and compare the effects of other amphiphilic biofuel molecules on domain organization. Further validation of the SFA's hypothesis will provide a more holistic understanding of solvent-membrane interactions and inform actionable approaches to mitigating toxicity and improving biofuel yields from fermentative microbes.

Tan, L., et al. 2023. "Amphiphilic Co-Solvents Modulate Structure of Membrane Domains," *ACS Sustainable Chemistry and Engineering* **11**(4), 1598–609. DOI:10.1021/acssuschemeng.2c06876.

**Funding Information:** This research is supported by the U.S. Department of Energy, Office of Science, through the Genomic Science program, Biological and Environmental Research (BER) Program, under FWP ERKP752. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725.

## Visualization of Solvent Disruption of Biomass and Biomembrane Structures in the Production of Advanced Biofuels and Bioproducts

**Brian H. Davison**<sup>1\*</sup> (davisonbh@ornl.gov), Barbara Evans<sup>1</sup>, Arthur J. Ragauskas<sup>1,2</sup>, Micholas Dean Smith<sup>1,2</sup>, Haden L. Scott<sup>1</sup>, Sai Venkatesh Pingali<sup>1</sup>, Hugh M. O'Neill<sup>1</sup>, Loukas Petridis<sup>1</sup>, Luna Liang<sup>2</sup>, Riddhi Shah<sup>2</sup>, Xianzhi Meng<sup>2</sup>, Yunxuan Wang<sup>2</sup>, Yunqiao Pu<sup>1</sup>, Jeremy C. Smith<sup>1,2</sup>

<sup>1</sup>Oak Ridge National Laboratory (ORNL); <sup>2</sup>University of Tennessee–Knoxville

#### https://sfa-biofuels.ornl.gov/

Goals: The Solvent Disruption of Biomass and Biomembranes Science Focus Area (SFA) provides fundamental knowledge about how solvents alter the structures of plant cell walls and microbial membranes. The project's overarching hypothesis is that knowledge of partitioning or binding of the solvent from the bulk phase to biomass or biomembranes can help predict maximal or minimal disruption. Solvents disrupt biological structures comprising amphiphilic molecules and polymers (e.g., membranes and biomass). Determining common biophysical principles of solvent disruption will lead to new understandings of how solvents affect the relevant structures. This information will help determine the ultimate microbial limits in tolerating specific solvents, as well as the eventual design of cosolvents best suited for pretreatment. The SFA will integrate the power of world-class neutron scattering capabilities and leadership-class supercomputing facilities available at ORNL. These capabilities are complemented by expertise in biodeuteration and biomembranes at ORNL, plant cell wall chemistry at the University of Tennessee, and interpreting small-angle neutron scattering (SANS) data at the University of Cincinnati.

A sustainable bioeconomy will undoubtably rely on the efficient production of lignocellulosic biofuels that can be combusted directly in automobile engines or catalytically upgraded to long-chain hydrocarbons for use as diesel and aviation fuels. Plant cell wall structure of biomass is an intricate design of several carbohydrate polymers encased in the hydrophobic lignin polymer to protect against degradation. The recalcitrance to deconstruction of lignocellulosic biomass due to the complex physicochemical structure of plant cell walls is a challenge in biological-based biorefinery systems due to the complex physicochemical structure of plant cell walls. Pretreatment and genetic modification are two approaches in biomass conversion that have succeeded in modifying the structure of lignocellulose to enable better enzymatic deconstruction. However, the structural differences among pretreatment-solubilized biomass biopolymers have not been extensively investigated. The SFA's goal is to understand the molecular-level mechanism which drive efficient biomass deconstruction. ORNL scientists have reported direct experimental and computational evidence of physical chemical principles underlying pretreatment. Here, the team will discuss the use of molecular dynamics (MD) simulations, experimental pretreatments with acids and with acidified solvents combine the scattering measurement to elucidate structural changes in the three key biomass polymers (cellulose, hemicellulose, and lignin). This will be illustrated by several examples.

Researchers determined that solvent mixtures with both hydrophilic and hydrophobic interactions are key for efficient deconstruction of biomass as revealed by neutron scattering and molecular simulation. The team elucidated the effect of tetrahydrofuran (THF)-water pretreatment on the nanoscale architecture of biomass and the role the co-solvents play in solubilizing lignin and cellulose (Pingali et al. 2020). In situ SANS determined temperaturedependent changes in biomass morphology; whereas lignin dissociates over a wide temperature range (>25°C), cellulose disruption occurs only above 150°C. SANS with contrast variation and MD simulations provided direct evidence for the formation of THF-rich nanoclusters (~0.5 nm) on the nonpolar cellulose surfaces and on hydrophobic lignin, and equivalent water-rich nanoclusters on polar cellulose surfaces.

In another example, three organosolv pretreatment systems—ethanol (EtOH), tetrahydrofuran, and  $\gamma$ -valerolactone, in dilute acidic aqueous—were used on wild-type and two transgenic switchgrasses with altered lignin. All organosolv pretreatments caused a significant reduction in the molecular weights of lignins particularly, and up to ~90% decrease was observed in EtOH pretreated lignin compared to untreated lignin. A correspondence was found between the molecular weight reduction of lignin molecules in the experiments and the number of hydrogen bonds between lignin and the organic solvents as calculated in the MD simulation, suggesting a connection between the depolymerization of lignin and its ability to hydrogen bond with the organic solvents.

To understand the role of noncellulosic switchgrass polymers on the overall efficiency of pretreatment, the structural evolution of the noncellulosic polymers of the plant cell wall was investigated during dilute acid pretreatment by employing *in situ* SANS on various polymer fractions from switchgrass (Yang et al. 2021). In this study, researchers observed real-time structural changes not possible to observe by any



**Fig. 11.** Local demixing of water (blue) and tetrahydrofuran (THF; yellow) on the surfaces of cellulose (green). Water and THF are mixed in the bulk away from the cellulose. [Courtesy Oak Ridge National Laboratory]

other technique. These interpretations were consistent with MD simulations. These results suggest that not only lignin but also hemicellulose can form aggregate particles within plant cell walls during pretreatment. These concepts can be employed to tune pretreatment technologies that maximize deconstruction of biomass and facilitate the separation of its components for upgrading to energy and materials.

Liang, L., et al. 2022. "Chemical and Morphological Structure of Transgenic Switchgrass Organosolv Lignin Extracted by Ethanol, Tetrahydrofuran, and  $\gamma$ -Valerolactone Pretreatments," *ACS Sustainable Chemistry and Engineering* **10**(28), 9041–52. DOI:10.1021/acssuschemeng.2c00948.

Pingali, S. V., et al. 2020. "Deconstruction of Biomass Enabled by Local Demixing of Cosolvents at Cellulose and Lignin Surfaces," *Proceedings of the National Academy of Sciences* **117**(29), 16776–81. DOI:10.1073/pnas.1922883117.

Yang, Z., et al. 2021. "Structural Reorganization of Noncellulosic Polymers Observed *In Situ* During Dilute Acid Pretreatment by Small-Angle Neutron Scattering," *ACS Sustainable Chemistry and Engineering* **10**(1), 314–22. DOI:10.1021/acssuschemeng.1c06276.

**Funding Information:** This research is supported by the U.S. Department of Energy (DOE), Office of Science, through the Genomic Science program, Biological and Environmental Research (BER) Program, under FWP ERKP752. Oak Ridge National Laboratory (ORNL) is managed by UT-Battelle, LLC for the U.S. DOE under Contract no. DE-AC05-00OR22725. It used neutron scattering resources at the High Flux Isotope Reactor, a DOE Office of Science, Scientific User Facility operated by the ORNL. SAXS measurements were performed at the LiX beamline of the National Synchrotron Light Source II, a U.S. DOE Office of Science User Facility operated by Brookhaven National Laboratory under Contract No. DE-SC0012704.
## **SCIENCE FOCUS AREA**

A Systems Biology Approach to Interactions and Resource Allocation in in Bioenergy-Relevant Microbial Communities

## Identifying Genomic and Metabolic Underpinnings of Algal-Bacterial Interactions via Metatranscriptomics, NanoSIMS Isotope Tracing, and Genome-Scale Metabolic Modeling

Megan Morris<sup>1</sup>, Vanessa Brisson<sup>1</sup>, Hyu Kim<sup>2</sup>, Jeffrey Kimbrel<sup>1</sup>, Austin Wessel<sup>1</sup>, Ali Navid<sup>1</sup>, Cullen Buie<sup>2</sup>, Peter Weber<sup>1</sup>, **Rhona Stuart<sup>1</sup>**, Xavier Mayali<sup>1\*</sup>

<sup>1</sup>Lawrence Livermore National Laboratory, <sup>2</sup>Massachusetts Institute of Technology

#### https://bio-sfa.llnl.gov

Goals: Algal and plant systems have the unrivaled advantage of converting solar energy and CO<sub>2</sub> into useful organic molecules. Their growth and efficiency are largely shaped by the microbial communities in and around them. The µBiospheres science focus area (SFA) seeks to understand phototroph-heterotroph interactions that shape productivity, robustness, the balance of resource fluxes, and the functionality of the surrounding microbiome. The team hypothesizes that different microbial associates not only have differential effects on host productivity but can change an entire system's resource economy. The approach encompasses single-cell analyses, quantitative isotope tracing of elemental exchanges, omics measurements, and multiscale modeling to characterize microscale impacts on system-scale processes. Team members aim to uncover crosscutting principles that regulate these interactions and their resource allocation consequences to develop a general predictive framework for system-level impacts of microbial partnerships.

The team's previous work has shown that heterotrophic bacteria influence algal productivity through complex metabolic interactions and resource utilization. However, those interactions, varying from mutualistic to antagonistic, may be context-dependent based on resource availability. Identifying these dependencies requires a better fundamental characterization of the interactions' genomic and metabolic underpinnings. To accomplish this, researchers present a study system comprised of a previously uncharacterized and uncultivated antagonistic parasitic bacterium that attaches to and crashes bioenergy-relevant algae in days. Team members hypothesized that the acute phenotypic changes would coincide with bacterial uptake of algal-derived substrates, and that bacterial metabolism during infection would indicate key resources required by the bacterium. To address these hypotheses, the team used a combination of amplicon sequencing and genome-resolved metagenomics, fluorescent *in situ* hybridization (FISH), stable isotope probing, metatranscriptomics, and metabolic modeling.

To set up the system, team members first identified a bacterial enrichment community that killed the alga Phaeodactylum tricornutum. Using 16S amplicon sequencing in tandem with genome-resolved metagenomics, researchers initially annotated the presumed parasitic bacterium as a previously uncharacterized *Rickettsiales* sp. Then, using custom FISH probes, positively identified the Rickettsiales sp. directly attached to algal cells. Researchers next quantified the ability of the parasitic Rickettsiales sp. to incorporate algal-derived substrates using stable isotope probing with <sup>13</sup>C bicarbonate and <sup>15</sup>N nitrate and measured single-cell uptake with nanoscale secondary ion mass spectrometry (NanoSIMS). Researchers found that algal cells with attached parasitic bacteria had lower single-cell carbon fixation, and when attached to the algal host, bacterial cells were enriched in both <sup>13</sup>C and <sup>15</sup>N, with some bacteria more highly enriched in <sup>15</sup>N compared to the host. This suggests that Rickettsiales sp. incorporated algal-derived substrates and may have the capacity to siphon newly metabolized N resources from its host.

In order to provide insight into the identity of these host resources, researchers used metatranscriptomics to quantify gene-level expression changes in P. tricornutum and Rickettsiales during infection. They found that prior to Rickettsiales attachment, it overexpressed genes for iron and trace metal scavenging, amino acid starvation, ribosomal hibernation, oxidative stress, and flagellar and pilin assembly, among others. Once attached, Rickettsiales upregulated genes for chemotaxis and signal transduction, antibiotic resistance, production of proteases and peptidoglycanases, type IV secretion system, gene transfer agents implicated in virulence, membrane transporters, and amino acid metabolism. Taken together, this suggests that free-living parasites are likely starved for nutrients, particularly amino acids and trace metals, and once attached can produce enzymes that degrade algal cell wall material, transfer virulence factors to the host, and potentially import and metabolize algal-derived N-rich amino acids and proteins. To better predict specific substrate utilization by bacteria, the team has curated a genome-scale model

from the near-complete metagenome-assembled genome (MAG) of the Rickettsiales sp. bacterium and integrated the gene-expression data into a metabolic flux balance analysis (GX-FBA). Optimization of the model is ongoing; however, researchers anticipate that it will putatively identify specific algal-derived substrates metabolized by the parasitic bacterium during the interaction, which can then be tested in subsequent experiments. To this end, the team has ongoing research to identify taxon-level resource partitioning of P. tricornutum exudates by its microbiome, using metabolomics, stable isotope probing, and the newly designed porous microplate incubation system. Thus far, researchers have conducted exometabolomics studies to characterize metabolite uptake profiles and potential resource partitioning among a suite of algal associated bacterial isolates, and further investigated those interactions with pairwise sequential growth experiments. Following this sequential interaction study, the team has confirmed the bacterial competition by measuring each growth response to algal exudates in situ using a co-culture porous microplate. Moving forward, the team will expand upon these simplified approaches to quantify how intimate algal-bacterial interactions change the flow of carbon system-wide using more complex microbial communities.

**Funding Information:** This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL Biofuels SFA, FWP SCW1039 LLNL-ABS-845199.

## The Context-Dependency of Plant-Microbial Interactions in the Bioenergy Resource Economy

Rachel Hestrin<sup>1,2\*</sup> (rhestrin@umass.edu), John Casey<sup>1</sup>, Rachel Hammer<sup>3</sup>, Noah Luecke<sup>3</sup>, Vanessa Brisson<sup>1</sup>, Megan Kan<sup>1</sup>, Rebecca Ju<sup>1</sup>, Jeffrey Kimbrel<sup>1</sup>, Christina Ramon<sup>1</sup>, Prasun Ray<sup>4</sup>, Katerina Estera-Molina<sup>1,6</sup>, Jessica Wollard<sup>1</sup>, Marissa Lafler<sup>1</sup>, Steven J. Blazewicz<sup>1</sup>, Kelly Craven<sup>4,5</sup>, Mary K. Firestone<sup>6</sup>, Peter Weber<sup>1</sup>, Ali Navid<sup>1</sup>, **Rhona Stuart<sup>1</sup>**, Christine Hawkes<sup>3</sup>, **Jennifer Pett-Ridge<sup>1</sup>**, **Erin Nuccio**<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory; <sup>2</sup>University of Massachusetts– Amherst; <sup>3</sup>North Carolina State University (NCSU); <sup>4</sup>Noble Research Institute; <sup>5</sup>Oklahoma State University; <sup>6</sup>University of California–Berkeley

#### https://bio-sfa.llnl.gov

**Goals**: Algal and plant systems have the unrivaled advantage of converting solar energy and  $CO_2$  into useful organic molecules. Their growth and efficiency are largely shaped by the microbial communities in and around them. The µBiospheres SFA seeks to understand phototroph-heterotroph interactions that shape productivity, robustness, the balance of resource fluxes, and the functionality of the surrounding microbiome. The team hypothesizes that different microbial associates not only have differential effects on host productivity but can change an entire system's resource economy. The approach encompasses single cell analyses, quantitative isotope tracing of elemental exchanges, omics measurements and multi-scale modeling to characterize microscale impacts on system-scale processes. The team aims to uncover crosscutting principles that regulate these interactions and their resource allocation consequences to develop a general predictive framework for system-level impacts of microbial partnerships.

The hyphosphere is a hotspot for multipartite interactions that shape critical terrestrial processes such as soil nutrient cycling, carbon (C) distribution, and plant growth. To investigate how water limitation impacts microbial dynamics and biogeochemistry in both the rhizosphere and hyphosphere, researchers inoculated P. hallii with one of two functionally different mycorrhizal partners (Rhizophagus irregularis and Serendipita bescii) and grew the plants under either water-limiting or water-replete conditions in <sup>13</sup>CO<sub>2</sub> labeling chambers to enable carbon tracking. After 3 months, researchers used H218O quantitative stable isotope probing (qSIP) to assess how water limitation impacted hyphosphere bacterial growth rates and diversity. The team found that both fungal partners helped sustain growth and diversity in hyphosphere bacterial communities exposed to water limitation relative to uninoculated controls. Of the bacterial taxa that responded positively to R. irregularis or S. bescii in water-limited soil, many belong to lineages that are considered droughtsusceptible, including Bacteroidetes, Planctomycetes, Verrucomicrobia, Proteobacteria, and Acidobacteria. The size of soil C pools and <sup>13</sup>CO<sub>2</sub> efflux from hyphosphere soil depended on soil moisture conditions, but exometabolite profiles and multimodal imaging suggest that the different mycorrhizal fungi also can influence C flow and soil biogeochemistry. Together, the findings indicate that mycorrhizal fungi can support biotic activity and resilience to water limitation.

In addition to mycorrhizal fungi, roots are often colonized by a diverse array of endophytic fungi. Historically, these fungi have been assumed to be largely commensal, but recent work by the NCSU team suggests that many confer nutritional benefits to the host plant. Using a panel of phylogenetically diverse root endophytes isolated from switchgrass, researchers demonstrated that these fungi broadly enable plant acquisition of organic N and P in soil. Compared to fungus-free controls, 30% of fungi (n=12) increased tissue N by 20-90% when provided with organic N, and 40% of fungi (n=16) increased tissue P by 25-80%. Most importantly, some fungi appear to substantially shift the N:P ratio (from N>>P to N=P). Fungi that aggressively consumed organic nutrients in culture were less beneficial for host acquisition of organic N or P (r = -0.39 to -0.57). Leveraging these results, team members are investigating C-nutrient trading in the root endophyte system. This work will substantially increase understanding of root endophyte contributions to host and ecosystem C and nutrient cycling.

In parallel with experimental work, team members are developing a plant-mycorrhizal-bacteria model to bridge cellular scale processes within a systems-level context. The model is a hybrid model that combines a lattice-free hyphal network and a co-localized diffusive/advective grid. The model is designed to enable interpretation of and integration with spatially resolved community flux balance simulations of mycorrhizal-bacterial communities using data from experimental studies. Environmental control on plant, mycelium, and bacterial dynamics are governed by a set of coupled differential equations that preserve C and nutrient mass balances. The team will explore the traits and tradeoffs that promote plant growth and total system biomass growth under different environmental conditions and will test these predictions in future experimental studies.

**Funding Information:** This work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and supported by the Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL Biofuels SFA, FWP SCW1039, LLNL-ABS-845280.

# **SCIENCE FOCUS AREA**

## QPSI: Quantitative Plant Science Initiative

## Integrating Functional Genomics with Molecular-Level Experimentation to Understand Adaptation to Nutrient Stress in Poplar and Sorghum

**Timothy Paape**<sup>1</sup>(tpaape@bnl.gov), Meng Xie<sup>1</sup>, Qun Liu<sup>1</sup>, Crysten E. Blaby-Haas<sup>1</sup>, Doreen Ware<sup>4,5</sup>, Daifeng Wang<sup>3</sup>, Samuel Seaver<sup>2</sup>, Lifang Zhang<sup>4</sup>, Dimiru Tadesse<sup>1</sup>, Michael Regulski<sup>4</sup>, Miriam Pasquini<sup>1</sup>, Changxu Pang<sup>1</sup>, Sunita Kumari<sup>4</sup>, Desigan Kumaran<sup>1</sup>, Chiarg Gupta<sup>3</sup>, Nicolas Grosjean<sup>1</sup>, Nicolas Gladman<sup>4</sup>, Aditi Bhat<sup>1</sup>, Jeremiah Anderson<sup>1</sup>

<sup>1</sup>Brookhaven National Laboratory; <sup>2</sup>Argonne National Laboratory; <sup>3</sup>University of Wisconsin—Madison; <sup>4</sup>Cold Spring Harbor Laboratory; <sup>5</sup>U.S. Department of Agriculture, Agricultural Research Service, Ithaca, NY

**Goals:** The Quantitative Plant Science Initiative (QPSI) is a capability that aims to bridge the knowledge gap between genes and their functions. A central aspect of QPSI strategy is combining genome-wide experimentation and comparative genomics with molecular-level experimentation (see Fig. 12, this page). In this way, researchers leverage the scalability of omics data and bioinformatic approaches to capture system-level information while generating sequence-specific understanding of gene and protein function. By incorporating molecular-level experimentation into the workflow, team members are addressing the question of how a protein functions and establishing mechanistic insight into how sequence variation impacts phenotype. This knowledge serves as a touchstone for accurate genomebased computational propagation across sequenced genomes and forms the foundation for robust predictive modeling of plant productivity in diverse environments.

Micro- and macronutrient stress is a growing importance in maximizing bioenergy/bioproduction crop yield in marginal soil. Bioavailability in the soil is dynamic and variable, and yield-impacting deficiencies are poorly understood. Because micronutrients are essential for the proper assimilation and metabolism of macronutrients such as nitrogen, metal deficiencies and other soil stresses can result in poor macronutrient availability. To support the development of bioenergy crops with improved nutrient stress resilience, the goal during the current 3-year period is to develop a genome-based, molecular-level and system-level understanding for the adaptation to micronutrient stress. Focusing on the bioenergy crops poplar and sorghum, researchers have completed a large-scale transcriptomics time-course experiment to understand how these plants respond to different nutrient stresses in their environment. Team members are also employing an interdisciplinary approach to provide a layer of experimentally grounded sequence-specific understanding of molecular-level functions for major players involved in plant homeostasis. Comparative genomics provides an *in silico* platform to generate protein function hypotheses. Hypotheses are tested with reverse genetics in model organisms and biochemical assays of protein family members. Structure-function studies supply mechanistic insight into how sequence space translates into molecular function. While working with micronutrient stresses in the current phase, there will be subsequent opportunities to incorporate other real-world conditions with the addition of field experiments addressing the impact of the soil geochemistry, microbiome and rhizosphere, and studying macro- and micro-nutrient interactions.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program.



Fig. 12. The Quantitative Plant Science Initiative (QPSI) science focus area multidisciplinary approach. [Courtesy Brookhaven National Laboratory]

## Functional Characterization of bHLH Transcription Factors Coordinating Abiotic Stress Response, Secondary Cell Wall Biosynthesis, and Metal Homeostasis in *Populus*

Dimiru Tadesse<sup>1</sup>, Estella Yee<sup>1</sup>, Desigan Kumaran<sup>1</sup>, Aditi Bhat<sup>1</sup>, Crysten E. Blaby-Haas<sup>2</sup>, Timothy Paape<sup>1</sup>, **Meng Xie**<sup>1\*</sup> (mxie@bnl.gov)

<sup>1</sup>Brookhaven National Laboratory; and <sup>2</sup>Lawrence Berkeley National Laboratory

Goals: The Quantitative Plant Science Initiative (QPSI) is a capability that aims to bridge the knowledge gap between genes and their functions. A central aspect of QPSI strategy is combining genome-wide experimentation and comparative genomics with molecular-level experimentation. In this way, the team leverages the scalability of omics data and bioinformatic approaches to capture system-level information while generating sequence-specific understanding of gene and protein function. By incorporating molecularlevel experimentation into the workflow, researchers are addressing the question of how a protein functions and establishing mechanistic insight into how sequence variation impacts phenotype. This knowledge serves as a touchstone for accurate genome-based computational propagation across sequenced genomes and forms the foundation for robust predictive modeling of plant productivity in diverse environments.

*Populus* is one of DOE's flagship bioenergy crops as the source of renewable energy and biobased products. Gene regulatory networks (GRNs) that describe the hierarchical regulatory relationships between transcription factors (TFs), associated proteins, and their target genes are fundamental for coordinating genome-wide gene expression

responses to environmental and developmental signals. The complex and dynamic behavior of plant TFs is crucial for the GRN plasticity for sensitive responses. However, it also poses a fundamental challenge in understanding molecular principles underlying GRN dynamics. Researchers previously identified two basic Helix-Loop-Helix (bHLH) TFs (PtrbHLH038 and PtrbHLH011) whose expressions were oppositely regulated in *Populus* leaves under iron deficiency treatment. Using the transactivation assay in protoplasts, researchers found that they are transcriptional repressors. Using their novel protoplast-based transient chromatin immunoprecipitation-sequencing (transient ChIP-seq) approach for mapping genome-wide binding targets of TFs in vivo, team members found and validated that PtrbHLH038 directly regulates PtrbHLH011 and abiotic stress-responsive genes. In contrast, PtrbHLH011 seems to have broader regulatory functions because its targets include metal transporters, growth regulators, and master regulators of secondary cell wall biosynthesis. More interestingly, protoplast-based approaches enabled the discovery that iron deficiency treatment can eliminate PtrbHLH011's binding and repression on its target genes. By performing TurboID-based proximity labeling in protoplasts, the team identified protein cofactors that form complexes with PtrbHLH011. Based on the results described above, researchers hypothesize that PtrbHLH038 and PtrbHLH011 form a regulatory hierarchy to coordinate abiotic stress responses, secondary cell wall biosynthesis, and iron homeostasis in Populus. Transgenic plants overexpressing PtrbHLH038 and PtrbHLH011 have been generated to test the team's hypothesis and study the biological impacts of these two transcription factors.

**Funding information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program as part of the Quantitative Plant Science Initiative (QPSI) Science Focus Area.



## **UNIVERSITY PROJECTS**

## Accelerating Discovery of Genes Regulating Stomatal Patterning and Water Use Efficiency in C<sub>4</sub> Crops with Novel High-Throughput Methods for Mutagenesis and Phenotyping

Grace Tan<sup>1\*</sup> (gracedt2@illinois.edu), Redeat Tibebu<sup>2</sup>, Erik Alan Myers<sup>2</sup>, Degao Liu<sup>2</sup>, Hui Jiang<sup>3</sup>, Collin Luebbert<sup>3</sup>, Daniel Tejeda-Lunn<sup>1</sup>, Colby G. Starker<sup>2</sup>, Andrew D. B. Leakey<sup>1</sup>, Daniel F. Voytas<sup>2</sup>, **Ivan Baxter**<sup>3</sup>

<sup>1</sup>University of Illinois–Urbana-Champaign; <sup>2</sup>University of Minnesota; and <sup>3</sup>Donald Danforth Plant Science Center

Goals: Bioenergy feedstocks need to be deployed on marginal soils with minimal inputs to be economically viable and have a low environmental impact. Currently, crop water supply is a key limitation to production. The yields of C<sub>4</sub> bioenergy crops such as Sorghum bicolor have increased through breeding and improved agronomy. Still, the amount of biomass produced for a given amount of water use (wateruse efficiency, or WUE) remains unchanged. Therefore, this project aims to develop novel technologies and methodologies to redesign the bioenergy feedstock sorghum for optimal WUE. Within this broader context, this subproject is using Setaria viridis as a rapid cycling model for gene discovery. The project aims to develop and demonstrate novel methods and resources to accelerate both the production of genetic variants and phenotyping of WUE traits as part of reverse and forward genetics approaches to discover genes regulating stomatal patterning and WUE.

Stomata regulate the exchange of CO<sub>2</sub> and water vapor between the leaf and atmosphere, and therefore play a key role in determining WUE. However, relatively little is known about the genes that regulate stomatal patterning and WUE in C<sub>4</sub> grasses. Previous work has identified several hundred candidate genes through a combination of genome-wide association study and transcriptome-wide association study. To validate these discoveries and advance efforts to engineer improved WUE of bioenergy crops, the team is developing novel methods to accelerate the use of forward and reverse genetics for gene discovery. Researchers conducted a forward genetic screen of 155 families of an N-nitroso-N-methylurea-mutagenized Setaria population, of which 100 lines show small stature and/or altered leaf color phenotypes. Whole-plant WUE is being assessed by imaging and automated lysimeters. These families are part of a larger population which is being fully sequenced by DOE

Joint Genome Institute to create a sequence indexed mutant population. This data is being paired with screening for abnormalities in stomatal patterning. To accelerate a reverse genetic screen, researchers developed new methods for viral delivery of mutagenesis reagents in both *Setaria* and tobacco and are using these methods for studying stomatal developmental genes through loss- and gain-of-function mutations. Both the forward and reverse genetic approaches utilize high-throughput optical tomography imaging to generate high-resolution images of the leaf surface.

A robust machine learning model was developed for identifying the size, shape, and number of epidermal cells in maize, and the team is adapting the model for *Setaria*. This work demonstrates a positive feedback loop of high-throughput phenotyping to genotyping in which genes of interest can be quickly identified and tested for a role in WUE. Success in this effort could be leveraged to accelerate research on a wide range of other traits and species.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0023160 and DE-SC0018277.

## Combining GWAS of Metabolomic and Transcriptomic Datasets to Accelerate Discovery of Genes Regulating the Effect of Drought on Plant Growth and Metabolism in *Sorghum* and *Setaria*

Allen Hubbard\*, Philip Ozersky, Louis Connelly, Xiaoping Li, Hui Jiang, Jennifer Barrett, Collin Lubbert, Maddison Pope, Shrikaar Kambhampati, **Ivan Baxter** 

**Donald Danforth Plant Science Center** 

Goals: Bioenergy feedstocks need to be deployed on marginal soils with minimal inputs to be economically viable and have a low environmental impact. Currently, crop water supply is a key limitation to production. The yields of C<sub>4</sub> bioenergy crops such as *Sorghum bicolor* have increased through breeding and improved agronomy. Still, the amount of biomass produced for a given amount of water use (wateruse efficiency, or WUE) remains unchanged. Therefore, this project aims to develop novel technologies and methodologies to redesign the bioenergy feedstock Sorghum for optimal WUE. Within this broader context, this subproject is leveraging the sorghum pangenome and large phenotypic datasets in Setaria viridis and S. bicolor to discover metabolically important genes for the regulation of WUE in the C<sub>4</sub> grasses. This project aims to develop and demonstrate novel methods and resources to accelerate the production of genetic variants and accelerate phenotyping in both reverse genetics and forward genetics approaches leading to discovery of genes regulating metabolic regulators of WUE.

Plants make an amazing array of metabolites to grow and respond to environmental change. The large number of compounds created by plants are poorly characterized, and the genetic programs controlling them are largely unknown. In order to better understand the metabolomic response of  $C_4$  plants to drought stress, researchers conducted parallel experiments in *Sorghum* and *Setaria* using diversity panels. Plants were grown in a controlled environment phenotyping system at two watering levels, and samples were harvested 6 days after the watering levels were set.

Metabolites for each sample were quantified in an untargeted fashion via liquid chromatography-mass spectrometry (LC-MS) using two different columns in both positive and negative mode to identify a large number of compound classes. A third of the samples were also profiled for RNA transcripts. ~3800 metabolomics samples, each run on two columns in two modes created an immense informatics challenge. To improve the sensitivity and accuracy of metabolite detection in similar large datasets, the team has developed a suite of three computational tools to overcome the challenges of unreliable algorithms and inefficient validation protocols: isolock, autoCredential, and anovAlign (IAA). Isolock uses metabolite-isotopologue pairs (isopairs) to calculate and correct for mass drift noise across LC-MS runs. AutoCredential leverages statistical features of LC-MS data to amplify naturally present <sup>13</sup>C isotopologues and validate metabolites through isopairs. AnovAlign, an anova-derived algorithm, is used to align retention time windows across samples to improve delineation of retention time windows for mass features. Using the IAA suite, researchers have quantified thousands of mass features across the 3,800 metabolomics samples. Genome-wide association study (GWAS) analysis has identified a large number of loci affecting these metabolites, including several loci in syntenic regions of the Setaria and Sorghum genomes for the same metabolite. Using informatics tools to harness the sorghum pangenome, researchers are combining the loci with transcriptomic and genomic data to identify candidate genes and alleles underlying the metabolomic response to water deficit, as well as leveraging tandem mass spectrometry to better characterize promising mass features.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0023160 and DE-SC0018277.

#### Understanding and Engineering Crown Root Development to Improve Water-Use Efficiency in Bioenergy Grasses

Willian Goudinho Viana<sup>1\*</sup> (viana@stanford.edu), Janina Tamborski<sup>1\*</sup> (jtambor@stanford.edu), Jose Sebastian<sup>2</sup>, Johannes Daniel Scharwies<sup>1</sup>, Taylor L. Clarke<sup>1</sup>, Keith E. Duncan<sup>3</sup>, Christopher N. Topp<sup>3</sup>, Hui Jiang<sup>3</sup>, Colby G. Starker<sup>4</sup>, **Jennifer A. N. Brophy<sup>1</sup>**, **José R. Dinneny<sup>1</sup>**, **Ivan Baxter<sup>3</sup>** 

<sup>1</sup>Stanford University; <sup>2</sup>Indian Institute of Science Education and Research; <sup>3</sup>Donald Danforth Plant Science Center; <sup>4</sup>University of Minnesota

Goals: This study aims to improve water-use efficiency in the bioenergy grass Sorghum bicolor by engineering crown root development. Crown roots play an important role in water and nutrient acquisition. To precisely modify crown root development, this project is engineering the location and gene expression level of a newly identified crown root regulator called *Crown Root Defective* using synthetic genetic circuits. Since synthetic genetic circuits have never been utilized in crop species before, researchers are currently testing and optimizing individual circuit building blocks in Setaria and Sorghum. The project goal is to enable the construction of circuits that drive gene expression in a predictable manner in these bioenergy grasses. Through the precise control of genes influencing crown root development and the identification of new root development regulators, the team aims to improve water-use efficiency.

S. bicolor is a biofuel crop that offers great potential because of its tolerance to drought, heat, and low cost of production as compared to other potential feedstocks. Despite yield gains through breeding, its productivity in suboptimal conditions is still limited. Acquisition of water and nutrients is mediated by the root, and in grasses the mature root system is primarily composed of crown roots (Viana, Scharwies, and Dinneny 2022). In the panicoid grass model Setaria viridis, soil moisture around the crown stimulates the development of crown roots, while drought conditions inhibit their growth, which facilitates the conservation of water (Sebastian et al. 2016). Despite their importance in crop productivity, the genetic mechanisms behind crown root development in response to environmental factors are not well understood. To engineer water-use efficiency in panicoid bioenergy grasses, researchers aimed to elucidate the key genetic mechanisms and subsequently alter crown root development. The team identified a mutant named crown root defective-1 (crd-1) that is specifically impaired in crown root development under well-watered conditions. Interestingly, this defect is rescued under drought stress conditions. Through Bulk Segregant Analysis, the gene was mapped to a single nucleotide polymorphism disrupting the splice site resulting in a premature stop codon of a gene encoding a WD-repeat protein. Researchers independently generated secondary alleles and complementation lines to confirm that the identified gene is in fact the causal gene for the phenotype.

Because of its promising phenotype, the team decided to leverage crd-1 to alter the number of crown roots by precisely modifying its expression level. In the past, synthetic genetic circuits were successfully used in combination with tissue-specific promoters to modify root architecture in Arabidopsis thaliana (Brophy et al. 2022). Despite success in developing tools and modifying model plants, in the past the transfer of these applications to crop plants has not always been successful. The team therefore established transient systems in Sorghum and S. viridis to test and optimize individual components of synthetic circuits for application in grass species. These circuits are built using synthetic transcription factors composed of bacterial DNA-binding proteins fused to transcriptional activation or repression domains. Through the use of transient protoplast expression, results have shown that the strength of the synthetic transcription factors is influenced by both the DNA-binding protein and activation domain. Researchers hypothesize this is due to the different conformation of each synthetic transcription factor resulting in altered accessibility of the activation domains. Moreover, research shows that the transcriptional activity of synthetic transcription factors is directly correlated to the number of binding sites utilized in the synthetic promoters. Together, the results from transient assays show that these parts can be modified in multiple aspects to drive gene expression in a predictable manner. Subsequently, genetic circuits were successfully built that implement the Boolean NOT implies logic operation in grasses.

In future research, the team aims to identify more factors influencing crown root development. WD-repeat proteins can act as scaffold proteins that organize multiprotein complexes and can also regulate gene expression. To understand the function of the WDR6 protein, researchers performed a yeast–two hybrid screen, which led to the discovery of several binding partners, including members of the Growth-Regulating Factor (GRF) family of transcription factors. The team is now working on validating these interactions and planning to conduct functional studies to further explore the role of the GRFs in crown root development. Through the identification of more players in the crown root development pathway and initial data from the genetic circuits, results indicate a promising future for engineering root architecture and other complex traits in bioenergy grasses.

Brophy, J. A. N., et al. 2022. "Synthetic Genetic Circuits as a Means of Reprogramming Plant Roots," *Science* **377**(6607), 747–51.

Sebastian, J., et al. 2016. "Grasses Suppress Shoot-Borne Roots to Conserve Water During Drought," *Proceedings of the National Academy of Sciences* **113**(31), 8861–66.

Viana, W. G., J. D. Scharwies, and J. R. Dinneny. 2022. "Deconstructing the Root System of Grasses Through an Exploration of Development, Anatomy and Function," *Plant, Cell and Environment* **45**(3), 602–19.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0018277 and DE-SC0023160.

## Developing Chassis for LDPE Upcycling from Microbes Native to the Gut Microbiome of Yellow Mealworms

Ross Klauer<sup>1,2\*</sup> (rklauer@udel.edu), Lummy Maria Oliveira Monteiro<sup>1,2</sup>, Jyoti Singh<sup>1,2</sup>, Bill Alexander<sup>3</sup>, Adam Guss<sup>3</sup>, Carrie A. Eckert<sup>3</sup>, Kevin Solomon<sup>1,2</sup>, **Mark Blenner<sup>1,2</sup>** 

<sup>1</sup>Delaware Biotechnology Institute; <sup>2</sup>University of Delaware; <sup>3</sup>Oak Ridge National Laboratory

https://www.blennerlab.com/

http://solomonlab.weebly.com

**Goals:** This project aims to enable the efficient depolymerization of polyolefins and upcycling to itaconic acid, through novel genomic insights into nutrient enhanced polyolefin degradation by the yellow mealworm gut microbiome and genetic tool development for gut microbiome isolates and engineered microbial communities.

Waste plastics represent a significant untapped source of carbon. Annually more than 200 million tonnes of plastic waste in the form of high-density polyethylene (HDPE), low-density polyethylene (LDPE), and polypropylene (PP) are generated and pollute soils, waterways, and bodies. No robust system exists to capture this carbon; however, in prior work in labs, researchers observed that yellow mealworm gut microbes are able to chemically modify these wastes suggesting an opportunity for biological upcycling. Microbial profiling of these communities reveals nonmodel taxa as the dominant microbes. Furthermore, the plastic degrading metabolic pathway remains unelucidated. A barrier to understanding plastic metabolism and engineering biological upcycling platforms is the lack of genetic tools to manipulate these species.

As a first step, the team previously enriched *Tenebrio molitor* guts for fast-degrading plastic microbial communities to identify optimal chasses for development for upcycling. Enrichment including growth on plastics and plastics with nutritional supplements such as oats, bran, and banana peels. Mealworm gut communities supplemented with oats were found to be optimal for plastics degradation rate. Researchers are further studying the effect of micronutrients and macronutrients in mealworm diets to enrich for an optimal polyolefin degrading community. The nutritional composition of the plastic and oats diet will be altered by supplementing with macronutrients such as protein and fats and with micronutrients such as nitrogen and potassium. From enriched communities, the team plans to cultivate key players in the degradation process and understand how the optimized community is structured, informing future creation of reduced complexity engineered microbial communities for plastic degradation and upcycling.

More than 300 microbial isolates were cultivated from microbial communities of LDPE, HDPE, and PP-fed mealworm guts, with and without oats supplementation, through standard microbiological methods and isolating morphologically distinct colonies from each gut condition. By screening these isolates for growth with polyethylene powder as the primary carbon source, 30 taxonomically unique isolates were found to use LDPE powder as a carbon source for growth in a liquid mineral medium. Improved growth with LDPE particles as the primary carbon source relative to a mineral medium suggested that microbes participate in the degradation process. LDPE degradation is demonstrated by scanning electron microscopy (SEM) through visualized microbial colonization of plastic particles, biofilm formation, and surface modifications to the particles. Contrarily, no surface modifications were observed via SEM on LDPE films treated by the same microbial isolates, indicating that degradation efficiency varies depending on polymeric mechanical and chemical properties such as form factors, additives, and processing conditions. From this characterization survey, and microbial profile abundance, five taxa were identified belonging to genera Staphylococcus, Enterococcus, Corynebacterium, Brevibacterium, and Kocuria as robust chassis for development as upcycling platforms.

To understand degradation and upcycling potential of these isolates, researchers bioinformatically screened these isolates for polyolefin degrading enzymes. Polyolefin backbone cleavage is anticipated to be initiated by secreted enzymes via oxygen radical chemistry due to their high reactivity on carbon-carbon bonds. Protein families (pfam) that perform oxygen radical chemistry such as monooxygenases, dioxygenases, and peroxidases were identified in isolate genomes in order to identify PE-active enzymes. Genomes of isolates that grow best on LDPE were mined to find pfam that have a high number of genes that perform said chemistry relative to closely related microbial taxa. Select genes were heterologously expressed into Escherichia coli, purified, and tested on plastic substrates. Plastic substrate modification as a result of enzymatic activity is observed via (%) crystallinity increase after enzyme treatment on LDPE films.

As a first step towards the development of robust genetic engineering tools for these isolates, the team is collecting genome and methylome sequence data. The genome sequences are a prerequisite for targeted genetic engineering. The methylome data will help researchers understand which restriction modification systems are present in each isolate and design genetic parts to work in the presence of these systems.

In summary, mealworm microbiomes were enriched for optimum plastics degradation by supplementing with various co-feeds. New co-feed studies will identify key nutrients that further improve plastic degradation. From previously enriched communities, over 300 organisms have been isolated, 30 of which are able to grow with plastic as their primary carbon source. Five organisms hailing from nonmodel genera appear able to grow on LDPE powder and are somewhat abundant in enriched communities, indicating these microbes are suitable chassis for engineered polyolefin degradation and upcycling. Enzymes initiating the first degradation step of the upcycling process are being mined bioinformatically from microbial isolates and communities. Degradation capability of these enzymes is being analyzed chemically by Fourier-transform infrared spectroscopy and mechanically by differential scanning calorimetry. In parallel, downstream metabolic degradation processes are being evaluated using metagenomics and metatranscriptomics. Upon elucidation of polyolefin deconstruction processes, said processes will be enhanced in isolated strains and reassembled back into reduced complexity communities, as well as engineered into a heterologous host for the eventual production of itaconic acid from plastic substrates via metabolically engineering of the host.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program, Genomic Science program under Award Number DE-SC0023085.

# Agent-Based Algal Modeling for the Rational Engineering of *Chlamydomonas reinhardtii*

Sandra Gomez Romero, Alex J. Metcalf, Nanette R. Boyle\*

**Colorado School of Mines** 

**Goals:** The overall research objective is to develop an experimentally validated multiparadigm multiscale modeling framework that will enable the most advanced and predictive metabolic modeling of diurnally grown photosynthetic organisms to date. The genome-scale metabolic model of *Chromochloris zofingiensis* will be embedded into an agentbased modeling framework to allow modeling of diurnal growth; the model will also be able to simulate intracellular fluxes, cell-to-cell interactions, cell-to-environment interactions, metabolite diffusion, and spatial distribution. This modeling approach will allow simulation of metabolic shifts that occur due to diel cycles and generation of rational engineering strategies to design production strains that are not impacted negatively by this natural phenomenon.

Economical algae production requires growth under outdoor light, but the diel nature of sunlight complicates modeling efforts. Researchers have developed a solution for that: a fully functional 3D agent-based model, capable of simulating algal growth under diurnal conditions. By combining systems biology data from Chlamydomonas reinhardtii grown in diurnal light (Strenkert et al. 2019) with agent-based modeling and detailed tracking of nutrient and light conditions, this model performs better than traditional steady-state metabolic models (Metcalf et al. 2022). In order to develop a model of growth during diurnal light, the team needed to decouple the standard biomass formation equation to allow different components of biomass to be synthesized at different times of the day. The model was able to more accurately predict qualitative phenotypical outcomes of the starchless mutant, sta6. The model then predicted growth of single-gene knockouts, and potential targets were identified for rational engineering efforts to increase productivity. The team will discuss recent advances in characterizing these mutants and further improvement of the model by including light and nutrient tacking. This model enables evaluation of the impact of genetic and environmental changes on the growth, biomass composition, and intracellular fluxes for diurnal growth.

Metcalf, A. J., et al. 2022. "Rhythm of the Night (and Day): Predictive Metabolic Modeling of Diurnal Growth in *Chlamydomonas*," *mSystems* 7(4), e00176-00122.

Strenkert, D., et al. 2019. "Multiomics Resolution of Molecular Events During a Day in the Life of *Chlamydomonas*," *Proceedings of the National Academy of Sciences* **116**, 2374–83.

#### Early Career

## Synthetic Membrane Biology of Microbial Cell Factories: Lipid Interactions that Shape the Inner Mitochondrial Membrane

Kailash Venkatraman<sup>1</sup>, Christopher Lee<sup>1</sup>, Guadalupe C. Garcia<sup>2</sup>, Arijit Mahapatra<sup>1</sup>, Mark Ellisman<sup>1</sup>, Padmini Rangamani<sup>1</sup>, **Itay Budin<sup>1\*</sup>** (ibudin@ucsd.edu)

<sup>1</sup>University of California–San Diego; <sup>2</sup>Salk Institute for Biological Studies

**Goals:** The goal of this Early Career Program project is to engineer the structure and properties of cell membranes to improve the performance of industrially relevant microbes. The project's first objective is to enhance the rate and

efficiency of the respiratory metabolism by engineering the organization of the electron transport chain. Engineering efforts will define the limits of respiratory metabolism and seek to increase the production of energy-intensive next-generation biofuels. The second objective is to apply the emerging biochemistry of intracellular lipid trafficking pathways to develop new transporters for the capture of valuable biochemicals produced by the engineered yeast.

The inner mitochondrial membrane (IMM) is the site of bulk ATP generation for yeast cell factories and is thus essential for production of high-energy biofuels and bioproducts. The IMM is defined by highly curved cristae membranes (CM), whose lipidome is composed of unsaturated phospholipids and cardiolipin (CL). Recent efforts combined experimental lipidome dissection with multiscale modeling to investigate how lipid interactions shape CM morphology and metabolic function. When modulating fatty acid unsaturation in engineered yeast strains, the team observed that loss of di-unsaturated phospholipids (PLs) led to a surprising breakpoint in IMM topology and respiratory capacity. PL unsaturation modulates the organization of ATP synthases that shape cristae ridges, phenocopying the loss of CM shaping proteins. Based on molecular modeling of mitochondrial-specific membrane adaptations, the team hypothesized that conical lipids like CL buffer against the effects of saturation on the IMM. Loss of CL was found to collapse the IMM at intermediate levels of PL saturation, a function that is independent of ATP synthase oligomerization. To explain this interaction, researchers employed a continuum modeling approach, finding that lipid and protein-mediated curvatures act in concert to form curved membranes in the IMM. These results suggested that fermentation conditions that alter the fatty acid pool, such as oxygen availability or overproduction of saturated fatty acids in engineered strains, define the CL function. While loss of CL only has a minimal phenotype in highly aerated shake flasks, research shows that its synthesis is essential in microaerobic fermenters, which promote saturated lipidomes. Lipid and protein mediated mechanisms of curvature generation thus act together to support mitochondrial architecture in industrially relevant environments.

**Funding Information:** This research is supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0022954.

#### BioPoplar: A Tunable Chassis for Diversified Bioproduct Production

**C. Robin Buell**<sup>1</sup>\*, Christopher Dardick<sup>2</sup>, Wayne Parrott<sup>1</sup>, Bob Schmitz<sup>1</sup>, Patrick Shih<sup>3</sup>, Chung-Jui Tsai<sup>1</sup>, Breeanna R. Urbanowicz<sup>1</sup>

<sup>1</sup>University of Georgia; <sup>2</sup>Agricultural Research Service, U.S. Department of Agriculture; <sup>3</sup>University of California–Berkeley

Domestication and breeding efforts have shown that selection of specific plant architecture traits across a wide array of plant species, both annuals and perennials, results in improved traits for human use, either for food, feed, or fuel. Similarly, selective breeding can yield distinct chemotypes of crops with desired chemical profiles or compositions. Today, precise knowledge of gene regulation and function can be generated through high-resolution omics technologies, and a synthetic biology toolkit can be constructed to engineer plant genomes at the DNA sequence, chromatin accessibility, and expression levels. Thus, society has entered an era where it is possible to model, design, and then engineer precise changes in plant genomes that will lead to predictive, modified traits.

In this project, researchers will re-engineer poplar as a multipurpose crop that can be used for bioenergy, biomaterial, and bioproduct production. A cell atlas will be generated that encompasses gene expression, gene regulatory networks, and cis-regulatory elements and is responsible for gene expression at the cell-type level, providing the requisite knowledge base and tools for precision bio-based design and fabrication of multipurpose poplar. The team will couple single-cell datasets with new genome and epigenome editing tools to develop new morphotypes of poplar that have altered tree and leaf architecture. These morphotypes will substantially improve biomass potential via increased stand density and tree integrity, photosynthetic capture, and trichome density, and serve as the foundational chassis. These chassis will have altered ratios of leaves to stems and/ or trichome density in which researchers can further engineer cell wall composition and/or novel molecules such as precursors for drop-in fuels, thus making chemotypes of poplar that are 'customized' to their biomaterial or bioproduct applications and simultaneously 'maximized' in optimal morphotypes. The project will employ an iterative design process in which metabolic pathways are optimized to create unique chemotypes with tailored biomaterial and bioproduct composition.

This project will yield poplar chassis with multipurpose uses including bioenergy, biomaterials, and bioproduct production. The generation of a robust cell-type–specific set of transcription factors and *cis*-regulatory elements, the ability to modulate gene expression in a high-resolution manner, i.e., that of specific cell types, will enable precision genome engineering of metabolism, a significant advancement in capabilities in modulating plant biochemistry. The change in architecture will be exploited to permit production of bioproducts (drop-in fuel precursors in leaves), biomaterials (modified wood composition) in wood, as well as changes in agronomic production practices such as increased stand density leading to increased yield. Collectively, these engineered chassis and tools provide the platform of a new era for poplar biology, agronomy, and processing.

# B5: Bigger Better *Brassicaceae* Biofuels and Bioproducts—An Overview

Doug K. Allen<sup>1,2</sup>, Malia Gehan<sup>1</sup>, Philip D. Bates<sup>3</sup>, Timothy P. Durrett<sup>4</sup>, Ruth Welti<sup>4</sup>, Jerome M. Fox<sup>5</sup>, Chaofu Lu<sup>6</sup>, Michael J. Smanski<sup>7</sup>, Trupti Joshi<sup>8</sup>, Jay J. Thelen<sup>8</sup>, Dong Xu<sup>8</sup>, **Edgar B. Cahoon<sup>9\*</sup>** (ecahoon2@unl.edu)

<sup>1</sup>Donald Danforth Plant Science Center; <sup>2</sup>Agricultural Research Center, U.S. Department of Agriculture; <sup>3</sup>Washington State University; <sup>4</sup>Kansas State University; <sup>5</sup>University of Colorado–Boulder; <sup>6</sup>Montana State University; <sup>7</sup>University of Minnesota; <sup>8</sup>University of Missouri; <sup>9</sup>University of Nebraska–Lincoln

#### https://fatplants.net/

**Goals:** The project addresses three goals:

- Systems-guided interrogation of the plastid bio-factory for enhanced production of fatty acids and predictable production of fatty acids with tailored chain-lengths
- Synthetic biology tool development for predictable and high-throughput oilseed crop engineering
- Integration of the redesigned plastid bio-factory with extra-plastidial metabolism for enhanced oils and biocontainment.

B5 will address the imperative need for sustainable liquid fuels and oils of defined structures desired by the U. S. bioenergy and oleochemical sectors. The project will integrate fundamental knowledge generation and synthetic biology tool development to predictably and more rapidly develop non-food *Brassicaceae* oilseeds that produce high quantities of oils and oils with tailored fatty acid compositions (see Fig. 13, p. 75). B5's multidisciplinary team will interrogate plastid metabolic circuitry for carbon flux through fatty acid biosynthesis in seeds of the *Brassicaceae*, pennycress, and *Camelina*. Focus on both species will generate rules for next-generation metabolic engineering of *Brassicaceae* oilseeds and provide higher-value and broader cover/ rotation crop options for U.S. farmers. B5 efforts will be guided by mathematical models as well as biochemical data



**Fig. 13.** B5: Bigger and better oil through integrated approaches across aims. [Courtesy B5 Team]

acquired from seeds of metabolically "extreme" species that produce exceptionally high levels of medium-chain fatty acids. In concert, B5 will develop synthetic biology tools to deliver transgene combinations into defined genome regions and advanced gene-editing methods for tunable upor down-regulation or replacement of endogenous genes. Aided by a comprehensive analytical learning platform and computational models, B5 will integrate data and toolsets to develop enhanced pennycress and Camelina germplasm through design, build, test, learn (DBTL) cycles. Given the central metabolic role of fatty acids in the cell, robust and integrated DBTL cycles will be key to discovering how plants "fight back" against lipid metabolic remodeling. The high-quality genome sequences, plethora of genomic resources, existing metabolic engineering toolbox, and simple Agrobacterium-based floral infiltration transformation systems make both pennycress and Camelina ideal crops for modified oil production. These attributes will also accelerate synthetic biology chassis optimization and introduction of genetic biocontainment technology for safe, sustainable production on marginal and underutilized land across wide portions of the United States. The availability of U.S. Department of Agriculture-Animal and Plant Health Inspection Service regulated field sites and a high-throughput camera-based phenotyping system will facilitate agronomic evaluation of engineered germplasm under diverse environments. In addition to significant fundamental and translational outputs, B5 will further develop extant databases (e.g., FatPlants, ARALIP) for the scientific community and train nine undergraduates, six graduate students, and 11 postdoctoral scientists as the next generation of investigators to tackle U.S. and global energy security, natural resources, and environmental challenges.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0023142.

#### FatPlants: A Comprehensive Information System for Lipid-Related Genes and Metabolic Pathways in Plants

Chunhui Xu<sup>1</sup>, Trey Shaw<sup>1</sup>, Sai Akhil Choppararu<sup>1</sup>, Yiwei Lu<sup>1</sup>, Matt Hudson<sup>1</sup>, Brock Weekley<sup>1</sup>, Michael Fisher<sup>1</sup>, Fei He<sup>1</sup>, Nicholas Wergeles<sup>1</sup>, Philip D. Bates<sup>2</sup>, Abraham J. Koo<sup>1</sup>, Doug K. Allen<sup>3,4</sup>, Timothy P. Durrett<sup>5\*</sup> (tdurrett@ ksu.edu), **Edgar B. Cahoon<sup>6\*</sup>** (ecahoon2@unl.edu), Jay J. Thelen<sup>1</sup>, Dong Xu<sup>1</sup>

<sup>1</sup>University of Missouri; <sup>2</sup>Washington State University; <sup>3</sup>Agricultural Research Service, U.S. Department of Agriculture; <sup>4</sup>Donald Danforth Plant Science Center; <sup>5</sup>Kansas State University; <sup>6</sup>University of Nebraska

#### https://fatplants.net/

**Goals:** The team will develop a dedicated web resource that provides a "one-stop" solution for plant acyl-lipid metabolism so that the research community can use it to study lipid science, model lipid networks, and pursue their own hypotheses.

Increasing seed oil content for biofuels and bioproducts by breeding and biotechnology has resulted in trade-offs or penalties with respect to protein content, seed size, or seed fitness. The molecular basis for this impasse is mostly speculative. Use of current global profiling approaches to better understand both the metabolic consequences of altered oil content and composition and the basis for reduced yield must also deal with off-target genetic mutations, ultimately confounding cause-effect interpretations. The team proposes a diverse, integrated strategy to study the consequences of higher and tailored lipid production by studying transgenic plants specifically engineered to produce altered seed oil content and composition. As a continuation of a prior project, researchers are developing a "one-stop-shop" community web resource for all data pertaining to modifying oil composition and increasing oil content in plants, and to leverage data generated from this project with curated forms of public data from other funded websites and the literature. The FatPlants framework and tools currently exist for a number of crop and model oilseeds, including Camelina sativa. As part of the B5 project, the team is expanding these resources to include pennycress (Thlaspi arvense) and Cuphea viscosissima, an "extreme" producer of seeds with medium-chain fatty-rich oils. Researchers will present all the known fatty acid related proteins and genes in these species and overlay these data with lipidomic measurements from seeds of B5 target species. As a comparative analysis tool, FatPlants includes pathway viewer, protein structure viewer, Basic Local Alignment Search Tool, protein-protein interaction viewer, and Gene Ontology enrichment viewer. To strengthen interactions

among B5 investigators, a user authentication internal data-sharing space has been provided to all collaborative labs. The website is publicly available as a community tool at www.fatplants.net.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0023142.

#### Genome-Wide Gene Regulation by Transcriptional CRISPRa/i Tools in Nonmodel Bacteria

Cholpisit Kiattisewee\* (cholpk@uw.edu), Ian D. Faulkner, Jesse G. Zalatan, **James M. Carothers** 

University of Washington

#### https://sites.google.com/view/carothersresearchgroup

**Goals:** CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi) are modular tools that can regulate gene expression of both heterologous and endogenous genes of microorganisms. The project's goal is to use CRISPRa/i to build large gene regulatory networks (GRNs) spanning more than 25 genes. This study will establish a new paradigm for genome-wide design and significantly improve the ability to engineer microbes for next-generation bioproduction applications.

CRISPRa, developed by this group, is an emerging tool for transcriptional regulation in bacteria providing the ability to modulate gene expression in trans without direct modification at the DNA target (Dong et al. 2018; Fontana et al. 2020). The team also characterized the rules for effective CRISPRa in Escherichia coli and Pseudomonas putida, potential chassis for aromatic compounds bioproduction, where the rules governing high-functional CRISPRa are portable across organisms (Fontana et al. 2020; Kiattisewee et al. 2021). With incorporation of protospacer adjacent motif-flexible dCas9 proteins, researchers can target almost any endogenous gene in the bacterial genome (Kiattisewee et al. 2022), and by combining with CRISPRi gene repression, gene expression can be fine-tuned by both up- and down-regulation to any desired expression level (Tickman et al. 2021). In this study, researchers have demonstrated that CRISPRa/i tools can be used to control heterologous gene expression for bioproduction of various fine chemicals. The team has also investigated CRISPRa/i of more than 25 endogenous genes related to carbohydrate, amino acid, and fatty acid metabolism. This CRISPRa/i platform should provide the ability to combine heterologous and endogenous gene regulations and further accelerate Design-Build-Test-Learn cycles of strains engineering in nonmodel bacteria.

Dong, C., et al. 2018. "Synthetic CRISPR-Cas Gene Activators for Transcriptional Reprogramming in Bacteria," *Nature Communications* 9(1), 2489. DOI:10.1038/s41467-018-04901-6.

Fontana, J., et al. 2020. "Effective CRISPRa-Mediated Control of Gene Expression in Bacteria Must Overcome Strict Target Site Requirements," *Nature Communications* **11**(1), 1618. DOI:10.1038/ s41467-020-15454-y.

Kiattisewee, C., et al. 2021. "Portable Bacterial CRISPR Transcriptional Activation Enables Metabolic Engineering in *Pseudomonas putida*," *Metabolic Engineering* **66**, 283–95. DOI:10.1016/j.ymben.2021.04.002.

Kiattisewee, C., et al. 2022. "Expanding the Scope of Bacterial CRISPR Activation with PAM-Flexible dCas9 Variants," *ACS Synthetic Biology* **11**(12), 4103–12. DOI:10.1021/acssynbio.2c00405.

Tickman, B. I., et al. 2021. "Multi-Layer CRISPRa/i Circuits for Dynamic Genetic Programs in Cell-Free and Bacterial Systems," *Cell Systems* **13**(3), 215–29.e8. DOI:10.1016/j.cels.2021.10.008.

**Funding Information:** This research was supported by the U.S. Department of Energy Bioenergy Technologies Office (BETO), grant No. DE-EE0008927 and DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0023091.

#### Multiomics-Driven Microbial Model Optimization

Janis Shin (jshin1@uw.edu)<sup>1\*</sup>, Herbert M. Sauro<sup>1</sup>, Jeremy Zucker<sup>2</sup>, **James M. Carothers**<sup>1</sup>

<sup>1</sup>University of Washington; <sup>2</sup>Pacific Northwest National Laboratory

https://sites.google.com/view/carothersresearchgroup

**Goals:** The project's goal is to create genome-scale models of endogenous metabolic pathways and develop metabolic sensitivity maps to identify reactions that dominate the control of flux. Specifically, the team aims to gain insight into the regulatory mechanisms within pathways and predict outcomes of metabolic interventions at the genome scale.

Biomanufacturing poses a sustainable approach to wean humanity's reliance on petrochemical-derived commodity products. Despite the advent of omics data and genomescale models, there is no straightforward process for integrating all this data to design biochemical pathways that produce chemicals at an industrial scale. To understand and engineer metabolism, researchers must identify which enzymes exert the most influence on metabolite concentrations and fluxes through the biochemical pathway. Theoretical work also suggests that the cast of enzymes exerting control over the pathway changes under different growth conditions. To identify these influential enzymes, steady state enzyme perturbation data is used within a genomescale context containing multiple metabolic engineering interventions. Researchers approximate Michaelis-Menten kinetics near the reference steady state through a lin-log model and supply these calculations to a Bayesian inference model. The inference model estimates each reaction's influence on the metabolic pathway and thus provides metabolic intervention targets for improving bioproduction titers and rates. This method was successfully applied to estimate sensitivities in yeast metabolism (St. John et al. 2019; McNaughton et al. 2021). The team is extending the approach to *Pseudomonas putida*.

McNaughton, A. D., et al. 2021. "Bayesian Inference for Integrating *Yarrowia lipolytica* Multiomics Datasets with Metabolic Modeling," *ACS Synthetic Biology*, **10**(11), 2968–81.

St. John, P. C., et al. 2019. "Bayesian Inference of Metabolic Kinetics from Genome-Scale Multiomics Data," *PLoS Computational Biology* **15**(11), e1007424.

**Funding Information:** This research was supported by the U.S. Department of Energy Bioenergy Technologies Office (BETO), grant No. DE-EE0008927 and DOE Office of Science, Biological and Environmental Research (BER) Program, under Funding Opportunity Announcement DE-FOA-0002600.

#### Pooled Microbial CRISPR Screens using Single-Cell RNA Sequencing

Jacob R. Brandner<sup>1\*</sup> (brandner@uw.edu), Quoc Tran<sup>1</sup>, Jesse G. Zalatan<sup>1</sup>, Georg Seelig<sup>1</sup>, **James M. Carothers**<sup>1</sup>, **Anna Kuchina**<sup>2</sup>

<sup>1</sup>University of Washington; <sup>2</sup>Institute For Systems Biology

http://www.carothersresearch.com

http://www.kuchina.isbscience.org

**Goals:** The goal is to design genome-wide CRISPRa/i programs for carbon-conserving bioproduction. To achieve this goal, researchers will develop new approaches for high-throughput analysis powered by single-cell RNA sequencing.

CRISPR activation and interference tools have transformed the ability to reprogram microbial hosts for bioproduction. However, building large genetic programs is a time-consuming and iterative process due to the limited understanding of host metabolic and transcriptional regulatory networks. The goal is to develop a custom bacterial single-cell RNA sequencing platform to profile the impact of multi-gene CRISPR gene regulatory programs on thousands of transcriptomes. This approach will provide a high-throughput, low-cost, and information-rich technology to investigate design rules for CRISPR activation and interference (CRISPRa/i), identify heterogeneity in engineered strains, and rapidly assess both intended and unintended transcriptional responses from CRISPR programs. In eukaryotic systems, similar approaches have transformed the ability to interrogate gene function and delineate regulatory networks, but these methods have not yet been implemented in bacteria.

Here, researchers have applied a custom microbial single-cell RNA sequencing platform (microSPLiT) to profile the impact of CRISPRa perturbations on transcriptomic states in engineered *E. coli*. To validate the platform, researchers targeted genes involved in aromatic amino acid biosynthesis. Single-cell analysis revealed distinct gene expression signatures and variable stress responses for the CRISPRa targets despite belonging to the same metabolic pathway. These results demonstrate that the platform can provide information-rich readouts from CRISPRa programs for high-throughput metabolic engineering in bacteria.

**Funding Information:** This research was supported by the U.S. Department of Energy Bioenergy Technologies Office (BETO), grant no. DE-EE0008927 and DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0023091.

#### A Pipeline for High-Throughput Genomic Recoding in Organisms Beyond *E. coli*

Huseyin Tas<sup>1,2\*</sup> (huseyin\_tas@hms.harvard.edu), Anush Chiappino-Pepe<sup>1,2</sup>, Max G. Schubert<sup>1,2</sup>, **George M. Church<sup>1,2</sup>** 

<sup>1</sup>Harvard Medical School; <sup>2</sup>Harvard University

http://arep.med.harvard.edu

**Goals:** This project presents a pipeline for construction of genomically recoded nonmodel organisms with a focus on a 5.9 Mb *Pseudomonas putida* genome to harbor 59 codons. The team designed the 59-codon genome with a newly developed computational model that considers the impact of individual and combinatorial codon changes in gene expression, gene function, and growth. Required input datasets have been obtained for the computational model, and new genome engineering tools have been developed to optimize the introduction of synonymous mutations.

Genomic recoding consists of removing a set of codons from the entire genome while maintaining the same protein sequences (Ostroy et al. 2016, 2020). Recoding provides attractive properties including tight biocontainment, virus resistance, and efficient nonstandard amino acid incorporation (Lajoie et al. 2013; Mandell et al. 2015; Robertson et al. 2021; Nyerges et al. 2022). However, to date, it has been only possible to fully recode the model organism *Escherichia coli* (Ostrov et al. 2020; Fredens et al. 2019). To enable recoding in nonstandard organisms, computational- and experimental-based broad-host methods are needed to fully leverage this enormous potential.

Here, the team presents the datasets, computational tools, and genome engineering technology that are being assembled to enable high-throughput genome recoding in non-standard organisms to validate in *P. putida*. Researchers first generated high-quality genome maps and datasets to identify all relevant genomic elements. The identification of promoters and ribosome-binding sites in the *P. putida* genome was emphasized.

Researchers next applied a computational framework that connects genome sequences with growth using information from metabolism, expression, and regulation to design a 59-codon of *P. putida*. This model predicts a minimum reduction in growth compared to the wild-type *P. putida* strain.

Finally, the team developed highly efficient recombineering tools in *P. putida*, including new recombinases and retronbased approaches to work with both single-stranded DNA and double-stranded DNA. In addition, researchers are working on extending the capacity for recombineeringbased approaches (e.g., CRISPR-associated nucleases).

In summary, this study presents a pipeline to recode organisms beyond *E. coli* and computational and experimental techniques that have been established. This work aims to: (1) validate first fully recoded nonstandard model organism *P. putida-59*; (2) achieve the first environmental bacterium to be biocontained and resistant to viruses; (3) establish design rules for computational and experimental expansion of recoding beyond *E. coli*; and (4) accelerate recoding processes.

Fredens, J., et al. 2019. "Total Synthesis of *Escherichia coli* with a Recoded Genome," *Nature* **569**, 514–18.

Lajoie, M. J., et al. 2013. "Genomically Recoded Organisms Expand Biological Functions," *Science* **342**, 357–60.

Mandell, D. J., et al. 2015. "Biocontainment of Genetically Modified Organisms by Synthetic Protein Design," *Nature* **518**(7537), 55–60.

Nyerges, A., et al. "Swapped Genetic Code Blocks Viral Infections and Gene Transfer," *bioRxiv*, Preprint.

Ostrov, N., et al. 2016. "Design, Synthesis, and Testing Toward a 57-Codon Genome," *Science* **353**(6301), 819–22.

Ostrov, N., et al. 2020. "Synthetic Genomes with Altered Genetic Codes," *Current Opinion in Systems Biology* **24**, 32–40.

Robertson, W. E., et al. 2021. "Sense Codon Reassignment Enables Viral Resistance and Encoded Polymer Synthesis," *Science* **372**(6546), 1057–62.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-FG02-02ER63445.

Dr. Church is a founder of companies in which he has related financial interests: GRO Biosciences, ReadCoor; EnEvolv; and 64-x Bio. For a complete list of Dr. Church's financial interests, see also v.ht/PHNc.

#### Construction of a Synthetic 57-Codon *E. coli* Chromosome to Achieve Resistance to All Natural Viruses, Prevent Horizontal Gene Transfer, and Enable Biocontainment

Akos Nyerges<sup>1\*</sup> (akos\_nyerges@hms.harvard.edu), Anush Chiappino-Pepe<sup>1</sup>, Regan Flynn<sup>1</sup>, Svenja Vinke<sup>1</sup>, Owen Spencer<sup>1</sup>, Shirui Yan<sup>1,3</sup>, Siân V. Owen<sup>1</sup>, Eleanor A. Rand<sup>1</sup>, Michael Baym<sup>1</sup>, Maximilien Baas-Thomas<sup>2</sup>, Nili Ostrov<sup>1</sup>, Alexandra Rudolph<sup>2</sup>, Yue Shen<sup>3</sup>, Ian K. Blaby<sup>4</sup>, Yasuo Yoshikuni<sup>4</sup>, Miranda Harmon-Smith<sup>4</sup>, Matthew Hamilton<sup>4</sup>, **George M. Church**<sup>1</sup>

<sup>1</sup>Harvard Medical School; <sup>2</sup>Harvard University; <sup>3</sup>BGI Group; <sup>4</sup>D0E Joint Genome Institute

http://arep.med.harvard.edu

**Goals:** The project is finalizing the construction of a fully recoded 3.97 Mb *Escherichia coli* genome that relies on the use of only 57 genetic codons. For this aim, the genome was computationally designed, synthesized, and assembled into 88 segments. In the final steps of genome construction, the team combines and optimizes these segments *in vivo* to assemble the fully recoded, viable chromosome. In parallel with the construction of this 57-codon organism, the team investigates whether mobile genetic elements and environmental viruses can overcome the genetic isolation of organisms bearing modified genetic codes.

The team presents the construction of a recoded, 57-codon E. coli genome, in which seven codons are replaced with synonymous alternatives in all protein-coding genes. For this aim, the entirely synthetic recoded genome was assembled as 88 episomal segments 25- to 48-kb, which were individually tested for functionality, and then integrated into the genome. Developing a specialized integration system and optimizing workflow enhanced integration efficiency to 100%, which resulted in an order-of-magnitude increase in construction speed. Researchers are now combining recoded genomic clusters with a novel technology that builds on the latest developments in recombineering and CRISPR-associated nucleases (Wannier et al. 2020, 2021). In parallel with genome construction, researchers developed novel experimental methods to identify fitness-decreasing changes and troubleshoot these cases. Leveraging massively parallel genome editing and accelerated laboratory evolution allowed correction of partially recoded strains' fitness within weeks (Nyerges et al. 2018). As the final assembly of this *E. coli* genome approaches, dependency on nonstandard amino acids is also implemented.

Previous experiments showed that rational genetic code engineering could isolate Genetically Modified Organisms (GMOs) from natural ecosystems by providing resistance to viral infections and blocking horizontal gene transfer (HGT); however, how natural mobile genetic elements and viruses could cross this genetic-code-based barrier remained unanswered. By systematically investigating HGT into E. coli Syn61 $\Delta$ 3, an *E. coli* strain with a synthetic, 61-codon genetic code, researchers discovered that transfer (t) RNAs expressed by viruses and other mobile genetic elements readily substitute cellular tRNAs and abolish genetic codebased resistance to HGT (Nyerges et al. 2022). The team also discovered 12 new bacteriophages in environmental samples that can infect and lyse this 61-codon organism. These viruses express 10 to 27 tRNAs, including functional tRNAs needed to replace the host's missing tRNA genes. Researchers also identified viruses with tRNAs that hold the potential to abolish the virus resistance of the 57-codon organism. These findings suggest that the selection pressure of organisms with compressed genetic codes can facilitate the rapid evolution of viruses and mobile genetic elements capable of crossing a genetic code-based barrier. Therefore, additional genetic biocontainment technologies were developed to simultaneously block GMOs' unwanted proliferation, eliminate viral infections, and prevent transgene escape (Nyerges et al. 2022).

In sum, this genome synthesis work will soon (1) demonstrate the first 57-codon organism, (2) establish a tightly biocontained chassis for new-to-nature protein production, and (3) open a new avenue for the bottom-up synthesis and refactoring of microbial genomes—both computationally and experimentally. Furthermore, this research demonstrates that horizontally transferred tRNA genes of mobile genetic elements and viruses can substitute deleted cellular tRNAs and thus rapidly abolish compressed genetic codes' resistance to viral infections and HGT.

Nyerges, A., et al. 2018. "Directed Evolution of Multiple Genomic Loci Allows the Prediction of Antibiotic Resistance," *Proceedings of the National Academy of Sciences* **115**(25), E5726–35. DOI:10.1073/ pnas.1801646115.

Nyerges, A., et al. 2022. "Swapped Genetic Code Blocks Viral Infections and Gene Transfer," *bioRxiv*, Preprint. DOI:10.1101/2022.07.08.499367.

Wannier, T. M., et al. 2020. "Improved Bacterial Recombineering by Parallelized Protein Discovery," *Proceedings of the National Academy of Sciences* **117**(24), 13689–98. DOI:10.1073/pnas.2001588117.

Wannier, T. M., et al. 2021. "Recombineering and MAGE," *Nature Reviews Methods Primers* 1(1), 1–24. DOI:10.1038/s43586-020-00006-x.

**Funding Information:** This project has been funded by DOE grant DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related financial interests: ReadCoor; EnEvolv (Ginkgo Bioworks); and 64x Bio. Harvard Medical School has filed provisional patent applications related to this work on which Akos Nyerges and George M. Church are listed as inventors. For a complete list of Dr. Church's financial interests, see also v.ht/PHNc.

## Domesticating Cyanobacteria Through Development of New Genome Engineering Tools and Isolation of New Bio-Prospected Model Strains

Max G. Schubert<sup>1\*</sup> (max\_schubert@hms.harvard.edu), Tzu-Chieh Tang<sup>1</sup>, Isabella Goodchild-Michelman<sup>1</sup>, Krista Ryon<sup>2</sup>, James Henriksen<sup>3</sup>, Paola Quatrini<sup>4</sup>, Marco Milazzo<sup>4</sup>, Gabriele Turco<sup>4</sup>, Davide Spatafora<sup>4</sup>, Braden Tierney<sup>2</sup>, Christopher Mason<sup>2</sup>, **George M. Church<sup>1</sup>** 

<sup>1</sup>Harvard University; <sup>2</sup>Weill Cornell Medical College; <sup>3</sup>Colorado State University; <sup>4</sup>Palermo University

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

**Goals:** Cyanobacteria are facile models of photosynthesis and chassis organisms for carbon-negative bioproduction. To help advance the development of cyanobacteria biotechnology, a team of researchers is both developing new genetic tools for cyanobacteria and bioprospecting novel cyanobacteria model strains. New genetic tools aim toward genome-scale study of cyanobacterial genomes, including applications like recoding and directed evolution of bacterial genomes, and toward improved performance in bioproduction. Novel cyanobacterial model strains possess unique differences to common model strains. They improve photosynthetic microbial bioproduction, thus further enabling the carbon-neutral bioeconomy.

Facile genome engineering tools further the ability to explore and apply biology, and previous work generating recoded and/or biocontained organisms, or performing directed evolution in bacteria, shows that recombineering is a critical tool (Mandell et al. 2015; Lajoie et al. 2013; Schubert et al. 2021). Recombineering is enabled by phage sequential structure alignment program (SSAP) proteins, which both produce and stabilize single-stranded DNA recombination intermediates and recruit them to the replisome (Filsinger et al. 2021; Caldwell et al. 2019). These proteins are known to have host-specific activity, and finding SSAPs that function efficiently in a specific clade of bacteria is a first step toward improving tools for precision genome engineering (Filsinger et al. 2021; Wannier et al. 2020, 2021). Researchers have identified 22 candidate SSAPs within both protein databases and metagenomic databases occurring within cyanobacteria or their phages. These proteins could extend recombineering approaches enabling multiplex engineering, recoding, and genome-scale remodeling into cyanobacteria and help form a roadmap for identifying efficient SSAPs in new model microbes.

Recombineering is one tool for engineering genomes that is well-suited to rational approaches. In contrast, transposon mutagenesis and transposon insertion sequencing (TnSeq) have been successful at improving understanding of genomes through irrational approaches and fast generation of large pools of genetic diversity (Gray et al. 2015). This project demonstrates that simple modifications to existing transposon mutagenesis procedures result in many random transposon insertions within cyanobacterial genomes. The resulting strains with large numbers of inactivated or upregulated genes suggest strategies for directed evolution and genome streamlining.

The project is also seeking to apply these technologies in unreported cyanobacterial model strains that could improve photosynthetic-based bioproduction. Indeed, various cyanobacterial model strains are used in the literature for these efforts as well as studying the fundamental processes of photosynthesis (Goodchild-Michelman et al. 2023). Together with the Two Frontiers project (twofrontiers. org), which aims to explore microbial communities in extreme environments (including high CO<sub>2</sub> environments), researchers have isolated two closely related cyanobacterial strains with promising growth phenotypes from seawater in the photic zone off the coast of Sicily. The two strains have approximately 99% nucleotide identity with each other with 32,897 single nucleotide polymorphisms differing between the strains and approximately 98% identity to the closest relative with genome sequence available, Cyanobacterium aponium PCC10605. Comparative genomics reveals each novel strain possesses 50 to 60 unique genes differing between both these strains and PCC10605, and 108 shared genes differing between them and PCC10605. One novel strain remarkably grows to a higher density in batch growth than Synechococcus sp. PCC11901, which holds the published record for high density batch biomass growth in cyanobacteria (Włodarczyk et al. 2020; Mills et al. 2022). Research shows that larger and more dense cells than common model strains may improve the economics of dewatering cyanobacterial biomass, and thus production of bioproducts. The second strain possesses unique characteristics such as programmed formation of large aggregates and phototactic motility. In sum, these strains obtained from a CO<sub>2</sub>-emitting volcanic vent are a promising new model for studies in cyanobacteria and possibly for photosynthetic bioproduction.

Caldwell, B. J., et al. 2019. "Crystal Structure of the Red $\beta$ C-Terminal Domain in Complex with  $\lambda$  Exonuclease Reveals an Unexpected Homology with  $\lambda$  Orf and an Interaction with *Escherichia coli* Single Stranded DNA Binding Protein," *Nucleic Acids Research* 47, 1950–63.

Filsinger, G. T., et al. 2021. "Characterizing the Portability of Phage-Encoded Homologous Recombination Proteins," *Nature Chemical Biology* **17**, 394–402.

Goodchild-Michelman, I. M., et al. 2023. "Light and Carbon: Synthetic Biology Toward New Cyanobacteria-Based Living Biomaterials," *Materials Today Bio* **19**, 100583. DOI:10.1016/ j.mtbio.2023.100583.

Gray, A. N., et al. 2015. "High-Throughput Bacterial Functional Genomics in the Sequencing Era," *Current Opinion in Microbiology* **27**, 86–95.

Lajoie, M. J., et al. 2013. "Genomically Recoded Organisms Expand Biological Functions," *Science* **342**, 357–60.

Mandell, D. J., et al. 2015. "Biocontainment of Genetically Modified Organisms by Synthetic Protein Design," *Nature* **518**, 55–60.

Mills, L. A., et al. 2022. "Development of a Biotechnology Platform for the Fast-Growing Cyanobacterium *Synechococcus* sp. PCC 11901," *Biomolecules* **12**(7), 872.

Schubert, M. G., et al. 2021. "High-Throughput Functional Variant Screens via *In Vivo* Production of Single-Stranded DNA," *Proceedings of the National Academy of Sciences* **118**(18), e2018181118.

Wannier, T. M., et al. 2020. "Improved Bacterial Recombineering by Parallelized Protein Discovery," *Proceedings of the National Academy of Sciences* **117**, 13689–98.

Wannier, T. M., et al. 2021. "Recombineering and MAGE," *Nature Reviews Methods Primers* 1, 1–24.

Włodarczyk, A., et al. 2020. "Newly Discovered *Synechococcus* sp. PCC 11901 is a Robust Cyanobacterial Strain for High Biomass Production," *Communications Biology* **3**, 215.

**Funding Information:** This project has been funded by DOE grant DE-FG02-02ER63445. Dr. Church is a founder of companies in which he has related financial interests: EnEvolv (Ginkgo Bioworks); and 64x Bio. For a complete list of Dr. Church's financial interests, see arep.med.harvard. edu/gmc/tech.html.

#### From Strand Design Principles to SNP Detection—Probing Oligonucleotide Hybridization at the Single Molecule Level

Johannes Stein<sup>1,2\*</sup> (johannes.stein@wyss.harvard.edu), Lorenzo Magni<sup>1</sup>, Peng Yin<sup>1</sup>, Chao-ting Wu<sup>2</sup>, **George M. Church<sup>1,2</sup>** 

<sup>1</sup>Harvard University; <sup>2</sup>Harvard Medical School

**Goals:** Advance understanding of oligonucleotide hybridization at the single molecule level in order to maximize detection sensitivity and throughput in DNA-based highly multiplexed fluorescence microscopy applications.

Recent advances in DNA nanotechnology have become a major driver of highly multiplexed and super-resolution fluorescence microscopy (Beliveau et al. 2012; Porreca et al. 2007; Boyle et al. 2011). Clever sequential schemes of *in situ* sequencing and barcoding together with both libraries of Oligopaint (oligonucleotide-based primary hybridization) probes (targeting DNA and RNA) and with DNA-functionalized antibodies (targeting proteins) can provide visual access to a vast number of targets in the genome, transcriptome, or proteome with subcellular resolution (Beliveau et al. 2012; Larsson, Frisén, and Lundeberg 2021; Bouwman, Crosetto, and Bienko 2022; Zhuang 2021; Jerkovic' and Cavalli 2021; Hickey et al. 2022). Signalamplification strategies such as rolling-circle amplification (Lee et al. 2014), linear appending (Kishi et al. 2019), or a minimum number of hybridization probes per RNA/DNA (Wang et al. 2016) target allow robust detection yet come at the cost of resolution and limited throughput. This work sets out to establish a simple single-molecule approach to assay the absolute efficiency of successful 1:1 probe hybridization events, given a known number of surface-immobilized DNA origami each carrying just a single copy of the complementary target site. Researchers further quantify the influences of oligo purification, directionality, and length of single-stranded overhang and derive design principles that maximize target hybridization and, hence, minimize the need for signal amplification. In a final proof-of-concept, the team demonstrates the versatility of this approach quantifying single nucleotide polymorphism detection efficiencies.

Beliveau, B. J., et al. 2012. "Versatile Design and Synthesis Platform for Visualizing Genomes with Oligopaint FISH Probes," *Proceedings* of the National Academy of Sciences **109**(52), 21301–06.

Bouwman, B. A. M., N. Crosetto, and M. Bienko. 2022. "The Era of 3D and Spatial Genomics," *Trends in Genetics* **38**(10), 1062–75.

Boyle, S., et al. 2011. "Fluorescence *In Situ* Hybridization with High-Complexity Repeat-Free Oligonucleotide Probes Generated by Massively Parallel Synthesis," *Chromosome Research* **19**, 901–09.

Hickey, J. W., et al. 2022. "Spatial Mapping of Protein Composition and Tissue Organization: A Primer for Multiplexed Antibody-Based Imaging," *Nature Methods* **19**, 284–95.

Jerkovic', I., and G. Cavalli. 2021. "Understanding 3D Genome Organization by Multidisciplinary Methods," *Nature Reviews Molecular Cell Biology* **22**, 511–28.

Kishi, J. Y., et al. 2019. "SABER Amplifies FISH: Enhanced Multiplexed Imaging of RNA and DNA in Cells and Tissues," *Nature Methods* **16**, 533–44.

Larsson, L., J. Frisén, and J. Lundeberg. 2021. "Spatially Resolved Transcriptomics Adds a New Dimension to Genomics," *Nature Methods* **18**, 15–18.

Lee, J. H., et al. 2014. "Highly Multiplexed Subcellular RNA Sequencing *In Situ*," *Science* **343**, 1360–63.

Porreca, G. J., et al. 2007. "Multiplex Amplification of Large Sets of Human Exons," *Nature Methods* **4**, 931–36.

Wang, S., et al. 2016. "Spatial Organization of Chromatin Domains and Compartments in Single Chromosomes," *Science* **353**, 598–602.

Zhuang, X. 2021. "Spatially Resolved Single-Cell Genomics and Transcriptomics by Imaging," *Nature Methods* **18**, 18–22.

**Funding Information:** This project has been funded by DOE grant DE-FG02-02ER63445 (to GMC), NIH SRM1HG011016-03 (to CTW) and by the European Molecular Biology Organization ALTF 816-2021 (to JS). Dr. Church is a founder of companies in which he has related financial interests: ReadCoor; EnEvolv; and 64-x. For a complete list of Dr. Church's financial interests, see also v.ht/PHNc. Dr. Wu holds or has patent filings pertaining to imaging, and her laboratory has held a sponsored research agreement with Bruker Inc. Although non-equity holding, Dr. Wu is a co-founder of Acuity Spatial Genomics and, through personal connections to George Church, has equity in companies associated with him, including 10x Genomics and Twist.

#### Developing Genome Engineering Technologies in *Cupriavidus necator* for Carbon-Negative Biomanufacturing

Akira Nakamura\* (akira.nakamura@yale.edu), Jaymin Patel, Joonseok Oh, Shenqi Wang, Laura Quinto, Jason Crawford, **Farren J. Isaacs** 

#### Yale University

The accelerating climate crisis combined with rapid population growth poses some of the most urgent challenges to humankind. A major contributing factor to this crisis is the unabated release and accumulation of CO<sub>2</sub> across the biosphere. Researchers can take advantage of this abundance of available CO<sub>2</sub> to transform the way the world produces and uses carbon by engineering CO<sub>2</sub>-fixing biosystems to produce commodity fuels and chemicals. Engineering efforts in CO<sub>2</sub>-fixing organisms are currently limited by an incomplete understanding of genotype-phenotype relationships and inefficient genome engineering tools to discern these relationships. To address such issues, researchers have developed an integrated computational and experimental workflow [computer-aided design of synthetic genetic elements (CAD-SGE)] to redesign multigene biological pathways for mobilization, expression, and characterization in versatile organisms (Patel et al. 2022). Heterologous expression and characterization of large pathways (i.e., tyrocitabine pathway, violacein pathway) have been demonstrated using this CAD-SGE technology in diverse prokaryotes and eukaryotes, including Escherichia coli, Pseudomonas putida,

Klebsiella aerogenes, Salmonella enterica, and Saccharomyces cerevisiae. Here, researchers demonstrate the expansion of the landing pad-based mobilization strategy into *Cupria-vidus necator*, an aerobic autotroph that utilizes the Calvin cycle. Genomic integration of the landing pad will allow streamlined integration and expression of large heterologous genetic elements in *C. necator*. By implementing the CAD-SGE technology in *C. necator* and other  $CO_2$ -fixing organisms, researchers hope to improve the understanding of  $CO_2$ -utilizing microbes and contribute to the development of sustainable and carbon-neutral technologies.

Patel, J., et al. 2022. "In Brief Cross-Kingdom Expression of Synthetic Genetic Elements Promotes Discovery of Metabolites in the Human Microbiome," *Cell* **185**. DOI:10.1016/j.cell.2022.03.008.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-FOA-0002600.

#### Novel Systems Approach for Rational Engineering of Robust Microbial Metabolic Pathways

Laura Jarboe<sup>1\*</sup> (Ijarboe@iastate.edu), Yannick J. Bomble<sup>2</sup>, Robert L. Jernigan<sup>1</sup>, Peter St. John<sup>2</sup>, Kejue Jia<sup>1</sup>, Pranav M. Khade<sup>1</sup>, Ambuj Kumar<sup>1</sup>, Onyeka Onyenemezu<sup>1</sup>, Jetendra K. Roy<sup>1</sup>, Chao Wu<sup>2</sup>

#### <sup>1</sup>Iowa State University; <sup>2</sup>National Renewable Energy Laboratory

Goals: The goal of this project is to develop and implement a process for improving bioproduction under conditions that are appealing for industrial processes, such as high temperature and low pH. This approach addresses the failure of metabolic reactions due to inhibition, denaturation, misfolding or disorder of enzymes. Researchers will develop and implement a framework that identifies these enzymes and then identifies their robust replacement enzymes. The engineering strategy of replacing enzymes to improve bioproduction is well established, but rarely applied to system-wide stressors. Researchers apply a systems-genomics approach to improve bioproduction, with Escherichia coli as the model organism. Butanol production at high temperature and succinate production at low pH are the model systems. This approach is complementary to improvement of microbial robustness by engineering the cell membrane and has advantages relative to evolutionary-based organism improvement by prioritizing bioproduction rather than growth.

**Temperature Tolerance:** Given the increase in thermostability data available from recent proteomics studies, researchers are taking a closer look at which enzymes are limiting growth at higher temperatures using metabolic models. While the melting temperature of an enzyme is useful, knowledge of the range of temperatures a given enzyme can withstand and remain active can help provide a more complete picture of an organism's sensitivity to temperature. Researchers are using embeddings from protein language models such as ESM-2 to predict enzyme temperature optimum (as well as the melting temperature when not available) and have found improvements over previous methods (e.g., ProTstab2, DeepET). Researchers are exploring ways to improve predictive models such as by integrating protein structures with embeddings from the language models. Growth-limiting enzymes in response to increasing temperature in *E. coli* have been previously identified through the use of genome-scale models (Chang et al. 2013). This prior analysis relied on estimated values for many of the key enzyme-specific temperature parameters. In the years since this original study, there have been multiple proteomic studies of enzyme melting temperature  $(T_m)$  and researchers have used this data, along with literature reports of enzyme assays and the ProTstab2 Tm predictors (Yang et al. 2019) to compile an updated set of *E. coli* T<sub>m</sub> values for use in assessment of enzyme thermosensitivity and prioritizing enzymes for experimental assessment.

Acid Tolerance: The pH tolerance efforts have prioritized modeling the effect of pH on the allocation of cellular resources. In the current version of the model, researchers assume that the cell interior is maintained at near-neutral pH by the energetic investment in proton extrusion via ATP. Researchers have completed a two-stage optimization model. In the first stage, cell growth is maximized with constraints of thermodynamics and total available enzyme protein allocation. The flux distributions associated with this maximum growth are then used in a minimization of overall enzyme protein cost. This model predicts a sharp decrease in specific growth rate when the extracellular pH drops below 4.9. Next, researchers investigated the thermodynamic and kinetic properties of individual enzymes in the model. Generally, low pH will lead to additional cost of enzyme protein to maintain a thermodynamic feasibility and reaction rate, represented by the thermodynamic item and kinetic item, respectively. Interestingly, the glucose transporter and glycolysis enzymes account for the majority of those with the highest enzyme protein requirement especially at low pH.

Researchers have compiled *E. coli* enzyme pH sensitivity data from the literature and are in the process of assessing the accuracy of the patcHwork sequence/structure-based tool for estimating enzyme pH sensitivity (https://patchwork. biologie.uni-freiburg.de/upload.php). Among the enzymes with literature data available, phosphatidylserine synthase (PssA) is especially pH sensitive. Researchers tested the effect of increased expression of PssA on growth at a range of pH values. At values as low as pH 4.0, the increased expression of PssA was beneficial for growth. Efforts to increase the pH tolerance of the PssA enzyme are ongoing. **Modeling of Enzyme Stability:** The investigations of existing predictions of protein stability found them to be unreliable. Researchers are applying protein language models such as ES-M1 to the problem but have found similar DNA codon language models (Outeiral and Deane 2022) that will be more effective for this type of prediction. The team is also developing new computational machine learning models using this DNA language model for predicting protein melting temperatures. (This relates to the importance of specific codons for the rate of translation as it affects cotranslational folding.) Researchers also plan to extend this type of model to the prediction of optimal pH for *E. coli* enzymes, which is a relatively under-developed problem.

This builds upon the team's own expertise in utilizing protein language models to make major improvements in protein sequence matching and function prediction (Kilinc, Jia, and Jernigan 2023).

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0022090.

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided U. S. Department of Energy, Office of Science, through the Genomic Science program of the Biological and Environmental Research program under FWP ERW3526. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.

#### **A Cell-Free Protein Evolution Platform**

Grant Landwehr<sup>1\*</sup> (grantlandwehr2024@u.northwestern. edu), Bastian Vögeli,<sup>1</sup> **Michael C. Jewett<sup>1,2</sup>** 

<sup>1</sup>Northwestern University; <sup>2</sup>Stanford University

The biosynthesis of high-value sustainable chemicals is a major goal of synthetic biology and securing the energy future. For many important molecular biotransformations, efficient enzymes have yet to be discovered or engineered. High-throughput methods to prepare enzyme mutants and measure their activity—as well as evaluate their engineering potential and transform them into industrial tools—are important to overcoming this bottleneck. Researchers addressed this challenge by developing a cell-free DNA assembly and protein-synthesis platform, allowing researchers to rapidly screen thousands of sequence-defined enzyme mutants in iterative design-build-test-learn cycles. These rich datasets are also amenable to machine-learning algorithms that try to capture the sequence-fitness landscape. As a model, researchers demonstrate the utility of the

platform by engineering a formyl-CoA synthetase for the activation of formate, a transformation that does not readily occur in nature and has implications in building synthetic carbon-fixing metabolic pathways. To design, researchers used molecular structure and sequence homology to guide selection of amino acid residues for creating new-to-nature enzymes that catalyze the desired chemistry. To build, researchers used site-saturation mutagenesis to create a DNA library of enzyme variants containing mutations at the identified residues. To test, researchers developed a visual detection scheme to bypass the need for traditional chromatography-based analytics. The team anticipates the work engineering a formyl-CoA synthetase will serve as a blueprint to enable future efforts aimed at the maturation of a diverse array of biocatalysts.

**Funding Information:** This work was supported by DOE Office of Science, Office of Biological and Environmental Research (BER), grant DE-SC0023278, and an NSF Graduate Research Fellowship to G.M.L. (DGE-1842165).

#### Accelerating Carbon-Negative Biomanufacturing Through Systems-Level Biology and Genome Optimization

Nicholas Fackler<sup>1\*</sup> (Nick.Fackler@lanzatech.com), Christine Brown<sup>1</sup>, Hannah Ranft<sup>1</sup>, Tadas Kolpakovas<sup>1</sup>, Alexander P. Mueller<sup>1</sup>, Heidi Schindel<sup>1</sup>, Vinicio Reynoso<sup>1</sup>, Shilpa Nagaraju<sup>1</sup>, Rasmus O. Jensen<sup>1</sup>, Ching Leang<sup>1</sup>, Michael Köpke<sup>1</sup>, **Michael C. Jewett<sup>2,3</sup>** 

<sup>1</sup>LanzaTech, Inc.; <sup>2</sup>Northwestern University; <sup>3</sup>Stanford University

Goals: The accelerating climate crisis combined with rapid population growth poses some of the most urgent challenges to humankind, all linked to the unabated release and accumulation of CO<sub>2</sub> across the biosphere. By harnessing the capacity to leverage chemoautotrophic gas-fermenting microorganisms, the project can begin to take advantage of this abundance of available  $CO_2$  to transform the way the world creates and uses carbon-based materials. This interdisciplinary project will address existing challenges around genomic optimization of chemoautotrophic gasfermenting microorganisms to establish versatile and efficient CO<sub>2</sub>-utilizing biosystems. The team will use in silico, in vitro, and in vivo methods to develop novel tools for genome engineering while advancing systems-level knowledgebases for industrially relevant CO<sub>2</sub>-fixing organisms. These new developments will be deployed towards generating streamlined genomes in Clostridium autoethanogenum under the contexts of bioproduction and biocontainment.

The project seeks to develop and integrate innovative cell-free tools, genome-engineering techniques, and

machine learning–based methods for predictive design of  $CO_2$ –fixing biosystems that deliver new routes to solve energy security and environmental stewardship challenges. The key idea is to develop (1) genome-scale cell-free tools for proteome study and design, (2) genome-engineering techniques for systems-level design, (3) high-throughput cell-free enzyme engineering approaches, (4) machine learning and molecular tools to guide enzyme and product selection and engineering, and (5) transcription factor–based biosensors for genetic regulation. The team then aims to apply these new tools across industrially relevant  $CO_2$ –fixing organisms.

Anaerobic acetogens (specifically the model acetogen, *C. autoethanogenum*) have emerged as sustainable biomanufacturing platforms capable of producing valuable chemicals from flexible nonfood and waste feedstocks that are already used at a commercial scale by LanzaTech today, building on their ability to natively ferment  $CO_2$  or carbon monoxide and produce a valuable product (ethanol) with high selectivity. Advances over the past decades including genome-engineering tools, cell-free prototyping, and metabolic models have enabled carbon-negative production of commodity chemicals such as acetone and isopropanol in engineered acetogens at industrially relevant productivities for prolonged periods (Liew et al. 2022).

LanzaTech has previously published a polished genome sequence along with metabolomic, proteomic, and transcriptomic profiles. These datasets have enabled synthesis of sophisticated genome-scale metabolic models and computational analysis for the genome of *C. autoethanogenum* (Brown et al. 2014; Tschaplinski et al. 2019). From these datasets, approximately 19% of the genome has been identified as essential for autotrophic growth (Woods et al. 2022). Even for scientists armed with developed models, automated geneengineering capabilities, and gene-essentiality information, generation of a genome-streamlined strain through iterative knockout is a daunting task that includes abundant possible permutations, which would be resource and time intensive.

To facilitate prioritization of gene-knockout targets for strain optimization, cell-free metabolic engineering (CFME) has been previously demonstrated (Liew et al. 2022). Building on this work, the team plans to further expand by identifying and removing protein effectors that have the greatest effects on  $CO_2$  to product metabolism. Additionally, using CFME, researchers can screen candidate effectors as combinations to prototype strains that have been iteratively reduced well before the engineered strain exists. This data will improve or validate gene ontology and develop genotype-phenotype linkages that inform *in silico* modeling of cellular processes.

While CFME has been demonstrated the ability to handle rapid, low-volume and high-throughput screens, genome engineering in the context of gas fermentation still relies on at least microliter volumes and generation times significantly longer than those of other model organisms. The team is able to leverage unique gas fermentation capabilities developed at LanzaTech, including the world's first and only automated biofoundry capable of anaerobic gas fermentation at high throughputs. This integrated system, developed with support from the Biological and Environmental Research (BER) Program's Genomic Science program DE-SC-0019090, enables genome engineering, colony picking, cultivation, screening, and strain repository at throughputs inaccessible to bench scientists. The large amount of generated data is managed in a custom-built laboratory information management system and provides the team with access to an ever-growing repository of host knockout genotypes for use as platforms for metabolic engineering.

In order to develop strains with increased genomic efficiency that are robust enough to tolerate industrial fermentation, the team, in partnership with the DOE Joint Genome Institute (JGI), has developed the largest dataset to date characterizing transcription factor binding sites (TFBS) across six *Clostridial* genomes. Here, the initial analysis of that dataset and some conserved TFBS motifs are demonstrated. Further work is planned to build and report on regulatory networks and to use that output to inform *in vitro* experiments. Validation through novel cellfree methods will aid in refining the knowledgebase through elucidating principles of regulation, promoter sequence 'leakiness,' dynamic range of the transcription factor in the presence of inducer, and ligand sensitivity.

Ultimately, the team will deliver novel *C. autoethanogenum* strains that have been streamlined to metabolize  $CO_2$  into products through continuous fermentation. These strains will have improved genomic efficiencies with increased functionally stability in the industrial bioproduction specific settings.

Brown, S. D., et al. 2014. "Comparison of Single-Molecule Sequencing and Hybrid Approaches for Finishing the Genome of *Clostridium autoethanogenum* and Analysis of CRISPR Systems in Industrial Relevant Clostridia," *Biotechnology for Biofuels and Bioproducts* 7, 40. DOI:10.1186/1754-6834-7-40.

Liew, F. E., et al. 2022. "Carbon-Negative Production of Acetone and Isopropanol by Gas Fermentation at Industrial Pilot Scale," *Nature Biotechnology* **40**, 335–44. DOI:10.1038/ s41587-021-01195-w.

Tschaplinski, T. J., et al. 2019. "Development of a Sustainable Green Chemistry Platform for Production of Acetone and Downstream Drop-In Fuel and Commodity Products Directly from Biomass Syngas via a Novel Energy Conserving Route in Engineered Acetogenic Bacteria," DOE EERE Technical Report. DOI:10.2172/1543199.

Woods, C., et al. 2022. "Required Gene Set for Autotrophic Growth of *Clostridium autoethanogenum*," *Applied and Environmental Microbiology* **88**, e02479-21. DOI:10.1073/pnas.2216244120.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant no. DE-SC0023278. DAP-seq work was supported by the Joint Genome Institute Community Science Program under award No. CSP-507323.

## Developing, Understanding, and Harnessing Modular Carbon/Nitrogen-Fixing Tripartite Microbial Consortia for Versatile Production of Biofuel and Platform Chemicals

Yanmeng Liu<sup>1</sup>, David Carruthers<sup>1</sup>, Yi Dai<sup>1</sup>, Josie McQuillan<sup>2</sup>, Maciek Antoniewicz<sup>1</sup>, Sujit Datta<sup>3</sup>, Jagroop Pandhal<sup>2</sup>, Andrew Allman<sup>1</sup>, **Xiaoxia (Nina) Lin**<sup>1\*</sup> (ninalin@umich.edu)

<sup>1</sup>University of Michigan; <sup>2</sup>University of Sheffield; <sup>3</sup>Princeton University

**Goals:** The overall goal of this project is to design, construct, analyze and optimize a synthetic microbial consortium system consisting of three closely interacting members: a CO<sub>2</sub>-fixing photosynthetic specialist, a nitrogen-fixing specialist, and a third specialist that can convert organic carbon and nitrogen generated by the first two specialists to synthesize a desired product. By integrating complimentary expertise from multiple research laboratories at three institutions, researchers are pursuing three specific objectives: (1) develop tripartite microbial consortia for carbon/nitrogen fixation and production of biomolecules with various nitrogen/carbon ratios;(2) investigate molecular and cellular mechanisms governing the tripartite consortia via omics study and predictive modeling; and (3) explore alternative spatial configurations and develop scalable design principles.

Microbial communities are ubiquitous in nature, exhibiting incredibly versatile metabolic capabilities and remarkable robustness. Inspired by these synergistic microbial ecosystems, rationally designed synthetic microbial consortia is emerging as a new paradigm for bioprocessing and offers tremendous potential for solving some of the biggest challenges society faces. In this project, the team focuses on a tripartite consortium consisting of a CO<sub>2</sub>-fixing photosynthetic specialist, a nitrogen-fixing specialist, and a third specialist that can convert organic carbon and nitrogen generated by the first two specialists to synthesize a desired product. In addition to CO<sub>2</sub> fixation, a noteworthy feature of this design is the elimination of the requirement for nitrogen fertilizer, which has been produced through ammonia synthesis using the Haber-Bosch process and accounts for an estimated 2% of global energy expenditure. Researchers aim to develop a modular and flexible model system capable of producing diverse biomolecules (varying C:N ratio) as advanced biofuel or platform chemicals to dissect this complex ecosystem using a spectrum of cutting-edge systems approaches and to ultimately derive scalable and broadly applicable design principles for maximizing the system performance.

The first prototype tripartite consortium employs genetically modified strains of photosynthetic cyanobacterium *Synechococcus elongatus* that secretes sucrose (Abramson et al. 2016) and nitrogen-fixing bacterium *Azotobacter*  *vinelandii,* which secretes ammonia (Barney et al. 2015) to form a symbiotic foundation for supporting a third producer member. Researchers demonstrated supported growth for a range of producer-strain candidates, including a sucrose-metabolizing *Escherichia coli* K-12 derivative strain, *Corynebacterium glutamicum,* and *Bacillus subtilis,* using a multichamber bioreactor system under continuous culture conditions (Carruthers et al. in preparation).

Ongoing investigations include: (1) development of predictive mathematical models of the triculture system to systematically explore the parameter space to understand how different biological parameters and operating strategies impact the system performance such as yield and productivity, and (2) omics analysis of monocultures and cocultures under controlled yet suboptimal/stressed conditions to identify the molecular bottlenecks limiting the performance of each member and hence the overall triculture.

Abramson, B. W., et al. 2016. "Increased Photochemical Efficiency in Cyanobacteria via an Engineered Sucrose," *Plant and Cell Physiology*, **57**(12), 2451–60.

Barney, B. M., et al. 2015. "Gene Deletions Resulting in Increased Nitrogen Release by *Azotobacter vinelandii*: Application of a Novel Nitrogen Biosensor," *Applied and Environmental Microbiol*ogy **81**(13), 4316–28.

Carruthers, D., et al. In preparation. "Engineering Modular Carbonand Nitrogen-Fixing Microbial Consortia for Sustainable Biochemical Production."

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0022136.

## A Gapless and Phased Diploid Genome Assembly for *Auxenochlorella protothecoides* Facilitates Metabolic Modeling and Proteomics Analyses

Rory J. Craig<sup>1</sup>, Dimitrios J. Camacho<sup>1</sup>, Sean D. Gallaher<sup>1</sup>, Carrie D. Nicora<sup>3</sup>, Samuel O. Purvine<sup>3</sup>, Jacob Tamburro<sup>2</sup>, Mark Vigliotti<sup>2</sup>, Jeffrey L. Moseley<sup>1</sup>, Mary S. Lipton<sup>3</sup>, Nanette R. Boyle<sup>2\*</sup> (nboyle@mines.edu), **Sabeeha S. Merchant**<sup>1</sup>

<sup>1</sup>University of California–Berkeley; <sup>2</sup>Colorado School of Mines; <sup>3</sup>Pacific Northwest National Laboratory

**Goals**: *Auxenochlorella protothecoides*, a Trebouxiophyte oleaginous alga, is a reference for discovery and a platform for photosynthesis-driven synthetic biology and sustainable bioproduction. Researchers will expand transformation markers, regulatory sequences; and reporter genes; improve transformation efficiency; and develop RNPmediated gene-editing methods for genome modification. Systems analyses and metabolic modeling approaches will inform genome modifications for rational improvement of photosynthetic carbon fixation and strain engineering to produce cyclopropane fatty acids. Regulatory factors and signaling pathways responsible for activating fatty acid and triacylglycerol biosynthesis will be identified, and researchers will manipulate them to increase lipid productivity. Non-photochemical quenching and a regulatory circuit for maintaining photosynthesis under Cu-limitation, both of which are absent in *A. protothecoides*, will be introduced to improve photosynthetic resilience. The performance of engineered strains will be modeled.

High-quality reference genomes and structural annotations are the foundation of many systems and synthetic biology approaches. Researchers have produced a gapless and phased genome assembly for the diploid A. protothecoides strain UTEX 250 using high-coverage Pacific Biosciences (PacBio) HiFi sequencing and Omni-C linked read sequencing. The haploid length of the UTEX 250 nuclear genome is 22 Mb, which is arranged on 12 chromosomes ranging from 0.5 to 4.1 Mb. The genome is GC-rich (64%) and generally highly heterozygous; the two haplotypes differ at approximately 3% of sites, enabling allele-specific transformation and allele-specific gene expression to be quantified. However, approximately a third of the genome is homozygous, including three entire chromosomes, suggesting widespread loss-of-heterozygosity events as observed in other vegetative diploids (e.g. yeasts). Complete circular plastome (84.6 kb) and mitogenome (54.0 kb) assemblies have also been produced.

To produce highly accurate structural annotations, researchers have sequenced PacBio Iso-Seq libraries from mixotrophic and heterotrophic conditions and approximately 60 million paired-end and stranded RNA-seq reads from several other growth conditions. Utilizing these data, researchers have annotated approximately 7,600 gene models per haploid genome, approximately 70% of which are supported by full-length Iso-Seq reads. More than 200 complex gene models were corrected by manual annotation. The quality of the annotations are supported by a Benchmarking Universal Single-Copy Orthologs (BUSCO) score of approximately 99% completeness, with all missing BUSCOs manually confirmed to be biologically absent in the genome. The A. protothecoides genome is remarkably compact with respect to gene content and features less than 5% repeats, although a handful of potentially active DNA transposons have been identified.

The *A. protothecoides* haploid gene number is less than half as many as the reference green alga *Chlamydomonas reinhardtii*. Researchers are presently performing comparative analyses among several high-quality algal genomes to functionally characterize gene presence and absence with particular focus on the GreenCut of proteins that are typically found in photosynthetic plants and green algae. Researchers are using the structural annotations to improve synthetic biology approaches, including the identification of potential condition-specific promoters and the refinement of Kozak sequence and codon bias optimization. *A. protothecoides* is also unusual among green algae in that several geographically and environmentally diverse isolates are available in culture. Researchers are targeting an approximately 20-strain pangenome to identify the standing genetic variation that may facilitate strain improvement (e.g., growth in brackish water).

The new genome sequence will be used to update and improve the first draft metabolic network of A. protothecoides that the team published previously. In order to model growth in a variety of conditions, a complete biomass analysis will be performed for growth in heterotrophic, autotrophic, and mixotrophic conditions. The experimentally obtained data will be used to develop accurate biomassobjective equations for each growth condition. Constraints based on metabolic-modeling approaches, such as flux balance analysis (FBA), will be used to predict growth and yields in different environmental and genetic backgrounds. These simulations will also be used to identify gene targets to further improve production of cyclopropane fatty acids. Isotope-assisted metabolic flux analysis (13C-MFA) will also be used to quantify fluxes in different growth conditions and mutant strains.

Assessment and quantification of protein production is a key step in evaluating engineering outcomes. Researchers have extended the pipelines to A. protothecoides using state-of-the-art technologies resident at Pacific Northwest National Laboratory and in a single pilot experiment the team captured more than 6100 A. protothecoides proteins, representing around 80% of the ~7600 proteins encoded in the genome. More detailed analyses are now possible due to the completion of the genome. For applications where it is critical to know the exact abundance of proteins, researchers will employ targeted proteomics with selected reaction monitoring (SRM) to determine the absolute amount of a subset of proteins. For instance, metabolic models of flux from carbon-fixation to triacylglycerol biosynthesis will be more accurate when they can incorporate the concentration of active sites for key pathway enzymes. Transgenes from engineering efforts, orthologs of fatty acid and lipid biosynthesis enzymes, CBC enzymes, PSI, PSII, and lightharvesting proteins in Auxenochlorella are candidate targets for the SRM approach.

**Funding Information:** This work was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0023027.

#### Deploying Top-Down and Bottom-Up Strategies for Genetic Engineering of *Auxenochlorella protothecoides* for the Production of Sustainable Biofuels

Derrick Chuang<sup>1</sup>, Shivani Upadhyaya<sup>1</sup>, Krishna K. Niyogi<sup>1,2</sup>, Setsuko Wakao<sup>2</sup>, Melissa S. Roth<sup>1\*</sup> (Mroth@berkeley.edu), **Sabeeha S. Merchant**<sup>1</sup>

<sup>1</sup>University of California–Berkeley; <sup>2</sup>Lawrence Berkeley National Laboratory

Goals: Auxenochlorella protothecoides, a Trebouxiophyte oleaginous alga, is a reference organism for discovery and a platform for synthetic biology driven by photosynthesis. Researchers will expand transformation markers, regulatory sequences, and reporter genes; improve transformation efficiency; and develop RNP-mediated gene-editing methods for genome modification. Systems analyses and metabolic modeling approaches will inform genome modifications for rational improvement of photosynthetic carbon fixation and strain engineering to produce cyclopropane fatty acids. The team will identify regulatory factors and signaling pathways responsible for activating fatty acid and triacylglycerol biosynthesis and will manipulate them to increase lipid productivity. Nonphotochemical quenching and a regulatory circuit for maintaining photosynthesis under copper limitation, both of which are absent in A. protothecoides, will be introduced to improve photosynthetic resilience. The performance of engineered strains will be modeled. Here, researchers focus on two genetic engineering strategies, improving core photosynthesis and upregulating de novo fatty acid biosynthesis.

Microalgae can play an integral role in a sustainable bioeconomy by helping to meet the rising demand for energy and products. Microalgae use solar energy to capture and convert CO<sub>2</sub> into biomass and achieve rapid growth without competing with food crops for land and water. However, there are considerable practical limitations in the photosynthetic production of biofuels from microalgae, resulting in low productivity and high costs. A. protothecoides has a highly flexible metabolism with rapid growth reaching high density under photoautotrophic and heterotrophic conditions, the latter in which cells accumulate large amounts of triacylglycerol (TAG). However, as with all photosynthetic organisms, one of the challenges that ultimately limits the theoretical yield is the low efficiency of photosynthetic energy conversion. Photosynthesis works efficiently when light is limiting. Under moderate to high light intensities, carbon fixation becomes limiting, and photoprotection mechanisms known as non-photochemical quenching (NPQ) are induced to minimize the generation of reactive

oxygen species and photoinhibition. To enhance photosynthesis efficiency, the team will increase sedoheptulosebisphosphatase activity to relieve the rate-limiting step of the Calvin-Benson (CB) cycle, thereby enhancing carbon fixation and decreasing the dissipation of light energy through the induction of NPQ.

Regions of non-coding DNA, cis-regulatory elements, play critical roles in transcriptional regulation in algae and plants and can be used in genetic-engineering approaches to increase the expression of fatty acid biosynthesis (FAS) genes. Here, researchers established a pipeline to extract transcription factors based on Interproscan IDs from the available A. protothecoides 0710 genome. Using available transcriptomic data, the team identified novel putative transcription factors involved in regulation of de novo FAS and glycolysis, which produce the precursor for FAS, pyruvate. The team then tested candidate transcription factors using time-resolved qPCR analysis to confirm their upregulation during TAG accumulation. Researchers are also using systems-biology tools, such as cis-regulatory element discovery using MEME suite (Bailey et al. 2015) to identify key regulatory motifs potentially involved in regulation of FAS. Simultaneously, researchers will (1) generate and integrate transcriptomic and proteomic data to create gene regulatory networks, (2) identify hub transcription factors, and (3) use DAP-Seq (Bartlett et al. 2017) to find and validate binding targets in promoters, untranslated regions (UTR), and introns. Altogether, this work will inform the engineering of strains to improve total lipid accumulation in *A. protothecoides*.

Strains with engineered SBPase will be characterized for their growth, biomass, and photosynthetic capacities to test whether they have enhanced carbon fixation. Metabolomics analysis of polar metabolites will inform researchers on how the flux through the CB cycle has changed and offer new strategies for strain improvement. Similarly, strains engineered with the transcription factors will be analyzed for lipid accumulation. The metabolome data will be fed to metabolic flux modeling to iteratively improve the engineering strategies through the design-build-test-learn cycle. While characterization of each of these strategies will be interesting and important to understand the regulation of each biochemical pathway, combining both in a single strain will likely be necessary to optimize lipid production.

Bailey, T. L., et al. 2015. "The MEME Suite," *Nucleic Acids Research* **43**(W1), W39–49.

Bartlett, A., et al. 2017. "Mapping Genome-Wide Transcription-Factor Binding Sites Using DAP-Seq," *Nature Protocols* **12**(8), 1659–72.

**Funding Information:** 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 No. DE-SC0023027.

#### Tools and Targets: Engineering Modified Fatty Acids and Improved Photosynthetic Resilience in *Auxenochlorella protothecoides*

Jeffrey L. Moseley<sup>1\*</sup> (jlmoseley@berkeley.edu), Dimitrios J. Camacho<sup>1</sup>, Rory J. Craig<sup>1</sup>, Marco A. Duenas<sup>1</sup>, Sean D. Gallaher<sup>1</sup>, Suzanne M. Kosina<sup>2</sup>, Kyle J. Lauersen<sup>3</sup>, Yang-Tsung Lin<sup>1</sup>, Radhika Mehta<sup>1</sup>, Crysten E. Blaby-Haas<sup>2</sup>, Trent R. Northen<sup>2</sup>, **Sabeeha S. Merchant**<sup>1</sup>

<sup>1</sup>University of California–Berkeley; <sup>2</sup>Lawrence Berkeley National Laboratory; <sup>3</sup>King Abdullah University of Science and Technology

**Goals:** *Auxenochlorella protothecoides*, a Trebouxiophyte oleaginous alga, is a reference for discovery and a platform for photosynthesis-driven synthetic biology and sustainable bioproduction. Researchers will expand transformation markers, regulatory sequences, and reporter genes; improve transformation efficiency; and develop RNP-mediated gene-editing methods for genome modification. Systems analyses and metabolic modeling approaches will inform genome modifications for rational improvement of photosynthetic carbon fixation and strain engineering to produce cyclopropane fatty acids. Regulatory factors and signaling pathways responsible for activating fatty acid and triacylglycerol biosynthesis will be identified, and researchers will manipulate them to increase lipid productivity. Nonphotochemical quenching and a regulatory circuit for maintaining photosynthesis under copper (Cu)-limitation, both of which are absent in A. protothecoides, will be introduced to improve photosynthetic resilience. The performance of engineered strains will be modeled.

Researchers have established transformation and targeted gene-replacement by homologous recombination in A. protothecoides and now seek to engineer strains with improved photosynthesis that produce modified fatty acids. As an essential preliminary step, researchers have resolved the organellar genomes and generated a gapless, ~22 Mbp (haploid size), phased diploid nuclear genome of strain UTEX 250. Iso-seq and RNA-seq analyses facilitated the annotation of more than 7500 gene models on 12 chromosomes. Elemental analysis by Inductively Coupled Plasma-Mass Spectrometry was used to optimize the defined growth medium, enabling luxury consumption of essential nutrients. Subsequent systems analyses will be carried out to identify the full complement of metalloproteins encoded in the genome and to understand how metal nutrients are allocated. Two biosafe transformation markers, Arabidopsis THIC, restoring thiamine prototrophy, and Saccharomyces SUC2, enabling sucrose assimilation, are in routine use, along with the *nptII* gene, conferring resistance to the aminoglycoside antibiotic, G418. Additional herbicide resistance transformation markers are under development, along with counter-selectable recyclable marker genes that will allow for scarless and marker-free integrations. Strong promoters from genes encoding an ammonium transporter, stearoyl-ACP desaturase, methionine synthase, RuBisCO small subunit, photosystem I, and light-harvesting complex components have been demonstrated to activate expression of codon-optimized transgenes. Chlamydomonas BKT1, encoding beta-carotene ketolase, Gaussia princeps luciferase, Venus, and mCherry have been used as quantifiable reporter genes for evaluating promoter strength, and plastid targeting of the fluorescent proteins was demonstrated. Researchers have also investigated the optimal arrangement of ORFs in polycistronic constructs. The results are consistent with more balanced translation when the shorter of the two ORFs is upstream. This is compatible with the finding that ORFs in endogenous polycistronic genes across the green lineage usually have the shorter ORF upstream.

One stated goal is to modify Auxenochlorella fatty acid and lipid biosynthesis to produce medium (mid)-chain length cyclopropane fatty acids (CPFAs) suitable as precursors for jet fuel. This will require chain-length control to increase mid-chain fatty acids, control of saturation level [since cyclopropane fatty acid synthases (CPS) compete with endogenous microsomal desaturases for mono-unsaturated substrates], and increased exchange between phospholipids (the site of CPFA synthesis) and Kennedy pathway intermediates in TAG biosynthesis (so that CPFAs are incorporated into storage lipids). Initial experiments established that accumulation of C12:0 and C14:0 fatty acids increased in strains expressing the Cuphea wrightii FATB2 thioesterase gene, driven by a SAD2 promoter that is activated during nitrogen-starvation and lipid production. C16:0 levels were increased by expressing FATB3 from Brassica juncea. Researchers will test whether mid-chain levels can be enhanced further by co-expressing CwFATB2 with a beta-ketoacyl-ACP synthase gene (CwKASA1) from the same species. To generate CPFAs, the team intend to first make Auxenochlorella strains with reduced  $\Delta^{12}$ -desaturase activity, thus removing the most significant competing activity for CPS. These strains will simultaneously overexpress phosphatidylcholine: diacylglycerol cholinephosphotransferase (PDCT), lysophosphatidylcholine acyltransferase (LPCAT), and lysophosphatidic acid acyltransferase (LPAAT) from Sterculia foetida and Litchi chinensis, which are both plant species that accumulate high amounts of CPFAs in their seed oils. The team will then screen CPS from E. coli and CPFA-accumulating plants or marine bacteria to identify the most active enzymes in Auxenochlorella.

Researchers will also introduce a regulatory circuit into Auxenochlorella to maintain photosynthetic resilience in response to Cu-deficiency. The major sink for Cu in plants is the thylakoid lumen protein plastocyanin, which transfers electrons from the cyt  $b_6/f$  complex to oxidized photosystem I. Consequently, Cu limitation, which is common in many environments, severely reduces plant photosynthesis and growth. Some algae and cyanobacteria acclimate to Cu deficiency by substituting a heme-containing cytochrome that performs the same electron carrier function as plastocyanin, thereby maintaining high rates of photosynthetic electron transfer and reducing their cellular Cu quota. In Chlamydomonas the CYC6 gene, encoding Cyt c<sub>6</sub>, is regulated by a Cu-sensing transcription factor CRR1. A potential CRR1 homolog is identified in the Auxenochlorella genome, and preliminary RNA-seq analysis indicates that a putative copper transporter gene, CTR1, is activated during Cu starvation. The team will use the CTR1 promoter to activate expression of a synthetic, codon-optimized CYC6 gene in Cu-deficient Auxenochlorella cells and test for improved photosynthetic performance and growth. Efficient import of heterologous Cyt c6 into the Auxenochlorella thylakoid lumen may require replacement of the native bipartite transit peptide with the transit peptide of an endogenous lumen-targeted protein, such as Auxenochlorella plastocyanin. This work will establish A. protothecoides as a powerful photosynthetically driven cell chassis for sustainable bioproduction of fuels and specialty products.

**Funding Information:** This work was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0023027.

#### Early Career

#### Systems Metabolic Engineering of *Novosphingobium aromaticivorans* for Lignin Valorization

Marco N. Allemann, Christopher C. Azubuike, Hannah R. Valentino, Gerald N. Presley, Leah Burdick, Richard J. Giannone, **Joshua K. Michener** 

#### Oak Ridge National Laboratory

**Goals:** To engineer a nonmodel bacterium, *Novosphingobium aromaticivorans,* for valorization of depolymerized lignin to value-added bioproducts. The project involves (1) discovery and optimization of pathways for assimilation of lignin-derived aromatic compounds, (2) engineering conversion pathways that match the stoichiometry of aromatic catabolism, and (3) development of genome-scale mapping techniques to identify new engineering targets in nonmodel bacteria. Lignin is one of the abundant renewable materials found in nature. This heterogeneous aromatic polymer is composed of a variety of *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) monomers that are connected by diverse chemical linkages. Lignin valorization would improve biofuel economics, potentially through bacterial conversion of thermochemically depolymerized lignin into valuable bioproducts. *Novosphingobium aromaticivorans* F199 is an Alphaproteobacterium capable of degrading G, S, and H monomers and, due to its genetic tractability and broad catabolic capabilities, is an emerging model organism for conversion of lignin-derived aromatic compounds. However, F199 cannot natively catabolize every component of depolymerized lignin, which limits conversion yields (Azubuike et al. 2022).

Researchers are identifying new aromatic degradation pathways to increase the catabolic potential of F199 using a combination of barcoded transposon insertion sequencing, proteomics, experimental evolution, and in vitro biochemistry. The team demonstrated this approach with the aromatic monomer syringate (Cecil et al. 2018), the  $\beta$ -1 linked dimer 1,2-diguaiacylpropane-1,3-diol (DGPD; Presley et al. 2021), and, more recently, the monomer guaiacol (Bleem et al. 2022). However, there are multiple lignin-derived aromatic compounds that F199 catabolizes poorly or not at all. Researchers have evolved F199 to rapidly and completely catabolize the common  $\beta$ -O-4 dimer guaiacylglycerol- $\beta$ guaiacyl ether (GGE). In the process, they identified an uncharacterized native catabolic pathway. Researchers have also isolated a Novosphingobium strain that can assimilate the  $\beta$ - $\beta$  linked dimer pinoresinol and have identified the relevant catabolic pathway. Current efforts include detailed characterization of the key pinoresinol pathway enzymes and transfer of the pathway into F199.

In addition to optimizing lignin assimilation, researchers are converting the resulting intermediates into value-added products, such as building blocks for bioderived polymers, and have demonstrated production from a model ligninderived aromatic compound, ferulate. To better understand the effect of host-genetic variation on pathway function, researchers are adapting a novel technique, bacterial quantitative trait locus (QTL) mapping, to F199. Researchers have demonstrated intraspecific recombination between strain of *N. aromaticivorans* and are currently studying and optimizing this process. By combining novel pathway discovery, heterologous expression, and optimization, researchers are engineering *N. aromaticivorans* F199 to efficiently valorize lignin-derived compounds.

Azubuike, C. C., et al. 2022. "Microbial Assimilation of Lignin-Derived Aromatic Compounds and Conversion to Value-Added," *Current Opinion in Microbiology* **65**, 64–72. Bleem, A., et al. 2022. "Discovery, Characterization, and Metabolic Engineering of Rieske Non-Heme Iron Monooxygenases for Guaiacol O-Demethylation," *Chem Catalysis* **2**, 1989–2011.

Cecil, J. H., et al. 2018. "Rapid, Parallel Identification of Catabolism Pathways of Lignin-Derived Aromatic Compounds in *Novosphingobium aromaticivorans,*" *Applied and Environmental Microbiology* **84**, AEM.01185-18.

Presley, G. N., et al. 2021. "Pathway Discovery and Engineering for Cleavage of a  $\beta$ -1 Lignin-Derived Biaryl Compound," *Metabolic Engineering* **65**, 1–10.

**Funding Information:** This work was supported by the U.S. Department of Energy Office of Science, Biological and Environmental Research (BER) Program's Early Career Research Program. Preliminary data was collected through the BioEnergy Science Center and Center for Bioenergy Innovation, Bioenergy Research Centers funded by BER.

# Engineering Novel Microbes for Upcycling Waste Plastic

Jinjin Diao\* (j.diao@wustl.edu), Yuxin Tian, Yifeng Hu, **Tae Seok Moon** 

Washington University-St. Louis

https://sites.wustl.edu/moonlab/

**Goals:** The goal is to develop a consolidated biological process to upcycle waste polyethylene terephthalate.

Polyethylene terephthalate (PET) represents 12% of global solid waste. PET chemical recycling has been an option to solve this global problem, but it suffers from its relatively high process cost and the extremely low price of virgin PET (~\$1/kg). One solution to address this issue is to upcycle waste PET rather than recycle it to generate the same PET typically with low quality. PET upcycling can be achieved by depolymerizing PET into terephthalic acid (TPA) and ethylene glycol (EG) and biologically converting these monomers into value-added products. However, there are only a handful of reports demonstrating microbes capable of growing on both TPA and EG generated from PET as sole carbon sources. To overcome this limitation, researchers have performed strain screening to discover a Rhodococcus strain RPET that can grow well on the alkaline hydrolysis products of PET as the sole carbon source without any purification step. Notably, this strain can grow on a mixture of TPA and EG at extremely high concentrations (up to 0.6M) and high osmolarity resulting from alkaline hydrolysis and pH neutralization. The resultant media supported RPET's growth without any purification and sterilization step except for their dilution. In addition, many synthetic biology tools, developed for a related species Rhodococcus opacus (DeLorenzo et al. 2018; DeLorenzo et al. 2021; Diao, Carr, and Moon 2022), were functional in RPET, facilitating its metabolic engineering. In this presentation, researchers discuss the effort to develop this novel chassis for waste PET valorization with PET conversion into carotenoids (up to \$7,500/kg) as a proof-of-concept demonstration (Moon et al. 2022). Specifically, researchers discuss lycopene production up to 1.3 mg/L from PET using this technology (Diao et al. 2023). Along with other efforts (Moon 2022), this technology can solve global plastic pollution issues and sustainable chemical production problems.

DeLorenzo, D. M., et al. 2018. "Molecular Toolkit for Gene Expression Control and Genome Modification in *Rhodococcus opacus* PD630," ACS Synthetic Biology 7(2), 727–38.

DeLorenzo, D. M., et al. 2021. "An Improved CRISPR Interference Tool to Engineer *Rhodococcus opacus*," ACS Synthetic Biology **10**(4), 786–98.

Diao, J., et al. 2023. "Upcycling of Poly(ethylene terephthalate) to Produce High-Value Bioproducts," *Cell Reports* **42**, 111908.

Diao, J., Carr and T. S. Moon 2022. "Deciphering the Metabolic and Regulatory Networks of Aromatic Catabolism using Synthetic Biology Tools," *Communications Biology* 5, 1109.

Moon, T. S., et al. 2022. PCT Patent Application.

Moon, T. S., 2022. "SynMADE: Synthetic Microbiota Across Diverse Ecosystems," *Trends in Biotechnology* **40**, 1405–14.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0022003.

## Broadening Aromatic Compound Degradation in *Acinetobacter baylyi* Using Synthetic Metabolic Pathways

Alyssa Baugh<sup>1\*</sup> (alyssabaugh@uga.edu), Melissa Tumen-Velasquez,<sup>2</sup> Ramesh K. Jha,<sup>3</sup> **Ellen L. Neidle**<sup>1</sup>

<sup>1</sup>University of Georgia; <sup>2</sup>Oak Ridge National Laboratory; <sup>3</sup>Los Alamos National Laboratory

**Goals:** This project expands the ability of the soil bacterium *Acinetobacter baylyi* ADP1 to degrade aromatic compounds. Applications range from lignin valorization to bioremediation of recalcitrant environmental pollutants. Biotechnology applications will benefit from new methods to design and evolve efficient pathways for the metabolism of specific compounds. Pathways targeting the degradation of aromatic compounds were constructed and evolved in ADP1, an ideal model organism for synthetic metabolic pathways. To enable consumption of target compounds, researchers are combining parts of characterized foreign pathways. For compounds without known metabolic routes, including pyrogallol and syringol, the synthetic pathways are improved via laboratory evolution, targeted gene amplification, and serial transfer, which enable growth-based selection. Novel metabolic functions can be demonstrated in ADP1 and exported to other microbes.

Lignin is a vastly underutilized, energy-rich renewable resource. Initial processing yields heterogenous mixtures of aromatic compounds that are naturally degraded slowly by microbial consortia. Industrial applications could benefit from combining the metabolic abilities of these bacteria into a single laboratory strain. Carbon funneled through central metabolism could then be used to synthesize a product. Researchers use A. baylyi ADP1, a genetically malleable bacterium in which chromosomal changes can be introduced by simply mixing cells and linear DNA fragments (Bedore et al. 2023). During adaptive evolution, genetic regions can be duplicated, facilitating the accumulation of beneficial mutations (Pardo et al. 2023). Together, these methods offer a generalizable system to create catabolic modules that can be mixed, matched, and applied to diverse aromatic compounds.

As the goal is the metabolism of compounds with unknown natural degradative routes, the team designed a pathway for one such compound: pyrogallol. Extradiol cleavage enzymes that use a structurally similar substrate, catechol, have been reported to cleave pyrogallol, albeit poorly. Bacteria have been discovered to use pyrogallol as a carbon source, but the natural catabolic pathway is unclear. As a first step in the synthetic pathway, researchers inserted genes from Pseudomonas putida into the A. baylyi chromosome. These genes encode extradiol catechol dioxygenases, enzymes that might be modified for pyrogallol cleavage. Two different enzymes, XylE and TodE, replaced a native catechol dioxygenase, which, unlike each foreign enzyme, catalyzes intradiol ring cleavage. Thus, the ring-cleavage products generated by the foreign enzymes are not normally encountered by A. baylyi. A second module is needed to route the non-native ringcleavage intermediates to central metabolism. This second module consists of a foreign catabolic pathway encoded by multiple pra genes. This foreign pra-gene pathway was previously expressed in A. baylyi, and its functionality was achieved during laboratory evolution. The pra-encoded enzymes can route the products of both catechol and pyrogallol cleavage to central metabolism.

There are several challenges to this approach. Enzymes often initially function suboptimally in new hosts. Therefore, functionality of XylE or TodE is not guaranteed, even for cleaving the natural substrate catechol. Additionally, in using catechol or pyrogallol as the carbon source, the catechol dioxygenase must work in conjunction with the *pra*-encoded enzymes. This pathway comes from a *Paeniba-cillus* species with a different natural substrate (protocate-chuate) and regulatory context. Growth requires balanced metabolism of many chemical intermediates, several of which are toxic. As a final hurdle, the function of a nonnative dioxygenase must be honed and altered to improve activity on pyrogallol rather than catechol.

Strategies to engineer pyrogallol degradation exploit growth-based laboratory evolution and methods to alter gene dosage. Genes encoding a catechol dioxygenase and the Pra pathway are first integrated in different positions in the *A. baylyi* chromosome. Independent gene amplification of either module can facilitate selection of functional strains (Tumen-Velasquez et al. 2018). Fluorescent biosensors can detect pathway intermediates to assist strain selection (Jha et al. 2015). After strains with novel capabilities are isolated, the contribution of mutations to the resulting phenotype identified via whole-genome sequencing—can be determined using transformation methods (Bedore et al. 2023).

In this project researchers evolve and select for *A. baylyi* strains with foreign extradiol catechol dioxygenases capable of cleaving aromatic growth substrates. In these strains, growth using catechol, anthranilate, and benzoate, each as a sole carbon source, required interdependent functions of a foreign catechol dioxygenase and *pra*-encoded enzymes. The current focus is altering substrate specificity of the dioxygenase to allow use of pyrogallol as the carbon source. Through strategic design of synthetic pathways, researchers are developing *A. baylyi* ADP1 for biotechnology applications. The unique genetic system of this strain allows researchers to rapidly engineer strains, interchanging different metabolic and regulatory modules.

Bedore, S.R., et al. 2023. "Natural Transformation as a Tool in *Acinetobacter baylyi*: Streamlined Engineering and Mutational Analysis," *Methods in Microbiology*. DOI:10.1016/bs.mim.2023.01.002.

Jha, R. K., et al. 2015. "Rosetta Comparative Modeling for Library Design: Engineering Alternative Inducer Specificity in a Transcription Factor," *Proteins: Structure, Function, and Bioinformatics* **83**(7), 1327–40.

Pardo, I., et al. 2023. "Natural Transformation as a Tool in *Acineto-bacter baylyi*: Evolution by Amplification of Gene Copy Number," *Methods in Microbiology*. DOI:10.1016/bs.mim.2023.01.001.

Tumen-Velasquez, M, et al. 2018. "Accelerating Pathway Evolution by Increasing the Gene Dosage of Chromosomal Segments," *The Proceedings of the National Academy of Sciences* **115**(27), 7105–10.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0022220. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). This program is supported by the U.S. Department of Energy Office of Science, Biological and Environmental Research (BER) Program under FWP LANLF32A. A.B. received support for some preliminary work from the DOE Office of Science Graduate Student Research (SCGSR) program.

#### Engineering Synthetic Anaerobic Consortia Inspired by the Rumen for Biomass Breakdown and Conversion

Elaina Blair<sup>1\*</sup>, Roy Kim<sup>2</sup>, Joel Howard<sup>2</sup>, Patrick Leggieri<sup>1</sup>, Christopher Lawson<sup>2</sup>, Scott Baker<sup>3</sup>, **Michelle O'Malley**<sup>1</sup>

<sup>1</sup>University of California–Santa Barbara, <sup>2</sup>University of Toronto, Canada; <sup>3</sup>Pacific Northwest National Laboratory

Goals: This project will leverage a synthetic rumen consortium composed of anaerobic fungi and chain-elongating bacteria to study which metabolites are shared and exchanged between microbes and identify strategies to bolster lignocellulose conversion to value-added products. The project's approach will develop high-throughput systems and synthetic biology approaches to realize stable synthetic consortia that route lignocellulosic carbon into short- and medium-chain fatty acids (SCFAs/MCFAs) rather than methane. Key research objectives are to (1) design and predict anaerobic fungal and bacterial consortia that efficiently convert lignocellulosic biomass into MCFAs, (2) understand how fermentation parameters and microbe-microbe interactions regulate and drive microbiome metabolic fluxes, and (3) use genomic editing to alter the fermentation byproducts of anaerobic fungi and bolster MCFA titers and yields.

Lignocellulose deconstruction and conversion in nature is driven by mixed microbial partnerships rather than the action of a single microbe. For example, microbes are particularly well optimized to recycle organic matter in anaerobic habitats, ranging from landfills to intestinal tracts, via interspecies hydrogen transfer and methane release. Compared to aerobic processes, anaerobic digestion can be far more efficient in converting substrate to chemical products, largely because far less carbon is funneled to cell growth resulting in higher yields, and far less energy inputs are required because pretreatment, aeration, mixing, and heat removal are greatly reduced. Compartmentalizing difficult biomass deconstruction and production steps among specialist anaerobes is an exciting new route to convert biomass into value-added products, especially if consortia can be built predictively and engineered for stability.

Previously, the team established a model bacterial consortium enriched from the rumen that converts lignocellulosic biomass into high titers of butyrate C4 volatile fatty acids (VFAs) based on a chain-elongation process that inhibits archaeal methanogenesis. Metagenomic and metatranscriptomic analysis identified key chain-elongating bacteria in these consortia that maintain high expression of the reverse β-oxidation pathway responsible for C4–C8 VFA production. This analysis also revealed several other bacterial species in the consortia that compete with chain elongators and reduce overall C4-C8 VFA yields by diverting carbon to unwanted products. Therefore, building synthetic consortia that eliminate these competing bacteria would bolster product yields and enable great control over VFA chain length. In parallel, the team also demonstrated that anaerobic rumen fungi within the Neocallimastix genus are superior biomass degraders compared to anaerobic bacteria from these enrichments. Moreover, the biomass degradation products lactate, acetate, and ethanol from Neocallimastix are optimal substrates for chain elongators.

Accordingly, partnering anaerobic fungi and chain-elongating bacteria in synthetic consortia represents a novel strategy for maximizing lignocellulose conversion to C4–C8 VFAs (see Fig. 14, p. 92).

Recently, the team screened multiple chain-elongating bacteria and identified candidates that grow robustly in culture media with known fungal metabolites and produce VFAs.



Two of these strains are Megasphaera elsdenii and Pseudoramibacter alactolyticus. The team paired Neocallimastix spp. with each of these strains and cultivated them for numerous passages on reed canary grass. These cocultures show promise in terms of stability because both members were present after multiple transfers. Lactate is the key metabolic intermediate in these consortia, where the fungi make lactate as they degrade grass and the chain elongators produce VFAs from lactate. High-performance liquid chromatography measurements showed that lactate from fungal cultures was consumed after chain elongators were added and butyrate was produced. Future plans include optimizing inoculation ratios of the different strains in consortia to maximize stability and VFA production. Researchers will employ quantitative PCR to evaluate abundances of the community members over time and thus elucidate insight into the culture's stability. Here, researchers further describe efforts to systematically characterize and model the production of VFAs from synthetic bacterial and fungal communities that have been grown on representative lignocellulosic grass substrates.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0022142.

## The MoonTag Programmable Transcription Activator and Synthetic Promoters: Tools for Engineering Oil Biosynthesis in *Camelina* and Pennycress

J. Armando Casas Mollano\* (jcasasmo@umn.edu), Jonathan Cors, **Michael J. Smanski** 

University of Minnesota–Saint Paul

https://fatplants.net

**Goals:** Leverage the use of sequence-programmable transcriptional activators (PTAs) and synthetic promoters to fine tune the expression of multiple endogenous genes and/ or transgenes to facilitate targeted lipid production.

PTAs are synthetic tools that enable tunable regulation of the expression of endogenous genes and transgenes in eukaryotic and prokaryotic organisms. PTAs are composed of activation domains fused to a dead Cas9 enzyme (dCas9) that lacks nuclease activity but can still bind DNA in a sequence-specific manner (Casas-Mollano et al. 2020). The main advantages of Cas9-based systems are that they can achieve high levels of gene activation and are very easy to program via base-pairing between the guide RNA (gRNA) and the DNA target strand (Casas-Mollano et al. 2020). The PTA described here, called MoonTag, is a second-generation system that uses a nanobody-antigen peptide interaction to recruit multiple copies of an activation domain to its target promoters (Casas-Mollano et al. 2023). MoonTag is capable of inducing high levels of transcription in reporter as well as in endogenous genes in the monocot model plant *Setaria*. MoonTag is also able to efficiently activate genes in eudicot species such as *Arabidopsis* and tomato (Casas-Mollano et al. 2023). In addition to its activation capabilities, MoonTag components are expressed in transgenic plants to high levels without any deleterious effects resulting from the expressed components being incompatible with plant cells. Thus, MoonTag is a new activator that could be used to regulate the transcription of endogenous genes in many plant species.

The team has also created a set of synthetic promoters that could be used together with MoonTag to coordinate the expression of multiple genes. The synthetic promoters were designed using a modular approach. They comprise a minimal promoter (the region immediately upstream the transcriptional initiation site where the transcription pre-initiation complex and RNA pol II binds) and a transactivation (TA) region upstream containing binding sequences for transcription factors that stimulate or repress transcription (Belcher et al. 2020). The TA region was designed to contain six gRNA binding sites for a Cas9-based PTA such as MoonTag. The designed synthetic promoters can be strongly activated by MoonTag and other Cas9-based activators. One advantage of these synthetic promoters is that they share a minimal amount of duplicated sequences, limited to the gRNA binding sites, allowing for the design of multigene vectors with limited sequence homology but with robust expression levels.

As proof-of-concept, researchers designed synthetic promoters driving the three genes from the betalain biosynthetic pathway (He et al. 2020). In the presence of the MoonTag PTA, these genes were simultaneously activated leading to the accumulation of betalains. Researchers also demonstrate tissue-specific regulation of the synthetic promoters by expressing MoonTag from a seed-specific promoter that led to the accumulation of betalains exclusively in the seeds. Thus, the use of synthetic promoters together with PTAs should allow for the deployment of multigene constructs such as those for metabolic pathways or trait stacking that can be directed to specifically be expressed in a particular tissue, organ, developmental time, or in response to external cues.

In the context of the B5 (Bigger Better *Brassicaceae* Biofuels and Bioproducts, p. 74) project the team will use the MoonTag activator and the synthetic promoters to achieve the precise coordination of gene expression of transgenes as well as endogenous genes in order to facilitate oilseed engineering in pennycress and Camelina. Overexpression of endogenous genes in seeds will be achieved by expressing

MoonTag driven by seed-specific promoters together with gRNAs targeting the promoters of the genes of interest. When necessary, transgenes driven by synthetic promoters activated by MoonTag will be used so that both transgenes and endogenous genes can be coordinately regulated in a particular seed developmental stage by where MoonTag is expressed. For overexpression, researchers will initially target the *FatB* gene to boost palmitic acid (16:0) content and WRI1, DGAT1, GDP1 to increase oil content. Oil seed engineering will also require repression of lipid biosynthetic genes. To target genes for repression, team members will generate MoonTag PTRs (programmable transcriptional repressors) bearing the repressive SRDX domain (Hiratsu et al. 2003). Alternatively, researchers will use the synthetic promoters to drive expression of double-stranded RNAs that target endogenous genes for downregulation through RNA interference.

Belcher, M. S., et al. 2020. "Design of Orthogonal Regulatory Systems for Modulating Gene Expression in Plants," *Nature Chemical Biology* **16**, 857–65.

Casas-Mollano, J. A., et al. 2020. "CRISPR-Cas Activators for Engineering Gene Expression in Higher Eukaryotes," *The CRISPR Journal* **3**, 350–64.

Casas Mollano, J. A., et al. 2023. "Efficient Gene Activation in Plants by the MoonTag Programmable Transcriptional Activator," *bioRxiv* 2023.02.15.528671.

He, Y., et al. 2020. "A Reporter for Noninvasively Monitoring Gene Expression and Plant Transformation," *Horticulture Research.* 7, 152.

Hiratsu, K., et al. 2003. "Dominant Repression of Target Genes by Chimeric Repressors that Include the EAR Motif, a Repression Domain, in *Arabidopsis*," *The Plant Journal* **34**, 733–39.

**Funding Information:** This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0023142.

#### Plastic Degradation by the Gut Microbiome of Yellow Mealworms

Lummy Maria Oliveira Monteiro<sup>1,2</sup>, Jyoti Singh<sup>1,2</sup>, Ross Klauer<sup>1,2\*</sup> (rklauer@udel.edu), Harrison Hall<sup>3</sup>, Sankarganesh Krishnamoorthy<sup>4</sup>, Aaron Wright<sup>3</sup>, Mark Blenner<sup>1,2</sup>, **Kevin Solomon**<sup>1,2</sup>

<sup>1</sup>Delaware Biotechnology Institute, <sup>2</sup>University of Delaware; <sup>3</sup>Baylor University; <sup>4</sup>Pacific Northwest National Laboratory

http://solomonlab.weebly.com

https://www.blennerlab.com

**Goals:** This project discovers and reconstructs the plastic degradation pathways distributed across the gut microbiome of yellow mealworms (larvae of *Tenebrio molitor*) to develop enhanced capabilities for biologically based polymer recycling.

Plastics, initially selected for their durability and environmental resiliency, pose a significant environmental challenge for modern economies. Polystyrene (PS), high- and lowdensity polyethylene (HDPE and LDPE), and polypropylene (PP) are produced at a rate of more than 228 million tonnes globally each year. However, none have robust infrastructures for mechanical or chemical recycling and ultimately become polluting waste streams. To address this need, the team will pursue biological strategies for plastics depolymerization. Researchers will focus on the microbiomes of insect larvae (colloquially called worms) as they degrade plastics more rapidly than microbial isolates and do not require clean plastics or pretreatment. In particular, the microbiome of yellow mealworms is unique in that its host does not appear to contribute to degradation of a wide range of plastics. While bacterial community members have been identified, the specific pathways responsible for biodegradation remain to be elucidated and the potential contributions of fungal members are unexamined. Additionally, emerging evidence suggests that nutrient supplementation enhances plastic metabolism up to 70% and gives rise to a gut community structure distinct from that without additional nutrients. However, it is unclear if nutrient supplementation induces microbes to participate in in the plastic degradation or if it supports an optimal community composition for function.

As a first step to address these gaps, researchers characterized the consumption rates of PS, LDPE, HDPE, and PP via *T. molitor* larvae in the presence and absence of co-fed oats as a nutritional supplement. The consumption rates of PS, LDPE, and HDPE were 20.4, 12, and 1.1 mg (100 larvae)<sup>-1</sup>d<sup>-1</sup>, respectively, in agreement with established studies. However, oat supplementation enhanced plastics consumption by ~160, 60, and 230%, respectively. These studies establish the use of oats as a potent supplement for enhancement of PS and LDPE consumption rates, up to double that obtained with established supplements, and validated HDPE consumption by *T. molitor*.

Worm-consumed plastics were chemically modified beyond simple mechanical degradation validating biological mechanisms for plastics depolymerization. Fourier transform infrared spectroscopy (FTIR) analysis of plastic extracted from the frass (excrement) of mealworms fed PS revealed incorporation of oxygen not found in untreated controls. Moreover, benzene ring cleavage was observed for treated PS samples. Similarly, FTIR spectra of extracted plastic from LDPE-fed mealworm frass revealed the incorporation of carbonyl and alcohol groups. Finally, gel permeation chromatography (GPC) of the ingested plastic from PS-fed T. molitor larvae confirmed a 40% decrease in polymer molecular weight while LDPE-fed worms were able to decrease the molecular weight by up to three orders of magnitude. Taken together, these results demonstrate that the plastics being ingested by the larvae are actively depolymerized and chemically modified.

Microbiome community analysis via 16s and ITS sequencing revealed a rich consortium of bacteria and fungi. The bacterial community was more diverse than the fungal community with observed taxa belonging to the bacterial phyla Firmicutes, Tenericutes, Proteobacteria, Actinobacteria, Spirochaetes, Bacteroidetes, and Fusobacteria, and fungal Ascomycota, Basidiomycota, and Mucoromycota. As expected, mealworm diet led to unique community structures adapted to degradation of the fed plastic substrate. However, oats cosupplementation frequently selected for taxa that were not observed in plastics-only or oats-only controls suggesting currently unrecognized interactions. Despite these unique community structures, microcosms of communities in planktonic culture selected for with LDPE, HDPE, PS, and PP diet were all able to grow on LDPE as a primarycarbon source. Finally, community analyses revealed many facultative and obligate anaerobic genera such as Spiroplasma associated with LDPE and PS degradation when supplemented with oats. Correspondingly, these communities were enriched with clusters of genes (COG) and protein family (pfam) for iron-dependent anaerobic oxidation enzymes and pathways, which may serve as novel oxygenindependent pathways for plastics depolymerization.

To determine microbes and active enzymes responsible for the degradation of PS, PE, and PP, researchers have developed a suite of photoreactive chemical probes that resemble oligomers of these polymers. These probes are fluorescently labeled, providing an avenue to selectively isolate microbes that take up these molecules via fluorescence-activated cell sorting and enable subsequent proteomic characterization of the proteins acting upon them. Team members have begun using these probes to characterize enzymes that bind them strongly in a series of microbial isolates and identify the taxonomy of cells capable of transporting these plastics in plastic-degrading worm gut microbiomes.

In summary, ongoing work has characterized plastic consumption rates in *T. molitor* microbiomes, revealing novel strategies to structure gut microbial populations for enhanced degradation. Plastics were noted to be metabolized and not only mechanically degraded by both bacterial and fungal communities that contribute to plastic degradation even independent of the host mealworm. Team members have also developed chemical probe analogs of common plastics to isolate plastic-binding microbes and proteins for study. Through these parallel efforts, researchers aim to generate systems-level insight into the metabolic pathways of plastic-degrading microbiomes and to develop consortia enriched in plastic degradation activity.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Genomic Science Program under Award No. DE-SC0022018. A portion of this research will be performed under the Facilities Integrating Collaborations for User Science (FICUS) initiative and use resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Both facilities are sponsored by the Office of Biological and Environmental Research and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL).

## Developing Anaerobic Fungal Tools for Efficient Upgrading of Lignocellulosic Feedstocks

Manish Pareek<sup>1\*</sup> (pareek@udel.edu), Radwa Hanafy<sup>1</sup>, Casey Hooker<sup>1,2</sup>, Ethan Hillman<sup>2</sup>, Javier Muñoz Briones<sup>2</sup>, **Kevin Solomon<sup>1, 2</sup>** 

<sup>1</sup>University of Delaware, <sup>2</sup>Purdue University

#### http://solomonlab.weebly.com

**Goals**: This project develops genetic and epigenetic tools for emerging model anaerobic fungi to identify the genomic determinants of their powerful biomass-degrading capabilities, facilitate their study, and enable direct fungal conversion of untreated lignocellulose to bioproducts.

Deconstruction of plant cell walls is a significant bottleneck to the economical production of affordable biofuels and bioproducts from abundant and renewable plant biomass. Anaerobic fungi (*Neocallimastigomycota*) from the digestive tracts of large herbivores have evolved unique abilities to degrade untreated fiber-rich plant biomass by combining hydrolytic strategies from the bacterial and fungal kingdoms (Haitjema et al. 2017). Anaerobic fungi secrete the largest known diversity of lignocellulolytic carbohydrate active enzymes (CAZymes) in the fungal kingdom, which unaided can degrade up to 60% of the ingested plant material within the animal digestive tract (Seppälä et al. 2017, Youssef et al. 2013). Unlike many other fungal systems, these CAZymes are tightly regulated and assembled in fungal cellulosomes to synergistically degrade plant material, including untreated agricultural residues, bioenergy crops, and woody biomass with comparable efficiency regardless of composition (Haitjema et al. 2017; Solomon et al. 2016; Solomon et al. 2018; Hooker et al. 2018).

The team's efforts to characterize gut fungal CAZymes reveal industrially relevant properties such as remarkable stability and activity towards untreated plant biomass (Hooker et al. 2018; Hillman et al. 2021). Gut fungal CAZymes liberate sugars from cellulosic substrates for over a week after inoculation with some sugar metabolized to organic acids. However, model bioproduction hosts such as Kluyveromyces marxianus can capture the carbon in these sugars and acids in a two-stage process to efficiently upgrade this carbon to high value solvents, fragrances and advanced fuels derived from esters and aromatic alcohols (e.g., ethyl-acetate and 2-phenylethanol; Hillman et al. 2021). Similarly, anaerobic fungal biosynthetic enzymes possess unique cofactor substrate preferences that support higher catalytic efficiencies, which are easily overlooked via heterologous expression due to the extremely high AT content (~83%) of gut fungal genomes and biased codon preferences (Hillman et al. 2021). Thus, there is an unmet need to build genetic tools and methods to study these enzymes natively in anaerobic fungi.

As a first step to tool development, the team sequenced the genomes of three novel anaerobic fungal isolates to enable part mining. High quality genomic DNA isolations were paired with PacBio long-read sequencing and Hi-C (chromosomal conformation capture) sequencing to achieve the first chromosomally resolved genomes for isolates belonging to the genera *Neocallimastix* and *Piromyces*. The team's assemblies incorporate more than 99% of the genome into 12 to 25 chromosomes with N50<10. Parts that regulate gene expression (*e.g.*, promotors and terminators) were then identified via sequence homology to assemble a nascent genetic toolbox for heterologous expression in anaerobic fungi.

To introduce these parts, the team optimized delivery of DNA with fluorescently labelled oligonucleotides and circular plasmids. Both *Piromyces* and *Neocallimastix* isolates were naturally competent for double-stranded DNA, taking up any DNA supplemented to the growth media and localizing it to the nucleus. Transformation was observed in multiple

life stages suggesting that natural competence is a robust property of anaerobic fungi and a facile method to introduce heterologous DNA. Using natural competency, team members then validated heterologous expression of anaerobic fluorescent reporters and codon-optimized antibiotic resistance markers expressed via identified constitutive promoters and terminators synthesized by the DOE Joint Genome Institute's Biological and Environmental Research Support Science (JGI-BERSS) program. Researchers also identified and validated nuclear localization sequences (NLS) from anaerobic fungal histone proteins, which displayed distinct sequence motifs from conventional NLSs used in model organisms. However, due to a lack of known autonomously replicating sequences, heterologous DNA must be supplemented to growth media daily to achieve stable phenotypes. Efforts are underway to achieve stable transformants via the use of Cas9 ribonucleoproteins (RNPs) together with splitmarker cassettes.

In conclusion, anaerobic fungi hold a wealth of potential for biocatalysis from renewable plant substrates. Leveraging natural competency to introduce exogenous DNA, researchers have achieved the first simple methods for targeted heterologous expression in anaerobic fungi. The team's growing toolbox for anaerobic fungi form foundational tools to generate a deeper systems-level understanding of anaerobic fungal physiology while establishing fundamental knowledge about regulation of gut fungal CAZymes. Ultimately, the team enables predictive biology in anaerobic fungi and derive insight into microbial plant deconstruction to advance the development of economical biofuels and bioproducts.

Haitjema, C. H., et al. 2017. "A Parts List for Fungal Cellulosomes Revealed by Comparative Genomics," *Nature Microbiology* **2**, 17087.

Hillman, E. T., et al. 2021. "Hydrolysis of Lignocellulose by Anaerobic Fungi Produces Free Sugars and Organic Acids for Two-Stage Fine Chemical Production with *Kluyveromyces marxianus*," *Biotechnology Progress* **37**(5), e3172.

Hooker, C., et al. 2018. "Hydrolysis of Untreated Lignocellulosic Feedstock is Independent of S-Lignin Composition in Newly Classified Anaerobic Fungal Isolate Piromyces sp. UH3-1," *Biotechnology for Biofuels* **11**:293.

Seppälä, Susanna et al. 2017. "The Importance of Sourcing Enzymes From Non-Conventional Fungi for Metabolic Engineering and Biomass Breakdown," *Metabolic Engineering* **44**, 45–59.

Solomon, K. V., et al. 2016. "Early-Branching Gut Fungi Possess a Large, Comprehensive Array of Biomass-Degrading Enzymes," *Science* **351**, 1192–1195.

Solomon, K. V., et al. 2018. "Catabolic Repression in Early-Diverging Anaerobic Fungi is Partially Mediated by Natural Antisense Transcripts," *Fungal Genetics & Biology* **121**, 1–9.

Youssef, N.H., et al. 2013. "The Genome of the Anaerobic Fungus *Orpinomyces* sp. Strain C1A Reveals the Unique Evolutionary History of a Remarkable Plant Biomass Degrader," *Applied and Environmental Microbiology* **79**(15), 4620–34. Funding Information: This project is supported by the Biological and Environmental Research Program through thetheDOE Office of Science under Award No. DE-SC0022206.nnA portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS)ddinitiative and used resources at the DOE Joint GenomeeInstitute and the Environmental Molecular Sciences Laboratory, which are DOE Office of Science User Facilities. Bothsfacilities are sponsored by the Biological and Environmen-p

tal Research Program and operated under Contract Nos. DE-AC02- 05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL).

#### Early Career

#### Encapsulin Nanocompartment Systems in *Rhodococcus opacus* for Compartmentalized Biosynthesis Applications

**Mimi Cho Yung**<sup>1\*</sup> (yung6@llnl.gov), Tiffany M. Halvorsen<sup>1</sup>, David F. Savage<sup>2</sup>, Timothy S. Carpenter<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory; <sup>2</sup>University of California–Berkeley

**Goals**: This project is focused on understanding how encapsulin nanocompartment systems can be used to enhance the biosynthesis of next-generation biomaterials in *Rhodococcus* species. The project seeks (1) to probe the mechanistic basis for how these compartments are regulated, biosynthesized, and maintained, and (2) to engineer these systems to achieve new biosynthetic functions (e.g., inorganic nanoparticle biosynthesis).

With recent innovations in synthetic biology, engineered microbes now have the potential to produce a wide variety of bioproducts from renewable sources (e.g., biomass) to support the U.S. bioeconomy. However, biosynthetic pathways leading to these products are often hindered by poor reaction efficiencies and toxicity, resulting in low yields.

Compartmentalization of these pathways could potentially overcome these challenges through colocalization, concentration, and sequestration. The goal of this Early Career research is to identify mechanisms for engineering compartmentalized biosynthesis in the emerging model bioproduction bacterium, *Rhodococcus opacus* PD630, using its native encapsulin nanocompartment system (herein called encapsulins). Toward this goal, the native regulation, biosynthesis, and maintenance of the encapsulin system will be investigated to uncover potential pathways for controlling encapsulin production. Gene-editing methods will be used to engineer encapsulins with novel structural properties for expanded bioproduction capabilities. As a case study, the *R. opacus* encapsulin system will be redirected to support and control the biosynthesis of cadmium sulfide (CdS) nanoparticle semiconducting materials used in optical and electronic applications (e.g., solar panels, light-emitting diodes). CdS nanoparticles have gained considerable interest as a bioproduction target because they can potentially be produced from upcycling cadmium-contaminated waste streams. Ultimately, this work will establish encapsulin compartmentalization systems as a means of improving yields and enabling new biosynthetic routes toward next generation bioproducts and biomaterials in support of DOE's mission to build a strong bioeconomy and thus enhance U.S. energy security.

**Funding Information**: Work at LLNL is performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-ABS-845214). This program is supported by the U.S. Department of Energy, Office of Science, through the Genomic Science Program, Biological and Environmental Research Program, under FWP SCW1770 (Early Career Award).

#### Microbial Guilds and Niches Enable Targeted Modifications of the Microbiome

Oriane Moyne\* (omoyne@ucsd.edu), Deepan Thiruppathy, Chloe Lieng, Manish Kumar, Eli Haddad, Grant Norton, **Karsten Zengler** 

#### University of California-San Diego

Goals: This project couples novel laboratory and field studies to develop the first predictive model of grassmicrobiomes based on new mechanistic insights into dynamic plant-microbe interactions in the grasses Sorghum bicolor and Brachypodium distachyon that improve plant nitrogen use efficiency (NUE). The results will be used to predict plant mutants and microbial amendments that improve low-input biomass production for validation in laboratory and field studies. To achieve this goal, the team will determine the mechanistic basis of dynamic exudate exchange in the grass rhizosphere with a specific focus on the identification of plant transporters and proteins that regulate root exudate composition and how specific exudates select for beneficial microbes that increase plant biomass and NUE. The team will further develop a predictive plant-microbe model for advancing sustainable bioenergy crops and will predictively shift plant-microbe interactions to enhance plant biomass production and nitrogen acquisition from varied nitrogen forms.

Microbiome science has contributed greatly to understanding of microbial life and provided insights on the essential roles of microbial communities, from global elements cycling to human health. However, a comprehensive understanding of how these communities are assembled, maintained, and function as a system is still lacking. In particular, the nature of microbe-microbe interactions and how microbial communities respond to perturbations remains poorly understood. As a result, current microbiome science is largely descriptive and correlation-based, rather than predictive and based on mechanistic understanding.

To achieve predictive microbiome science, it is necessary to comprehensively elucidate the metabolic role of each microbe and its interactions with others. Such knowledge would enable the manipulation of a microbe's trajectory within a community, for example by selectively promoting or limiting its growth.

In this study, the team presents a new method that integrates transcriptional and translational regulation measurements to reveal how each microbe allocates its resources for optimal proteome efficiency. Protein translation is the most energyintensive process in a cell, and microbes closely regulate their resource allocation by prioritizing essential functions through differential translational efficiency (TE). Direct measurement of TE in a microbial community sample would provide insights into the metabolic role of each member of the community and allow for a better understanding of interactions with other members.

The team performed metatranscriptomics and metatranslatomics analysis to directly measure TE in situ in a 16-member synthetic community (SynCom) composed of rhizosphere isolates grown in a complex culture medium. This approach allowed us to perform a guild-based microbiome classification, grouping microbes according to the metabolic pathways they prioritize independent of their taxonomic relationships. Team members demonstrated that guilds predicted competition between members of the same guild with 100% sensitivity and 74% specificity (77% accuracy) in the SynCom. Furthermore, gene-level analysis of TE allowed us to predict each microbe's substrate preferences, (i.e., their niche in the community). Such Microbial Niche Determination (MiND) predicted which particular microbes would benefit from substrate supplementation with 54% sensitivity and 83% specificity (78% accuracy) in the SynCom.

It is worth noting that as microbes adapt their translational regulation to community settings, such predictions could not be achieved using axenic culture approaches (e.g., phenotypic microarray or growth curves) or partially functional measurements (e.g., metagenomics or metatranscriptomics).

By combining TE-based MiND and guilds predictions, researchers were able to selectively manipulate the Syn-Com, by increasing or decreasing the abundance of targeted members, either by providing preferred substrates or by giving an advantage to their competitors. Importantly, this method is scalable to more complex, natural samples. Team members applied MiND to native soil samples and demonstrated its applicability in predicting changes and manipulating microorganisms in complex microbiomes.

In conclusion, the method presented in this study represents a significant step towards achieving predictive microbiome science by providing a comprehensive understanding of the metabolic role of each microbe and its interactions with others. The guild-based microbiome classification and MiND approach allows for the manipulation of microbial communities and has potential applications in various fields such as agriculture, biotechnology, and human health.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Awards DE- SC0021234 and DE-SC0022137. Furthermore, the development of the technologies described in this article were in part funded through Trial Ecosystem Advancement for Microbiome Science Program at Lawrence Berkeley National Laboratory funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research Awards DE-AC02- 05CH11231. The work was also supported by the UC San Diego Center for Microbiome Innovation (CMI) through a Grand Challenge Award.

#### Model-Guided Design of Synthetic Microbial Consortia for Next-Generation Biofuel Production

Deepan Thiruppathy<sup>1\*</sup> (dthirupp@eng.ucsd.edu), Neal N. Hengge<sup>2</sup>, Lizzette Moreno García<sup>2</sup>, Gustavo Lastiri-Pancardo<sup>1</sup>, Violeta Sànchez i Nogué<sup>2</sup>, Jeffrey G. Linger<sup>2</sup>, Michael T. Guarnieri<sup>2</sup>, **Karsten Zengler<sup>1</sup>** 

<sup>1</sup>University of California–San Diego; <sup>2</sup>National Renewable Energy Laboratory

**Goals:** The team propose to lay the foundation for bioproduction using multifaceted microbial communities. Researchers will build metabolic community models of increasing complexity by integrating multiomics datasets. These models will guide engineering designs for optimized production of biofuels from lignocellulosic biomass. Furthermore, researchers will use innovative approaches to augment existing communities for improved bioproduction and complete conversion of different biomass feedstock. Overall, these strategies will provide knowledge of the functional metabolic exchanges driving interspecies interactions in microbial communities, thus providing insights
into fundamental biological processes. Lessons learned here would be crucial for designing stable microbial communities for various biotechnology applications in the future.

Microbial communities are everywhere, and their influence on the environment is gaining recognition for their industrial potential, such as bioenergy production. The multiplicity of intertwined, interspecies metabolite interactions within these communities regulates their ultimate functional organization and assembly. This allows them to perform complex functional tasks unreachable by axenic systems such as the breakdown of hardy lignocellulosic materials into high-energy volatile fatty acids (VFAs).

Bioproduction of one such fatty acid, butyrate (BA), from sustainable lignocellulosic sources has gained attention owing to butyrate's versatile applications as a precursor for jet-fuel, polymers, fibers, and even cosmetics. However, current industrial processes have been forced to rely on monoculture setups requiring expensive enzymatic raw-material preprocessing. Thus, there is a need to rationally design reproducible tunable consortia that can replicate the collective capabilities of natural communities, thereby negating the need for expensive preprocessing steps and significantly intensifying the economic benefit of using cheap feedstocks.

Here, team members characterized the metabolic interactions of the mutualistic co-culture of Clostridium thermocellum and Clostridium thermobutyricum, recently shown to be effective in converting lignocellulosic biomass to butyrate (Chi et al. 2018) and identified bottlenecks that could be relieved by augmentation with additional microbes to increase biomass conversion performance and efficiency. The approach is two-pronged. Researchers first used high-quality manually curated genome-scale metabolic models (GEMs) for both species to unravel the metabolic exchange network of the co-culture, compartmentalized as a community model containing 1,777 reactions, 1,679 metabolites, and 1,569 genes. This allowed the team to computationally identify metabolic bottlenecks responsible for the co-culture's limited butyrate production efficiency. Simultaneously, team members experimentally identified substrate inefficiencies in the co-culture setup by measuring solids deconstruction percentages and characterizing the monomeric and oligomeric sugars in the substrate left unused. Researchers tested the co-culture on raw corn stover and deacetylated-milled corn stover (DMR) and determined highest butyrate production from DMR (without enzymatic preprocessing), with a solid deconstruction of 83.1%. The largest percentage of leftover unused sugar moieties were xylose oligomers along with some arabinose and glucose oligomers.

To identify candidate isolates that could augment the co-culture to improve carbohydrate utilization, researchers collected various soil samples and enriched them on raw

corn stover, bagasse, switchgrass, poplar, and pine substrates as well as on supernatants from the DMR/co-culture experiments. The enrichments were carried out under anoxic conditions at 55°C, identical to those used with the *Clostridia* co-culture. They were passaged multiple times into fresh media, keeping the raw lignocellulosic plant materials as their sole carbon source, to ensure the selected members are producing butyrate, which was validated by highperformance liquid chromatography.

Following isolation, the isolates will be tested for compatibility with the co-culture using a high-throughput community design and construction method (Coker et al. 2022). These constructed consortia will then be further engineered to optimize production of butyrate. Engineering strategies will be guided by community metabolic models, ensuring the collective capability of the designed community is reproducible and optimized toward bioproduction. This study will lay the foundation for advanced bioproduction using multifaceted microbial communities inspired from nature and will expand knowledge on intra-community microbial interactions.

Chi, X., et al. 2018. "Hyper-Production of Butyric Acid from Delignified Rice Straw by a Novel Consolidated Bioprocess," *Bioresource Technology* **254**, 115–120.

Coker, J., et al. 2022. "A Reproducible and Tunable Synthetic Soil Microbial Community Provides New Insights into Microbial Ecology," *mSystems* 7, e00951-22.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award DE- SC0022137.

## Reproducible Plant Growth in Fabricated Ecosystems (EcoFAB 2.0) Reveals that Nitrogen Supply Modulates Root Exudation

Vlastimil Novak<sup>1</sup>, Peter F. Andeer<sup>1</sup>, Yi Zhai<sup>1</sup>, John P. Vogel<sup>1,2,3</sup>, Trent R. Northen<sup>1,2\*</sup> (trnorthen@lbl.gov), **Karsten Zengler<sup>4</sup>** 

<sup>1</sup>Lawrence Berkeley National Laboratory; <sup>2</sup>DOE Joint Genome Institute; <sup>3</sup>University of California–Berkeley; <sup>4</sup>University of California–San Diego

**Goals:** This project couples novel laboratory and field studies to develop the first predictive model of grassmicrobiomes based on new mechanistic insights into dynamic plant-microbe interactions in the grasses *Sorghum bicolor* and *Brachypodium distachyon* that improve plant nitrogen use efficiency (NUE). The results will be used to predict plant mutants and microbial amendments that improve low-input biomass production for validation in laboratory and field studies. To achieve this goal, the team will determine the mechanistic basis of dynamic exudate exchange in the grass rhizosphere, with a specific focus on the identification of plant transporters and proteins that regulate root exudate composition and how specific exudates select for beneficial microbes that increase plant biomass and NUE. Researchers will further develop a predictive plant-microbe model for advancing sustainable bioenergy crops and will predictively shift plant-microbe interactions to enhance plant biomass production and nitrogen (N) acquisition from varied N forms.

Understanding plant-microbe interactions requires examination of root exudation under nutrient stress using standardized and reproducible experimental systems. Researchers grew Brachypodium distachyon hydroponically in novel fabricated ecosystem devices (EcoFAB 2.0) under three inorganic N forms (NO<sub>3</sub>-, NH<sub>4</sub>+, or NH<sub>4</sub>NO<sub>3</sub>), followed by N starvation. In liquid chromatography with tandem mass spectrometry (LC-MS/MS), analyses of exudates, biomass, medium pH, and nitrogen uptake showed EcoFAB 2.0's low intra-treatment data variability. Furthermore, the three inorganic N forms caused differential exudation, generalized by abundant amino acids, peptides, and alkaloids. Comparatively, N-deficiency decreased N-containing compounds but increased shikimates and phenylpropanoids. Subsequent bioassays with two shikimates and phenylpropanoids (shikimic and p-Coumaric acids) on the rhizobacterium Pseudomonas putida or Brachypodium seedlings revealed that shikimic acid promoted bacterial and root growth, while *p*-Coumaric acid–stunted seedlings. The next objective was to identify transport mechanisms for organic acids and inorganic nitrogen by creating plant mutants with knockout ABC or nitrogen transporters. These mutations caused significant phenotypic and exometabolic changes. In conclusion, results suggest: (1) Brachypodium alters exudation in response to nitrogen status, which can affect rhizobacterial growth; (2) EcoFAB 2.0 is a valuable standardized plant research tool; and (3) the plant root exudation can be altered by membrane transport engineering.

**Funding Information**: The team gratefully acknowledges funding from the U.S. Department of Energy (DOE) Office of Science, Biological and Environmental Research Program. The research described was funded under contract DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory as part of a project lead by UC–San Diego (DE-SC0021234). The EcoFAB 2.0 was developed as part of the Trial Ecosystem Advancement for Microbiome Science (TEAMS) under contract DE-AC02-05CH11231 to Lawrence Berkeley National Laboratory project.

## Unraveling Metabolic Interactions in a Rhizosphere Microbial Community Through Genome-Scale Modeling

Manish Kumar\* (m3kumar@health.ucsd.edu), Maxwell Neal, Deepan Thiruppathy, Oriane Moyne, Gabriela Canto Encalada, Anurag Passi, Rodrigo Santibañez, Diego Tec-Campos, **Karsten Zengler** 

#### University of California-San Diego

Goals: This project couples novel laboratory and field studies to develop the first predictive model of grassmicrobiomes based on new mechanistic insights into dynamic plant-microbe interactions in the grasses Sorghum bicolor and Brachypodium distachyon that improve plant nitrogen (N) use efficiency (NUE). The results will be used to predict plant mutants and microbial amendments that improve low-input biomass production for validation in laboratory and field studies. To achieve this goal, researchers will determine the mechanistic basis of dynamic exudate exchange in the grass rhizosphere with a specific focus on the identification of plant transporters and proteins that regulate root exudate composition and how specific exudates select for beneficial microbes that increase plant biomass and NUE. The team will further develop a predictive plant-microbe model for advancing sustainable bioenergy crops and will predictively shift plant-microbe interactions to enhance plant biomass production and N acquisition from varied N forms.

The team presents manually curated genome-scale metabolic models for 17 bacteria commonly found in the rhizosphere. These bacteria isolated from the switchgrass rhizosphere represent dominant members of the rhizosphere of grasses belonging to various genera, such as Arthrobacter, Bacillus, Bosea, Bradyrhizobium, Brevibacillus, Burkholderia, Chitinophaga, Lysobacter, Methylobacterium, Mucilaginibacter, Mycobacterium, Niastella, Paenibacillus, Rhizobium, Rhodococcus, Sphingomonas, and Variovorax. The models were generated using standard pipelines and incorporated a total of 3,877 reactions and 2,663 metabolites. Each model contains between 790 to 1,788 genes, covering 15% to 30% of the respective microbial genomes. The models were curated using information from literature and public database s such as KEGG, UniProt, and MetaCyc. All models were simulated on 215 different carbon and nitrogen sources and results were compared with experimental measurements and to improve model predictability. The models accurately predicted the growth phenotypes of rhizosphere bacteria 90% of the time. Specifically, there were 1,475 truepositive predictions (correctly identified growth), 1,542 true-negative predictions (correctly identified no growth), 388 false-positive predictions (incorrectly identified

growth), and 250 false-negative predictions (incorrectly identified no growth).

Next, the team deployed these highly curated metabolic models to study the interactions in a synthetic microbial community (SynCom) of the rhizosphere. The team developed a computational framework for community metabolic models. For this the team included members in the community reconstruction with relative abundance above a minimum cutoff (>0.01%) under any tested condition. Each member's metabolic model was treated as a separate compartment linked to a shared metabolic pool with connections refined by experimentally determined phenotypic data (i.e., Biolog plates). Flux balance analysis was used to simulate the growth of individual members as well as that of the entire community. The biomass of each member and the community was optimized during the simulation. To further improve the computational framework, the team added a module that constrains the activity of reactions using metatranscriptomics and metatranslatomics data, reflecting internal resource allocation for each bacteria. This analysis predicts metabolic exchanges between community members and uncovers the nature of interactions, such as competition and cooperation, between rhizosphere microbes.

**Funding Information**: This work is supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award DESC0021234.

## A Gene-Editing System for Large-Scale Fungal Phenotyping in a Model Wood Decomposer

**Jiwei Zhang**<sup>1\*</sup>(zhan3437@umn.edu), Jonathan S. Schilling<sup>1\*</sup>(schillin@umn.edu), Igor Grigoriev<sup>2</sup>, Daniel Cullen<sup>3</sup>, Hugh D. Mitchell<sup>4</sup>

<sup>1</sup>University of Minnesota–Saint Paul; <sup>2</sup>Joint Genome Institute; <sup>3</sup>USDA Forest Products Laboratory and University of Wisconsin–Madison; <sup>4</sup>Environmental Molecular Sciences Laboratory at Pacific Northwest National Laboratory

**Goals:** The team combines CRISPR/Cas9-based genomeediting and network analysis for large-scale phenotyping in a model wood decomposer fungus relevant to the DOE mission area. The overall goal is to develop a high-throughput genetic platform that can allow the discoveries of distinctive genes and genetic features enabling the fast wood degradation in brown rot fungal species. Through this research, the team hopes to provide stand-alone tools and resources for discovering novel fungal genetic mechanisms that can be used in combination to advance relevant plant biomass conversion research in the post-genomic era. This research focuses on a group of unique wood decomposer basidiomycete fungi-brown rot fungi, which harbor the industrially relevant pathways to extract carbohydrates from lignocellulose and have broad relevance to global carbon cycling. Distinct from other fungi, brown rot species use nonenzymatic reactive oxygen species (ROS) mechanisms to modify lignin and selectively extract sugars. Their degradative mechanisms, from a process-efficiency standpoint, represent a pathway upgrade relative to the ancestral approaches in white rot species (Hibbett and Donoghue 2001; Eastwood et al. 2011). Fungi obtained this capacity evolutionarily by shedding rather than gaining carbohydrateactive enzymes (CAZys) repertoire genes (Martinez et al. 2009; Floudas et al. 2012; Riley et al. 2014). This paradox therefore makes brown rot fungi a promising candidate for discovering unknown genetic mechanisms governing plant biomass degradation. Although DOE mission relevance is clear, and the team has made major genomically informed advances in brown rot, progress is limited by an inability to manipulate genes in any brown rot fungal strain.

Reshuffling of fungal genome and gene regulation might play key roles in determining brown rot efficacy (Zhang et al. 2016; Zhang et al. 2019; Zhang et al. 2017). Using functional genomic tools, recently team members elucidated a staggered two-step (i.e., oxidation-then-hydrolysis) gene regulation model for brown rot (Zhang et al. 2016; Zhang et al. 2019). Although these genomic studies have greatly advanced understanding of brown rot, its genetic basis remains uncharacterized and unharnessed. Targeted gaps are still remaining for understanding its genetic mechanism. For example, (1) the functions of genes involved in the two-step model remain unverified and ambiguous, (2) the gene regulatory mechanism used to control and consolidate two steps is unclear, and (3) the functions of the majority of genes identified by multiomics are either hypothetical or unknown and are waiting for interrogation. The existence of these gaps is, to a large extent, due to the lack of a robust genome-editing tool that can validate and discover brown rot genetic features.

The team plans to integrate systems biology, genome-editing, and network modeling to address these key gaps. Progress toward accomplishing the project goals:

**Objective 1**: Create the CRISPR/Cas9-based editing system to validate brown rot gene functions.

To make the genetic manipulation available in brown rot fungal species, researchers first built a genetic transformation platform in a model species: *Gloeophyllum trabeum*. Researchers then developed a gene reporting system that relies on a laccase reporter gene and its rapid, colorimetric detection method for optimizing expression elements and transformation procedures (Li et al. 2023). With this, a codon-optimized eSpCas9 gene retrieved from *Streptococcus*  pyogenes was fused to eGFP and expressed in the *G. trabeum* nucleus led by a nucleoplasmin nucleus localization signal. The sgRNAs targeting a cellulase gene Cel5A (Gene ID 57704) were then expressed in the *G. trabeum*-Cas9 mutant for plasmid-based CRISPR-Cas9 gene-editing. Three U6 promoters from *Aspergillus niger*, *Trametes versicolor*, and *G. trabeum* were tested for their efficiencies in driving the sgRNA's expression. Parallelly, a preassembled Cas9-sgRNA ribonucleoprotein method was also tried in the same brown rot species for the scarless gene disruption. Once the gene-editing tool is fully built, researchers will continue to work on not functionally validated genes with the goal of generating a first-ever single-gene mutant library for brown rot phenotypic studies.

**Objective 2**: Build a carbon-utilization network to discover key genetic features used by brown rot.

To build a gene coexpression network for discovering novel brown rot genetic features, the team measured the transcriptomes of brown rot species in response to a broad spectrum of lignocellulose-derivative carbon sources. Two brown rot species, *G. trabeum* and *Rhodonia placenta*, were used for cross-species comparisons for discovering shared or distinct mechanisms. Using genome-wide co-expression modeling, key modules, and hub genes associated with lignocellulose polymers or monomers were discovered. DNA affinity purification (DAP-seq) was then used to identify the cis- and trans-regulatory elements involved in the carbon signaling pathway and revealed the regulatory machineries unique to brown rot (in preparation). Connecting to the whole project, this objective will complement the gene pool, aiming for large-scale phenotypic screening.

**Objective 3**: Integrate network analysis with CRISPR/Cas9 library for large-scale phenotypic screens.

This objective aims to develop a pipeline to use the multiplexing sgRNA library for genome-editing and mutant library construction for large-scale phenotypic screens, followed by next-generation sequencing to discover key functional genes. Tens of plasmid constructs were used for simultaneous delivery into *G. trabeum* and were combined with the laccase reporter to optimize a high-throughput transformation procedure. When this ongoing trial is successful, the team will continue to integrate it with the plasmid-based CRISPR/Cas9 method for large-scale gene editing.

Project researchers anticipate providing stand-alone tools and resources to elucidate fundamental microbial processes relevant to the DOE mission of advancing new engineering designs for lignocellulose bioconversion.

Eastwood, D. C., et al. 2011. "The Plant Cell Wall-Decomposing Machinery Underlies the Functional Diversity of Forest Fungi," *Science* **333**(762). DOI:10.1126/science.1205411.

Floudas, D., et al. 2012. "The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes," *Science* **336**(6089), 1715–19. DOI:10.1126/science.1221748.

Li, W., et al. 2023. "A Laccase Gene Reporting System That Enables Genetic Manipulations in a Brown Rot Wood Decomposer Fungus *Gloeophyllum trabeum*," *Microbiology Spectrum* **11**(1), e04246-22. DOI:10.1128/spectrum.04246-22.

Martinez, D., et al. 2009. "Genome, Transcriptome, and Secretome Analysis of Wood Decay Fungus *Postia placenta* Supports Unique Mechanisms of Lignocellulose Conversion," *Proceedings of the National Academy of Sciences* **106**(6), 1954–59. DOI:10.1073/ pnas.0809575106.

Riley, R., et al. 2014. "Extensive Sampling of Basidiomycete Genomes Demonstrates Inadequacy of the White-Rot/Brown-Rot Paradigm for Wood Decay Fungi," *Proceedings of the National Academy of Sciences* **111**(27), 9923–28. DOI:10.1073/pnas.1400592111.

Zhang, J., et al. 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**(39), 10968–73. DOI:10.1073/pnas.1608454113.

Zhang, J., et al. 2017. "Role of Carbon Source in the Shift from Oxidative to Hydrolytic Wood Decomposition by *Postia placenta*," *Fungal Genetics and Biology* **106**, 1–8. DOI:10.1016/j.fgb.2017. 06.003.

Zhang, J., et al. 2019. "Gene Regulation Shifts Shed Light on Fungal Adaption in Plant Biomass Decomposers," *mBio* **10**(6), e02176-19. DOI:10.1128/mBio.02176-19.

**Funding Information:** This research is supported by the U. S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0022151.

## **SCIENCE FOCUS AREA**

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

## IMAGINE BioSecurity: Genome-Scale Engineering and Modeling for Secure Biosystems Design

Jianping Yu<sup>1</sup>, Ayako Murao<sup>2</sup>, Jacob Sebesta<sup>1</sup>, Bin Yang<sup>1</sup>, Chao Wu<sup>1</sup>, Diana Hernandez Hernandez<sup>2</sup>, Melissa Amezola<sup>2</sup>, Lin Ding<sup>2</sup>, Gabriella Li<sup>1</sup>, Kathleen L. Arnolds<sup>1</sup>, Rodrigo Santibañez<sup>3</sup>, Juan D. Tibocha-Bonilla<sup>3</sup>, Karsten Zengler<sup>3</sup>, Wei Xiong<sup>1</sup>, Jeffrey G. Linger<sup>1</sup>, Yo Suzuki<sup>2\*</sup> (YSuzuki@jcvi.org), **Michael T. Guarnieri**<sup>1</sup>

<sup>1</sup>National Renewable Energy Laboratory; <sup>2</sup>J. Craig Venter Institute; <sup>3</sup>University of California–San Diego

**Goals:** The Integrative Modeling and Genome-scale Engineering for Biosystems Security (IMAGINE BioSecurity) Science Focus Area project seeks to establish an understanding of the behavior of engineered microbes in controlled versus environmental conditions to predictively devise new strategies for responding to biological escape. To this end, the IMAGINE team integrates core capabilities in synthetic and applied systems biology to develop a high-throughput platform for the design, generation, and analysis of biocontainment strategies in industrially relevant and emerging, next-generation microbes.

Genetically modified organisms (GMOs) have emerged as an integral component of a sustainable bioeconomy, with an array of applications in agriculture and bioenergy. However, the rapid development of GMOs and associated synthetic biology approaches raises several biosecurity concerns related to environmental escape of GMOs, detection thereof, and impact upon native ecosystems. To establish a secure bioeconomy, novel biocontainment strategies—informed by a fundamental understanding of systems-level governing mechanisms—are needed. To this end, the IMAGINE team is developing an array of passive and active synthetic-biocontainment strategies to effectively minimize laboratory-escape frequency while concurrently maintaining maximal laboratory performance. Researchers have selected a series of nonmodel, industrial, and next-generation microbial hosts to serve as chassis for secure biosystems design, including *Pseudomonas putida*, *Synechocystis sp.* PCC6803, *Clostridium ljungdahlii*, *Mycoplasma mycoides*, and *Saccharomyces cerevisiae*.

To facilitate the analysis of combinatorial constructs in the target organisms, a method termed combinatorial genetics en masse (CombiGEM; Wong et al. 2016) for generating combinatorial genotypes en masse and tracking them in mixed populations using DNA barcodes and nextgeneration sequencing was implemented. Combinatorial biocontainment strategies are being developed and evaluated for the capacity to reduce GMO-escape frequency in laboratory and environmental simulation settings. Additional efforts to target synthetic carbon, nitrogen, and phosphorus storage auxotrophies are under development. In parallel, researchers have initiated assessment of the metabolic burden associated with implementation of these strategies with the goal of maximizing biocontainment while maintaining optimal microbial fitness in deployment settings. Engineered strains are experimentally analyzed via growth, escape-frequency, and bioproductivity using high-throughput screening in laboratory and environmental mesocosm settings. Strains are concurrently subjected to fitness and escape frequency screening assays to assess the effect of genetic safeguards on strain fitness and biocontainment efficacy. Researchers have also initiated the assessment of the robustness and fitness of the engineered microbes via computational robustness and genome-scale metabolic modeling to understand the underlying mechanisms that govern the efficacy of biocontainment and metabolic fitness.

Systems-level analyses of these hosts in the absence and presence of biocontainment constraints will elucidate principles that (1) govern effective biocontainment and laboratory performance and (2) drive biological systems in their natural environments. These learnings will establish an extensive library of biocontainment modules and strains, testing platform, and systems knowledgebase, which will lay the foundation for predictive design of biocontainment strategies with enhanced stability and resilience in diverse microbial hosts. Combined, these efforts will reduce the risk associated with deployment of GMOs, ultimately forwarding a secure bioeconomy.

Arnolds, K. L., et al. 2021. "Biotechnology for Secure Biocontainment Designs in an Emerging Bioeconomy," *Current Opinion in Biotechnology* **71**, 25–31.

Wong, A. S. L., et al. 2016. "Multiplexed Barcoded CRISPR-Cas9 Screening Enabled by CombiGEM," *Proceedings of the National Academy of Sciences* **113**, 2544–49.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, Genomic Science program, Secure Biosystems Design Science Focus Area | IMAGINE BioSecurity: Integrative Modeling and Genome-Scale Engineering for Biosystems Security, under contract No. DE-AC36-08GO28308.

## IMAGINE BioSecurity: Mesocosm-Based Methods to Evaluate Biocontainment Strategies and Impact of Industrial Microbes Upon Native Ecosystems

Kathleen L. Arnolds, Riley C. Higgins, Gabriella Li, Jeffrey G. Linger, **Michael T. Guarnieri**\* (Michael.Guarnieri@nrel.gov)

National Renewable Energy Laboratory

**Goals:** The Integrative Modeling and Genome-Scale Engineering for Biosystems Security (IMAGINE BioSecurity) Science Focus Area project seeks to establish an understanding of the behavior of engineered microbes in controlled versus environmental conditions to predictively devise new strategies for responding to biological escape. To this end, the IMAGINE Team integrates core capabilities in synthetic and applied systems biology to develop a high-throughput platform for the design, generation, and analysis of biocontainment strategies in industrially relevant and emerging next-generation microbes.

Genetically modified industrial-production microbes and their associated bioproducts have emerged as an integral component of a sustainable bioeconomy. However, the rapid development of these innovative technologies raises biosecurity concerns, namely, the risk of environmental escape. Thus, the realization of a bioeconomy hinges not only on the development and deployment of microbial production hosts but also on the development of secure biosystems and biocontainment designs. Current laboratory-based biocontainment testing systems do not accurately reflect complexities found in natural environments, necessitating an environmentally relevant analysis pipeline that allows for the detection of rare escapees, the effect of associated bio-products, and the impact on native ecologies. To this end, researchers have developed an approach that utilizes soil mesocosms and integrated systems analyses to evaluate the efficacy of novel biocontainment strategies and to assess the impact of production systems upon terrestrial microbiome dynamics. Researchers demonstrate the utility of this approach by modeling a perturbation event by contaminating the mesocosms with the industrial microbial chassis, Saccharomyces cerevisiae. The resultant data demonstrate that researchers can track the fate of the contaminating microbe with high sensitivity in the soil, as well as monitor broader impacts of the perturbation on the underlying soil microbiome with a high degree of spatial and temporal resolution. The findings presented here support the use of this mesocosm-based approach to assess the environmental impact of industrial microbes and to validate biocontainment strategies.

Arnolds K. L., et al. 2021. "Biotechnology for secure biocontainment designs in an emerging bioeconomy," *Current Opinion in Biotechnology* **71**, 25–31.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), Genomic Science program, Secure Biosystems Design Science Focus Area | IMAGINE BioSecurity: Integrative Modeling and Genome-scale Engineering for Biosystems Security, under contract No. DE-AC36-08GO28308.

## **SCIENCE FOCUS AREA**

From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms

## **Engineering Overlapping Genes in Bacteria**

Sean P. Leonard<sup>1</sup>, Jennifer L. Chlebek<sup>1</sup>, Jose Manuel Martí<sup>1</sup>, Hunter Nisonoff<sup>2</sup>, Chloe Hsu<sup>2</sup>, Charlotte Rochereau<sup>3</sup>, Christina Kang-Yun<sup>1</sup>, Harris Wang<sup>3</sup>, Jennifer Listgarten<sup>2</sup>, Jonathan Allen<sup>1</sup>, Dante P. Ricci<sup>1</sup>, Dan Park<sup>1</sup>, **Yongqin Jiao<sup>1\*</sup>** (jiao1@llnl.gov)

<sup>1</sup>Lawrence Livermore National Laboratory; <sup>2</sup>University of California– Berkeley; <sup>3</sup>Columbia University

https://sc-programs.llnl.gov/biological-and-environmentalresearch-at-llnl/secure-biosystems-design

Goals: A primary goal of the science focus area (SFA) is to establish genetic sequence entanglement—in which two genes are encoded within the same DNA sequence through use of alternative reading frames—as a generalizable biocontainment strategy to protect engineered functions against mutational inactivation and to mitigate the horizontal transfer of invasive genes. Achieving sequence entanglement remains a significant challenge due to sequence constraints that necessitate large-scale redesign of the entangled proteins. Through design-build-test-learn (DBTL) iterations using the Constraining Adaptive Mutations using Engineered Overlapping Sequences, eXtended (CAMEOX) algorithm (Blazejewski, Ho, and Wang 2019), high-throughput functional assays, and state-of-the-art machine learning and protein-structural-prediction algorithms, researchers aim to improve the accuracy of entanglement designs and expand the application to a broad range of microbes.

To generate an initial data set for model training and development, researchers initiated a DBTL campaign of an entanglement pair composed of *infA* and *aroB*. *infA* encodes the translation initiation factor 1 (72 AA) that is essential for growth and *aroB* encodes 3-dehydroquinate synthase (362 AA) that is required for aromatic amino acid biosynthesis. CAMEOX was used to generate 130,000 entanglement designs, among which, 2,000 designs were selected for experimental testing. The functionality of *infA* and *aroB* were assayed separately through selection. The results revealed that between ~10 t0 30% of *aroB* variants and ~25% of *infA* variants were highly enriched in the surviving population, indicative of protein function. Researchers found that the gene fitness–scoring metric generated by CAMEOX— pseudolikelihood score—correlates well with the experimental enrichment scores, confirming the pseudolikelihood score as a reliable indicator for protein fitness. Combing the results from both assays for *infA* and *aroB*, 14 variants were identified with high enrichment values for both genes and are being tested for functionality within the entangled context.

Using the experimentally measured fitness data for *infA* and *aroB*, the team trained random forest (RF) classifiers based on amino acid composition and used the model to predict the fitness of CAMEOX-designed entanglement solutions. Researchers found that simple classifiers such as the frequency of certain amino acids were able to predict variant fitness with high accuracy, even when a small number of measured variants was used in the training set. Using these RF models, researchers screened the complete set of 130,000 *infA/aroB* CAMEOX designs and identified 29 with potential functionality for both genes. Experimental testing of these variants is underway.

In addition to RF models, researchers have leveraged AlphaFold to rank CAMEOX variants according to how foldable their structures are. Relying on AFRank (Roney and Ovchinnikov 2022), the team predicted the structures of CAMEOX variants and used the predicted confidence metrics as a proxy for variant fitness. This approach complements sequence-based screening methods to in silico select the best variants for experimental testing. Besides better ranking the variants, researchers have also modified the algorithm to expand the diversity of proposed CAMEOX solutions. Researchers have developed gradient-based Markov Chain Monte Carlo (MCMC) methods for designing the entangled nucleotide sequences. This new optimization protocol generally improves over the previous greedy optimization algorithms and enables the generation of more diverse sequences with better fitness scores.

In addition to engineering entanglement for specific gene pairs, the team seeks to comprehensively assess entanglement feasibility of a wide array of gene pairs to identify features of DNA/protein sequences that make genes co-encoding more amenable. Leveraging improved speed and automation of CAMEOX, the team undertook a campaign to computationally generate entanglement designs for nearly all conditionally essential genes in *Escherichia coli* (94) by entangling them with one another. An additional 24 genes of interest (positive controls with naturally entangled phiX174 genes, reporter genes that allow for quantitative phenotypic characterization, and antibiotic resistance cassettes) were also included. This *in silico* campaign yielded >8 million pairwise entanglements solutions in both +1 and +2 reading frames. Using custom evolutionary models of the parental proteins, the team developed a scoring rubric for CAMEOX designs that allows the ability to quantify and compare the entangle-ability of individual genes as well as the compatibility of gene pairs. By generating and testing specific designs, the team found CAMEOX can successfully design functional protein variants with low sequence conservation (<50% identify) to their natural orthologs. Furthermore, researchers have identified specific features of proteins that correlate with better entanglement outcomes such as enrichment for amino acids with a higher degree of codon degeneracy, a property also observed with naturally occurring entangled genes.

Blazejewski, T., H.-I. Ho, and H. H. Wang. 2019. "Synthetic Sequence Entanglement Augments Stability and Containment of Genetic Information in Cells," *Science* **365**, 595–598.

Roney, J. P., and S. Ovchinnikov. 2022. "State-of-the-Art Estimation of Protein Model Accuracy using Alpha Fold." *Physical Review Letters* **129**(23), 238101.

**Funding Information:** This work is supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Lawrence Livermore National Laboratory Secure Biosystems Design SFA "From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms." Work at LLNL is performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

## Engineered Overlapping Genes Paired with Directed Evolution Prolongs the Evolutionary Stability of a Genetic Circuit

Jennifer L. Chlebek, Christina Kang-Yun, Sean P. Leonard, Dante P. Ricci, Mimi Cho Yung, Yongqin Jiao, **Dan Park**\* (park36@llnl.gov)

Lawrence Livermore National Laboratory

https://sc-programs.llnl.gov/biological-and-environmentalresearch-at-llnl/secure-biosystems-design

**Goals:** A primary goal of this science focus area (SFA) project is to establish genetic sequence entanglement as a generalizable biocontainment strategy to protect engineered functions against mutational inactivation and to mitigate the horizontal transfer of invasive genes. Sequence entanglement was inspired by overlapping genes found in many viral genomes and involves the synthetic encoding (entangling) of two genes within the same DNA sequence space through use of alternative reading frames. As such, mutations within the entangled region likely impact the function of both genes, providing a mechanism to constrain the allowable mutational space. The team thus hypothesize that by entangling a gene-of-interest (GOI) with an essential gene, the evolution

of the GOI can be constrained by rendering mutations in one frame nonpermissible due to deleterious mutations in the frame encoding the essential gene. As a proof-of-concept, researchers assessed the utility of an entanglement design in which a toxin is entangled with an essential gene, to improve genetic stability of a kill-switch circuit.

The development of synthetic biological circuits that maintain functionality over application-relevant timescales remains a significant challenge. Synthetic circuits are often burdensome to cellular fitness and are subject to evolutionary pressures, which select for mutated and nonfunctional circuits. In this study, researchers employed a gene overlap technique called synthetic sequence entanglement—in which one protein is encoded entirely within an alternate reading frame of another gene-to enhance the sequence stability of a burdensome engineered genetic circuit. Specifically, the toxin-encoding *relE* gene was entangled within *ilvA*, which encodes threonine deaminase, an enzyme essential for isoleucine biosynthesis. This pairing allows the ability to test whether an essential function (isoleucine biosynthesis) can increase the mutational robustness of a gene prone to mutational inactivation (e.g., *relE*).

Starting from a partially functional entanglement design in which significant missense mutations (~79% of entangled residues) were introduced within the *ilvA* sequence to accommodate a wild-type amino acid sequence for RelE, the team made targeted modifications of an internal ribosome-binding site that simultaneously enhanced the expression of the RelE toxin and the function of IlvA. Using this optimized design, researchers show that entanglement of *relE* with *ilvA* significantly increased the evolutionary stability of the toxic *relE* gene, which retained function for >130 generations. This stabilizing effect was achieved through a complete alteration of the allowable mutational landscape such that mutations inactivating both entangled gene products were disfavored. Instead, small deletions, insertions, and point mutations accumulated within the regulatory region of *ilvA* for the majority of lineages. By reducing baseline *relE* expression, these more benign mutations lowered circuit burden, which suppressed the accumulation of *relE* inactivating mutations, thereby prolonging kill-switch function. Overall, this work demonstrates the utility of sequence entanglement to increase the evolutionary stability of burdensome synthetic circuits.

**Funding Information:** This work is supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Lawrence Livermore National Laboratory Secure Biosystems Design SFA "From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms". Work at LLNL is performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-ABS-845282).

## **SCIENCE FOCUS AREA**

Persistence Control of Engineered Functions in Complex Soil Microbiomes

## Development of Synthetic Microbial Communities to Study Consortium Engraftment Dynamics and to Improve the Yield Performance of Sorghum

Citlali Fonseca<sup>1</sup>, Ryan McClure<sup>2</sup>, Dean Pettinga<sup>1</sup>, Andrew Wilson<sup>2</sup>, Henri Baldino<sup>2</sup>, Ritu Shrestha<sup>2</sup>, Jesus Sanchez<sup>1</sup>, Claudia Castro<sup>1</sup>, Joshua Elmore<sup>2</sup>, Andrew Frank<sup>2</sup>, Pubudu Handakumbura<sup>3</sup>, Devin Coleman-Derr<sup>1</sup>, **Robert Egbert<sup>2\*</sup>** (robert.egbert@pnnl.gov)

<sup>1</sup>University of California–Berkeley; <sup>2</sup>Pacific Northwest National Laboratory (PNNL); <sup>3</sup>Environmental Molecular Sciences Laboratory *https://qenomicscience.energy.gov/pnnlbiosystemsdesign/* 

Goals: The PNNL Persistence Control Science Focus Area (SFA) aims to gain a fundamental understanding of factors governing the persistence of engineered microbial functions in rhizosphere environments. With this understanding, the SFA is investigating design principles to control the environmental niche of native rhizosphere microbes. Researchers are examining the efficacy of genome reduction and metabolic addiction to plant-root exudates in environmental isolates as persistence-control strategies using the bioenergy crop sorghum and defined microbial communities as a model ecosystem. The engraftment dynamics of non-native microbes into a reduced-complexity microbial community and the establishment of defined-isolate synthetic communities in field environments are two major areas of current investigation. Effective persistence control will lead to secure plant-microbe biosystems that promote stress-tolerant and highly productive biomass crops.

In the past 2 decades, research of the plant microbiome has shown the importance of plant-associated microbes (PAM) in modulating crop performance (Compant et al. 2019). These studies have paved the way for the use of PAM to provide economic and sustainable solutions to current bioenergy cropping and, more generally, agricultural challenges. However, applying PAM within the context of field agriculture has met with mixed success, in part because the introduced microbes must persist within the context of a resilient existing microbiome to be successful. PAM engineering may help overcome these limitations, including via the use of strategies like genome reduction to control the environmental niche of target microbes in agricultural soils (Ke, Wang, and Yoshikuni 2021). A key step to implementing this approach is understanding how engineered microbes may co-colonize or be suppressed by the native microbiome.

In the Persistence Control SFA, researchers have developed two representative, reduced-complexity microbial communities to aid understanding of the colonization dynamics of engineered microbes in field-like conditions. First, a naturally evolved consortia of ~50 species was developed through repeated dilutions and plate passaging on synthetic growth media emulating the rhizosphere nutrient environment of sorghum. Here, with this reduced-complexity community, the team describes new assays examining the engraftment efficiency of an engineered host that is not part of the enrichment community, as well as taxonomic differences in communities that allowed or rejected colonization of the engineered host. These analyses showed that engraftment was possible but appeared to be the exception rather than the rule. In addition, successfully engrafted species showed strong co-abundances with other members of the community, pointing to possible points of microbial interaction that may drive engraftment. These studies begin to reveal the mechanisms behind how addition of a microbe to an existing community as a method of PAM engineering might take place.

To investigate the environmental persistence and plantgrowth promotion in field environments, researchers developed a defined synthetic community from sorghum rhizosphere microbiome isolates and tested this community with sorghum in growth chambers and the field. In contrast to the reduced-complexity enrichment community, this community was established using network co-abundance analysis from Sorghum field 16S taxonomic surveys by co-culturing Sorghum isolates of 56 member strains representing 18 bacterial genera. The project demonstrates that this synthetic community can stably colonize the rhizosphere and roots of sorghum during lab-based in planta experiments, and that it enhances overall shoot biomass compared to mock treated controls. Remarkably, field experiments replicate the findings observed in the labbased in planta data. These results reveal that the synthetic community is a stable and reproducible community that colonizes Sorghum plants and unexpectedly improves their performance in agricultural soils. The team anticipates the enrichment and defined communities will reveal the drivers of isolate colonization into rhizosphere microbiomes and enable the scaling of rhizosphere synthetic biology from laboratory to field settings. This knowledge will promote the responsible deployment of engineered microbial functions in cropping settings to reduce nutrient inputs, promote drought resilience, and suppress plant pathogens.

Compant, S., et al. 2019. "A Review on the Plant Microbiome: Ecology, Functions, and Emerging Trends in Microbial Application," *Journal of Advanced Research* **19**, 29–37.

Ke, J., B. Wang, and Y. Yoshikuni. 2021. "Microbiome Engineering: Synthetic Biology of Plant-Associated Microbiomes in Sustainable Agriculture," *Trends in Biotechnology* **39**, 244–61.

**Funding Information:** This research was supported by the U.S. Department of Energy's (DOE) Biological and Environmental Research (BER) Program as part of BER's Genomic Science program (GSP) and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design SFA "Persistence Control of Engineered Functions in Complex Soil Microbiomes." A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by BER and located at PNNL. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.

## Refactoring Metabolism to Control the Persistence of Genetically Engineered Microorganisms in the Environment

Joshua Elmore<sup>1\*</sup> (joshua.elmore@pnnl.gov), Adam Guss<sup>2</sup>, Ryan McClure<sup>1</sup>, Gara Dexter<sup>2</sup>, Henri Baldino<sup>1</sup>, Jay Huenemann<sup>2</sup>, Ryan Francis<sup>1</sup>, George Peabody V<sup>2</sup>, Jessica Martinez-Baird<sup>2</sup>, Lauren Riley<sup>2</sup>, Tuesday Simmons<sup>3</sup>, Devin Coleman-Derr<sup>3</sup>, Andrew Wilson<sup>1</sup>, Elise Van Fossen<sup>1</sup>, Ritu Shrestha<sup>1</sup>, Andrew Frank<sup>1</sup>, Valentine V. Trotter<sup>4</sup>, Brenton Poirier<sup>1</sup>, Young-Mo Kim<sup>1</sup>, Adam M. Deutschbauer<sup>4</sup>, Yasuhiro Oda<sup>5</sup>, William C. Nelson<sup>1</sup>, Yuliya Farris<sup>1</sup>, Aaron Ogden<sup>1</sup>, Caroline Harwood<sup>5</sup>, William Alexander<sup>3</sup>, **Robert Egbert<sup>1</sup>** 

<sup>1</sup>Pacific Northwest National Laboratory (PNNL); <sup>2</sup>Oak Ridge National Laboratory; <sup>3</sup>University of California–Berkeley; <sup>4</sup>Lawrence Berkeley National Laboratory; <sup>5</sup>University of Washington

**Goals:** The Persistence Control Science Focus Area (Per-Con SFA) at PNNL is focused on developing fundamental understanding of factors governing the persistence of engineered microbial functions in rhizosphere environments. From this understanding, the team will establish design principles to control the environmental niche of native rhizosphere microbes for the model bioenergy crop sorghum through data-driven genome reduction and engineered metabolic addiction to plant-root exudates. These principles will lead to secure plant-microbe biosystems that promote secure, stress-tolerant, and highly productive biomass crops. Persistence control is an engineering approach in which survival of genetically modified microorganisms is restricted to a target environmental niche. Refactoring the catabolic repertoire of these organisms such that they have reduced fitness outside of the target niche and enhanced fitness within the target niche is one of several powerful tools for achieving persistence control (see Fig. 15, p. 109). Within the PerCon SFA, researchers are refactoring the catabolism of plant growth–promoting rhizobacteria (PGPR) to use crop-specific root exudate compounds and not use abundant, wide-spread soil compounds (e.g., lignocellulose, chitin, and non-crop-specific root exudates) as carbon sources. This has the potential to enable responsible application of beneficial genetically modified PGPR in the environment, while preventing their uncontrolled spread outside of the crop rhizosphere. Through the use of Sorghum bicolor (sorghum) as a model crop, and with random barcode transposon mutagenesis (RB-TnSeq), researchers have identified genes whose removal from the model sorghum rhizobacteria Pseudomonas facilor TBS28 will abolish its use of common carbon sources. Sorgoleone (a lipophilic benzoquinone) and MHPP (a phenylpropanoid methyl ester) are two compounds entirely or almost entirely unique to sorghum root exudates, respectively. This makes them excellent model nutrients for expansion of the TBS28 catabolic repertoire, but the catabolic pathways for their consumption are unknown. Researchers isolated three bacteria, including a new species, that utilize sorgoleone and found that the model rhizobacteria Pseudomonas fluorescens SBW25 utilizes MHPP. Using RB-TnSeq and RNAseq individually, dozens of genes were found that were potentially involved in sorgoleone and MHPP catabolism. When used together, the number of candidate genes was reduced by an order of magnitude, enabling detection of key enzymes in each catabolic pathway. A core set of four highly conserved genes are essential for use of sorgoleone, serve as a biomarker for sorgoleone catabolic function, and are enriched in members of the sorghum rhizosphere. With a combination of genetic characterization and biochemical enzyme assays, the team identified two enzymes responsible for funneling MHPP into a common aromatic catabolism pathway. This includes a novel family of esterases that demethylates MHPP and other plant-derived phenylpropanoid methyl esters. To enable facile and efficient site-specific integration of non-native catabolic pathways and other heterologous genetic programs into bacterial genomes, researchers developed serine recombinase-assisted genome engineering (SAGE; Elmore et al. 2023). For this project, the team is actively using a multiplexed version of SAGE to evaluate genetic parts for rational control of gene expression in diverse bacteria, complement gene deletions for pathway discovery, introduce machinery for CRISPR interference, and engineer PGPR that can consume MHPP and sorgoleone as carbon sources.



Fig. 15. Refactoring the metabolism of bacteria for persistence control. [Courtesy Pacific Northwest National Laboratory]

Elmore, J. R. et al. "The SAGE Genetic Toolkit Enables Highly Efficient, Iterative Site-Specific Genome Engineering in Bacteria," *Science Advances*. In press.

**Funding Information:** This research was supported by the U.S. Department of Energy's (DOE) Biological and Environmental Research (BER) Program as part of BER's Genomic Science program (GSP) and is a contribution of the Pacific Northwest National Laboratory (PNNL) Secure Biosystems Design SFA "Persistence Control of Engineered Functions in Complex Soil Microbiomes." A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by BER and located at PNNL. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.

## **SCIENCE FOCUS AREA**

## SEED: Secure Ecosystem Engineering and Design

## Secure Ecosystem Engineering and Design (SEED) to Enable Safe Biodesign of Novel Plant-Microbe Interactions

Xiaohan Yang\*, Yang Liu, Mengjun Shu, Joanna Tannous, Tomas Rush, Kuntal De, Wellington Muchero, Jay Chen, **Paul E. Abraham** 

Oak Ridge National Laboratory (ORNL)

https://seed-sfa.ornl.gov/

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

Advancements made in plant engineering are necessary to address future challenges associated with climate change and food security. CRISPR/Cas9-based genome engineering now provides novel methods for accelerating high precision engineering in nonmodel plants. Yet nearly all genetic editing is created through tissue culture-based plant transformation systems, and these are often poorly developed in nonmodel plant species. Moreover, it is currently difficult to predict the activity of CRISPR using existing bioinformatic methods. Virus-mediated delivery of CRISPR/Cas systems has great potential to improve the delivery needed to expedite and maximize the usefulness of this technology. Because it is a challenge to deliver an entire CRISPR/Cas tool using RNA viruses, researchers recently developed an intein-mediated split-nCRISPR/Cas9 technology to deliver an entire based-editing CRISPR/Cas system into plants (Yuan et al. 2021). Biocontainment of these advanced genome-engineering tools is important to mitigate risks of unwanted genome engineering. Therefore,

the team developed a biosensor for real-time detection of active CRISPR/Cas tools *in planta* and an anti-CRISPR (Acr) protein countermeasure to limit unwanted CRISPR/ Cas9-based genome editing activity *in planta* (Yuan et al. 2022; Liu et al. 2023). These advancements are important steps towards safe, high-throughput plant biodesign and genome engineering.

Targeted genome editing of plants alone may not facilitate the advancements necessary to achieve the Department of Energy's climate and economic competitiveness goals. Emerging research on plant holobiont theory and microbial invasion ecology emphasizes the importance of plantmicrobe interactions. However, researchers currently lack the knowledge necessary to successfully introduce beneficial alterations, prevent undesired modifications, or assess the risks of proposed ecosystem engineering efforts. Therefore, advancements are being made to detect and control novel plant-microbe interactions for safe biodesign. Researchers are currently developing plant-based biosensors to detect the establishment of fungi on poplar, and the Plasminogen-Apple-Nematode (PAN) domain was recently recognized for its important role in plant host-cell invasion, which will serve as a useful target for engineering plants to control microbial invasion. Lastly, plant-delivered in situ engineering is being developed to control root-associated microbes through the delivery of small-secreted proteins. Preliminary results indicate these advancements have potential for engineering plants to detect and control associated microbes and thus facilitate new opportunities of safe ecosystem engineering.

Liu, Y., et al. 2023. "Expanding the Application of Anti-CRISPR Proteins in Plants for Tunable Genome Editing," *Plant Physiology* kiad076. DOI:10.1093/plphys/kiad076.

Yuan, G., et al. 2021. "Plant-Based Biosensors for Detecting CRIS-PR-Mediated Genome Engineering," *ACS Synthetic Biology* **10**(12), 3600–03.

Yuan, G., et al. 2022. "An Intein-Mediated Split–nCas9 System for Base Editing in Plants," *ACS Synthetic Biology* **11**(7), 2513–17.

**Funding Information:** The SEED SFA is sponsored by the Genomic Science program (GSP), U.S Department of Energy, Office of Science, Biological and Environmental Research (BER) Program, under FWP ERKPA17. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract No. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through GSP, BER Program, under FWP ERKP123.

## Secure Ecosystem Engineering and Design (SEED) to Mitigate the Impacts of Non-Native Fungal Pathogens on Managed Ecosystems

Joanna Tannous<sup>1\*</sup> (tannousj@ornl.gov), Tomas Rush<sup>1</sup>, Dana L. Carper<sup>1</sup>, Alyssa A. Carrell<sup>1</sup>, Cole Sawyer<sup>1</sup>, Miranda Clark<sup>1</sup>, Wellington Muchero<sup>1</sup>, Jared LeBoldus<sup>2</sup>, David Kainer<sup>1</sup>, Daniel Jacobson<sup>1</sup>, **Paul E. Abraham<sup>1</sup>** 

<sup>1</sup>Oak Ridge National Laboratory (ORNL); <sup>2</sup>Oregon State University

#### https://seed-sfa.ornl.gov/

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

Invasion of non-native fungal species is acknowledged as one of the major external drivers altering the structure, biodiversity, and function of ecosystems (Pyšek and Richardson 2010). Understanding the mechanisms of establishment of these invaders and developing mitigation approaches to manage them is a critical aspect of sustaining native biodiversity and normal ecosystem functions. The fungal pathogen Sphaerulina musiva is a well-characterized example of an invasive species spread unintentionally by human activities (Abraham et al. 2018). Originally native to Eastern North America, S. musiva was only recently introduced and established into the Pacific Northwest of North America, resulting in deleterious effects on susceptible Populus species/ genotypes, a foundational bioenergy crop, and a keystone tree species in forested ecosystems (Herath et al. 2016; Sakalidis et al. 2016).

Within the SEED SFA, a goal is to identify genetic determinants that alter S. musiva establishment, spread, and overall impact in DOE-managed Populus ecosystems to guide engineering and risk mitigation strategies for more sustainable and productive systems. As a critical first step, researchers are developing an S. musiva pangenome representing isolates collected across the United States and generating genomewide association study resources for high-throughput genotype-to-phenotype discovery. These resources have already uncovered genetic associations for several important establishment and pathogenicity traits that can be exploited by developed genomic engineering approaches such as CRISPR-enabled gene drives. To enable future biodesign on S. musiva, the team developed the first transformation system using a protein-based version of the CRISPR-Cas9 genome editing system.

Ultimately, the ongoing research will address fundamental knowledge gaps related to anthropogenic-assisted microbial invasions and guide future biosystems design strategies that safely prevent undesired modifications in DOE-managed ecosystems.

Abraham, N. D., et al. 2018. "Multiplex qPCR for Detection and Quantification of *Sphaerulina musiva* in *Populus* Stems," *Plant Pathology* 67, 1874–82.

Herath, P., et al. 2016. "Anthropogenic Signature in the Incidence and Distribution of an Emerging Pathogen of Poplars," *Biological Invasions* 18, 1147–61.

Pyšek, P., and D. M. Richardson. 2010. "Invasive Species, Environmental Change and Management, and Health," *Annual Review of Environment and Resources* **35**, 25–55.

Sakalidis, M. L., et al. 2016. "Genetic Patterns Reveal Historical and Contemporary Dispersal of a Tree Pathogen," *Biological Invasions* **18**, 1781–99.

**Funding Information:** The SEED SFA is sponsored by the Genomic Science program (GSP), U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program, under FWP ERKPA17. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract No. DE-AC05-00OR45678. This program is supported by the U. S. Department of Energy, Office of Science, through GSP, BER Program, under FWP ERKP123.

# OTHER

## InCoGenTEC

## **Integrases on Demand**

Kelly Williams<sup>1\*</sup>(kpwilli@sandia.gov), Jesse Cahill<sup>1</sup>, Lauren Clark<sup>2</sup>, Catherine Mageeney<sup>1</sup>, Laura Quinto<sup>4</sup>, Michael C. Jewett<sup>2,3</sup>, Farren J. Isaacs<sup>4</sup>, **Joe Schoeniger**<sup>1</sup>

<sup>1</sup>Sandia National Laboratories; <sup>2</sup>Northwestern University; <sup>3</sup>Stanford University; <sup>4</sup>Yale University

The Intrinsic Control for Genome and Transcriptome Editing in Communities (InCoGenTEC) project sponsored by the BER BSSD Secure Biosystems Design initiative conducts mechanistic studies of gene flows between bacteria encompassing broad phylogenetic diversity and evolutionary time, focusing on mechanisms that permit natural gene delivery and gene integration. In prokaryotes, DNA rearrangements are physiologically and ecologically important and an important potential source of genome instability and loss of biocontainment of engineered features. Goals include comprehensively mapping and classifying bacterial genomic islands, analyzing mechanisms of mobility, and identifying routes of horizontal gene transfer. The team will also use this information to identify prophages suitable as vectors for editing microbial community members and to improve design of synthetic genetic elements that can be integrated into target organisms.

Recent advances in genome editing have stimulated a renewed interest in microbial DNA integrases. Previously the team published bioinformatic methods for precise high-throughput definition of bacterial and archaeal genomic islands that contain integrases (Mageeney, et al. 2020). These methods enable pairing of integrases with their specific attachment site (Att) sequences. As part of InCo-GenTEC, the team has mapped about 1 million genomic islands in approximately 350,000 bacterial genomes, yielding tens of thousands of unique integrase-Att pairs.

Researchers have implemented assays for high-throughput in vitro and in vivo characterization of integrase activity to verify integrase activity at predicted Att sites as well as investigate integrase dependence on host-species factors that might limit the range of horizontal gene transfer. Considerable work has been published on the use of serine integrases for synthetic biology and genome editing of target organisms, typically requiring first the introduction of a heterologous Att site (landing pad) into the target genome. Less attention has been paid to the much more abundant tyrosine integrases because of their putative dependences on microbial-host factors. Researchers have used the project's assays to demonstrate activity in E. coli of both serine and tyrosine integrases drawn from broad phylogenetic spans. On the other hand, the team also demonstrated that its methods can be used to identify from near neighbors of a target strain integrases that utilize native Att sites in the target genome, simplifying and potentially greatly increasing the efficiency of bacterial genome editing. Finally, team members have used modified versions of its assays to mine difficult to identify directionality factors (excisionases) from genomic islands. These methods have the potential to significantly improve genome editing efficiency and enable layered genome restructuring in diverse organisms while improving understanding of the biochemical determinants of the specificity of integrase activity and mobile element host-range.

Mageeney, C. M., et al. 2020. "New Candidates for Regulated Gene Integrity Revealed Through Precise Mapping of Integrative Genetic Elements," *Nucleic Acids Research* **48**(8), 4052–65.

**Funding Information:** Funding was provided by the U.S. Department of Energy, Office of Science, through the Genomic Science Program, Biological and Environmental Research Program, under the Secure Biosystems Design Initiative project Intrinsic Control for Genome and Transcriptome Editing in Communities (InCoGenTEC); Sandia National Laboratories is managed by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy [DE-NA-0003525].



## **UNIVERSITY PROJECTS**

## Conversion of Lignocellulosic Plant Biomass into Industrial Chemicals via Metabolic Engineering of Two Extreme Thermophiles, *Caldicellulosiruptor bescii* and *Pyrococcus furiosus*

Ryan G. Bing<sup>1\*</sup> (rgbing@ncsu.edu), Tania N. N. Tanwee<sup>2</sup>, Hailey O'Quinn<sup>2</sup>, Jason Vailionis<sup>3</sup>, Gina L. Lipscomb<sup>2</sup>, James Crosby<sup>1</sup>, Tunyaboon Laemthong<sup>1</sup>, Ke Zhang<sup>3</sup>, Ying Zhang<sup>3</sup>, Dmitry Rodionov<sup>4</sup>, Robert M. Kelly<sup>1</sup>, **Michael W. Adams**<sup>2</sup>

<sup>1</sup>North Carolina State University; <sup>2</sup>University of Georgia; <sup>3</sup>University of Rhode Island; <sup>4</sup>Sanford Burnham Prebys Medical Discovery Institute

**Goals:** This project aims to metabolically engineer two extreme thermophiles, *Caldicellulosiruptor bescii* ( $T_{max}$  90°C) and *Pyrococcus furiosus* ( $T_{max}$  103°C), to convert lignocellulosic plant biomass into industrial chemicals including acetone, 2,3-butanediol, 1-propanol, 3-hydroxypropionate, and ethanol. This work includes efforts to reincorporate  $CO_2$ , which is generated by fermentation, into desired products powered by energy recovered from hydrogen, which is also produced during fermentation. Some of the native enzymes of *C. bescii* that degrade lignocellulose will be expressed in *P. furiosus* to allow growth on cellulose and xylan. Additionally, system-wide metabolic and regulatory models for both organisms will be leveraged to optimize conversion, yield, and selectivity of plant biomass to industrial chemicals.

Conversion of lignocellulosic plant biomass into industrial chemicals has potential to provide renewable, sustainable sources of non-electrifiable fuels and chemicals. Enzymatic degradation and conversion of lignocellulose into desirable chemicals offers possible energy and monetary savings compared to chemical or mechanical methods. Consolidated bioprocessing aims to combine biological deconstruction and conversion of plant biomass into a single step to further increase these savings. Some extreme thermophiles, such as C. bescii, excel at deconstruction of plant biomass which is in part aided by the high temperatures (Straub et al. 2018; Bing et al. 2021). Labs have shown how extremely thermophilic fermentation temperatures (>70°C) offer additional specific advantages including reduced contamination risk and opportunities for novel product separations (Bing et al. 2022, 2023). Previously, C. bescii was metabolically engineered to produce ethanol, acetone, and various alcohols, but not yet at industrially relevant titers (Williams-Rhaesa et al. 2018; Straub et al. 2020; Rubinstein et al. 2020). As

such, the project is currently working to improve selectivity, yield, and titers of these products in *C. bescii* as well as looking at additional target products such as 2,3-butanediol and 3-hydroxypropionate.

Additional work involving the hyperthermophilic archaeon P. furiosus is also underway with the aim to leverage its high thermophily, efficient and established genetic system, and its unique CO<sub>2</sub>-fixing and energy-conservation enzymes (Hawkins et al. 2015; Keller et al. 2015, 2017). This work not only includes similar efforts to produce industrial chemicals (3-hydroxypropionate, 1-propanol, and ethanol), but also to leverage knowledge of C. bescii to engineer P. furiosus with (hemi)cellulases from C. bescii to enable growth on cellulose and xylan. Likewise, researchers are using enzymes from P. furiosus in C. bescii to improve production of target chemicals. Research is ongoing to engineer an NADPH regenerating soluble hydrogenase I from P. furiosus into C. bescii to increase redox factors for different target chemicals, such as 3-hydroxypropionate. Throughout all this work, system-wide metabolic and regulatory models for the organisms were created to evaluate, guide, assist, and optimize the production of target chemicals. The C. bescii models that were created previously continue to be updated, and P. furiosus models were created more recently as part of this project (Rodionov et al. 2021; Zhang et al. 2021).

Bing, R. G., et al. 2021. "Thermophilic Microbial Deconstruction and Conversion of Natural and Transgenic Lignocellulose," *Environmental Microbiology Reports* **13**, 272–93.

Bing, R. G., et al. 2022. "Plant Biomass Fermentation by the Extreme Thermophile *Caldicellulosiruptor bescii* for Co-Production of Green Hydrogen and Acetone: Technoeconomic Analysis," *Bioresource Technology* **348**, 126780.

Bing, R. G., et al. 2023. "Fermentative Conversion of Unpretreated Plant Biomass: A Thermophilic Threshold for Indigenous Microbial Growth," *Bioresource Technology* **367**, 128275.

Hawkins, A. B., et al. 2015. "Bioprocessing Analysis of *Pyrococcus furiosus* Strains Engineered for CO<sub>2</sub>-Based 3-hydroxypropionate Production," *Biotechnology and Bioengineering* **112**, 1533–43.

Keller, M. W., et al. 2015. "A Hybrid Synthetic Pathway for Butanol Production by a Hyperthermophilic Microbe," *Metabolic Engineering* **27**, 101–06.

Keller, M. W., et al. 2017. "Ethanol Production by the Hyperthermophilic Archaeon *Pyrococcus furiosus* by Expression of Bacterial Bifunctional Alcohol Dehydrogenases," *Microbial Biotechnology* **10**, 1535–45.

Rodionov, D. A., et al. 2021. "Transcriptional Regulation of Plant Biomass Degradation and Carbohydrate Utilization Genes in the Extreme Thermophile *Caldicellulosiruptor bescii,*" *mSystems* **6**, e01345-01320.

Rubinstein, G. M., et al. 2020. "Engineering the Cellulolytic Extreme Thermophile *Caldicellulosiruptor bescii* to Reduce Carboxylic Acids to Alcohols Using Plant Biomass as the Energy Source," Journal of Industrial Microbiology and Biotechnology 47, 585–97. Straub, C. T., et al. 2018. "Biotechnology of Extremely Thermophilic Archaea," *FEMS Microbiology Reviews* **42**, 543–78.

Straub, C. T., et al. 2020. "Metabolically Engineered *Caldicellulosir-uptor bescii* as a Platform for Producing Acetone and Hydrogen from Lignocellulose," *Biotechnology and Bioengineering* **117**, 3799–808.

Williams-Rhaesa, A. M., et al. 2018. "Engineering Redox-Balanced Ethanol Production in the Cellulolytic and Extremely Thermophilic Bacterium, *Caldicellulosiruptor bescii*," *Metabolic Engineering Communications* 7, e00073.

Zhang, K., et al. 2021. "Genome-Scale Metabolic Model of *Caldicellulosiruptor bescii* Reveals Optimal Metabolic Engineering Strategies for Bio-Based Chemical Production," *mSystems* **6**, e0135120.

**Funding Information:** This material is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program, Genomic Science program under Award No. DE-SC0022192.

#### **Early Career**

## Metabolome-Informed Proteome Imaging of Lignocellulose Decomposition by a Naturally Evolved Fungal Garden Microbial Consortium

Marija Velickovic<sup>1</sup>, Ruonan Wu<sup>1</sup>, Margaret W. Thairu<sup>2</sup>, Yuqian Gao<sup>1</sup>, Dusan Velickovic<sup>1</sup>, Carrie D. Nicora<sup>1</sup>, Jennifer E. Kyle<sup>1</sup>, Nathalie Munoz<sup>1</sup>, Chaevien S. Clendinen<sup>1</sup>, Aivett Bibao<sup>1</sup>, Priscila M. Lalli<sup>1</sup>, Kevin Zemaitis<sup>1</sup>, Rosalie K. Chu<sup>1</sup>, Daniel Orton<sup>1</sup>, Sarai Williams<sup>1</sup>, Ying Zhu<sup>1</sup>, Rui Zhao<sup>1</sup>, Matthew E. Monroe<sup>1</sup>, Ronald J. Moore<sup>1</sup>, Bobbie-Jo M. Webb-Robertson<sup>1</sup>, Lisa M. Bramer<sup>1</sup>, Cameron R. Currie<sup>2</sup>, Paul Piehowski<sup>1</sup>, **Kristin E. Burnum-Johnson<sup>1\*</sup>** (kristin.burnum-johnson@pnnl.gov)

#### <sup>1</sup>Pacific Northwest National Laboratory; <sup>2</sup>University of Wisconsin–Madison

**Goals:** The objective of this Early Career Research project is to gain transformative molecular-level insights into microbial lignocellulose deconstruction through comprehensive and informative review of underlying biological pathway data yielded by the integration of spatiotemporal multiomic measurements (i.e., proteomics, metabolomics, and lipidomics). One of this project's goals is to uncover the mechanisms that drive cooperative fungal-bacterial interactions that result in the degradation of lignocellulosic plant material in the leafcutter ant–fungal garden ecosystem. The project's approach will enrich the current knowledgebase needed for a predictive systems-level understanding of the fungal-bacterial metabolic and signaling interactions that occur during cellulose deconstruction in an efficient, natural ecosystem. The leafcutter ant fungal garden is known as a natural model system for efficient plant-matter degradation. The degradation processes are largely mediated by the symbiotic fungal and bacterial members within the complex microbial consortium. These symbiotic microbes with unique metabolic capabilities, however, are heterogeneously spatially organized in the sample. Previous mass spectrometry (MS) studies profiled molecules from bulk fungal garden samples; thereby, averaging the biological processes across the ecosystem and masking their spatial localization, biological origin, and molecular dynamics (Khadempour et al. 2021). To overcome this limitation, researchers performed microscale imaging across 12 µm-thick fungal garden serial sections by applying a metabolome-informed proteome imaging (MIPI) approach. This approach combines two spatial multiomics MS modalities that enable obtaining comprehensive molecular characterization across and through the fungal garden. Matrix-assisted laser desorption/ ionization (MALDI) imaging profiled metabolites with a spatial resolution of 50 µm and correlated morphologically unique features with metabolome profiles of interest (i.e., lignocellulose degradation). The identified regions of interest (ROIs) were selected for subsequent microdissection and microscale proteomic imaging using microPOTS (microdroplet processing in one pot for trace samples) with an integrated metaproteomic approach to detect metabolic activities and identify microbial community members.

Untargeted MALDI-Fourier-transform ion cyclotron resonance-MS imaging analysis revealed heterogeneous spatial distribution of various molecular features across the fungal garden sections. Researchers leveraged the METASPACE annotation platform to search against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and tentatively annotate 650 unique metabolites, which colocalized metabolomic signatures with distinct microscopic features. The MALDI images mapped the presence of phenylpropanoids, benzaldehydes, flavonoids, plant hormones, Krebs cycle compounds, sugars, amino acids, and other molecules that were produced by the complex community in the fungal garden. Differential relative abundance and accumulation of low-molecular-weight lignin products (e.g., coniferyl alcohol, coniferyl aldehyde, sinapoyl aldehyde, cinnamate, ferulate, caffeate, and vanillin) were observed indicating specific spatial patterns. Another observed spatial pattern colocalized with a unique ant winglike feature that was characterized mainly by primary metabolites, such as soluble sugars, amino acids, and fatty acids. Informed by the metabolomic specific features, the team selected the wing and three additional ROIs characterized as lignocellulose degradation hotspots for subsequent micro-POTS metaproteomics analyses. Selected ROIs and their biological replicates were dissected using a laser capture microdissection system and collected in individual wells of

the microPOTS chip. Peptides resulting from on-chip sample preparation were analyzed by liquid cryotomography (LC)-MS/MS analyses. For the metaproteomic analyses, a reference database was first curated from 50 million proteins of known members in the consortium that were grouped into >24 million clusters based on sequence similarity to annotate the high-resolution tandem MS spectra with stringent matching criteria. A total of 7,392 nonconservative and taxon-specific peptides that mapped to 2,239 unique protein clusters were detected, unveiling a complex community with relatively high representation of arthropod peptides (5,178) observed only in the wing ROI, while fungal peptides (1,825) and comparatively low abundant plant (552) and bacterial (47) peptides were localized in the other three ROIs mapped as lignocellulose degradation zones. Metaproteomics data revealed the presence of a fungal ligninolytic auxiliary enzyme and several fungal carbohydrate-active enzymes such as hemicellulases, cellulases, pectinases, and amylases in the lignocellulose degradation hotspots ROIs. The metabolic functions detected at the microscale provide more direct evidence that fungi cultivated by leafcutter ants such as Leucoagaricus gongylophorus degrade plant cell walls in the leafcutter ant garden ecosystem.

Leveraging MIPI capability to spatially profile a plethora of metabolites and peptides provided some molecular insights and understanding of species-specific activities in this multimember heterogenous ecosystem. Integration of MIPI data unraveled some of the processes in this complex ecosystem by reconstructing crucial parts of the lignocellulose decomposition pathways in distinct microscopic ROIs. MIPI enzyme-metabolites integration showed a strong correlation comparing abundance and spatial localization between two omics modalities. Mechanistic understanding of this symbiotic system can aid in the biological production of biofuel precursors and bioproducts from plant biomass. This novel MS micron-scale multiomic workflow can be applied to other complex and heterogenous biological systems to enhance the understanding of community member interactions and dynamics.

Khadempour, L., et al. 2021. "From Plants to Ants: Fungal Modification of Leaf Lipids for Nutrition and Communication in the Leaf-Cutter Ant Fungal Garden Ecosystem," *mSystems* 6(2). DOI:10.1128/mSystems.01307-20.

**Funding Information:** Pacific Northwest National Laboratory (PNNL) is a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy (DOE) under Contract DE-AC05-76RLO 1830. This program is supported by the DOE, Biological and Environmental Research (BER) Program under the Early Career Award

Program. A part of this work was performed in the Environmental Molecular Sciences Laboratory (EMSL), a DOE Office of Science user facility at PNNL.

## Unraveling Spatiotemporal and Physicochemical Constraints on Soil Viral Community Composition and Viral Particle Integrity

Anneliek M. ter Horst, Christian Santos-Medellín, Sara E. Geonczy, Jane D. Fudyma, **Joanne B. Emerson**\* (jbemerson@ucdavis.edu)

University of California–Davis

Goals: The overarching goal of this project is to assess and compare the contributions of active, infectious viruses and inert viral particles to biogeochemistry across diverse terrestrial ecosystems. Using a multiomics approach, the team seeks to establish spatiotemporal patterns in soil-viral community composition and activity linked to host carbon and nitrogen metabolism in grasslands, shrublands, woodlands, and wetlands. Leveraging a prescribed forest fire and a peatland temperature and atmospheric CO<sub>2</sub>-manipulation experiment also allows exploration into feedbacks between soil viruses and carbon dynamics in response to environmental change. Through laboratory experiments, researchers are investigating the chemical composition, fate, transport, and integrity of viral particles in soil. By integrating field and laboratory experiments across a variety of soil edaphic properties and spatiotemporal scales, this project is expanding understanding of the soil virosphere and its influence on carbon and nutrient cycling.

Viruses have been recognized as highly abundant but poorly characterized members of the soil microbiome. By infecting soil microbes, viruses likely have substantial impacts on terrestrial biogeochemical processes under their hosts' control. Viral particles (virions) may also play more direct roles in soil biogeochemical cycling as packets of carbon, nitrogen, and phosphorus. But the time scales and environmental conditions that determine virion infectivity, transport, and sorption to soil particles are unknown. This project uses a combination of field, laboratory, and computational approaches to distinguish between infective and inert virions and to assess their respective contributions to soil biogeochemical cycling.

Using a post-0.22  $\mu$ m viral size–fraction metagenomics (viromics) approach, researchers are exploring the conditions and temporal scales over which virions are produced, remain infective, and decay in soil. Research has shown that viromes can recover about 500 times more viral sequence than total metagenomes. At the start of this project, it was unknown whether viromes reflected recent or long past infections. Results suggest that soil viromes generally capture very recently active viral communities—particularly in moist soil—but can reflect earlier infections in less active communities (e.g., seasonally dry soils) and can be dominated by compromised (i.e., inert) viral particles after extreme temperature perturbations (e.g., heating to 90°C or freezing).

The project's interpretation that most soil viromes capture a short window of recent activity is consistent with repeated findings of highly divergent soil-viral communities over spatial distances as short as 1 m in nearly all habitats explored. Results from the first 2 years of this award suggested that soil-viral communities were so distinct by site on a regional scale that more localized-habitat comparisons would be more tractable (Durham et al. 2022). Seven distinct wetland habitats were thus sampled over a 0.6 km<sup>2</sup> area in the Bodega Bay Natural Reserve on the California Pacific Coast. Viral communities were most distinct by site, with few populations shared between sites. Secondarily, viral communities with similar habitat characteristics (e.g., plant community composition and/or salinity) were most similar. Although reducing the spatial area of the study and selecting seemingly similar habitats (all wetlands) improved resolution of viral ecological patterns, ongoing efforts are focused on further reducing complexity.

Centered on a highly spatiotemporally resolved viromic study of two habitats in the Jepson Prairie grassland (eight locations, 30 time points since November 2020), ongoing work seeks to unravel the relative contributions of space, time, habitat, and dispersal on patterns of soil viral community composition. Briefly, viromes at Jepson Prairie were most distinct by habitat (between mounds and their adjacent swales, defined by differences in topography, plant community composition, and hydrology), but viral community compositional patterns over time were different within each habitat. The four swales exhibited similar viral community successional patterns over time, likely reflecting greater mixing in the swale habitats via intermittent flooding, whereas the four mounds (which rise ~0.5 m above swales and never flood) were more distinct over space, reflecting dispersal limitation. Analysis of the full dataset of >300 viromes is ongoing.

To investigate physicochemical constraints on virion integrity and viral community composition, researchers are analyzing data from three burned habitats and a laboratory temperature-manipulation experiment. In shrublands and woodlands that burned during the dry season in the LNU Lightning Complex fires in August 2020, the team is characterizing viral community successional dynamics after fire. A prescribed burn in a mixed conifer forest in Spring 2021 was also leveraged to compare burned and unburned soil-viral communities. Preliminary results suggest that habitat and location differentiate viral community composition more than the impact of fire, but viral richness was lower post fire than in unburned or pre-fire soils. The degree of virion inactivation and the timing of viral community recovery post fire seem to depend on soil moisture and depth. Briefly, heating experiments are revealing that viral particle survival thresholds are similar to those known for bacteria, with a reduction in survival at 60°C and nearly complete removal of intact virions (with some inert, compromised virions remaining) at 90°C.

Together, these results are revealing that the relative importance of spatial distance (and dispersal), time, and environmental conditions in structuring viral communities varies. With substantial differences in environmental parameters, habitat seems to trump all other factors at both local and global scales, but in local environments under similar conditions, space and time can be important. Results from this project are facilitating a better understanding of viral contributions to terrestrial biogeochemical cycling, both as dynamic components of soil organic matter and through their infection of hosts responsible for carbon and nutrient cycling.

Durham, D. M., et al. 2022. "Substantial Differences in Soil Viral Community Composition Within and Among Four Northern California Habitats," *ISME Communications* **2**, 100.

**Funding Information:** This research was supported by the U.S. Department of Energy's (DOE) Biological and Environmental Research (BER) Program under grant No. DE-SC0021198.

## Metagenomic Insights into Microbial Traits Influencing Community Dynamics and Nitrogen Cycling Post Fire

Amelia R. Nelson<sup>1\*</sup> (amelia.nelson@colostate.edu), Michael J. Wilkins<sup>1</sup>, Steven D. Allison<sup>2</sup>, Joanne B. Emerson<sup>3</sup>, Peter M. Homyak<sup>4</sup>, **Sydney I. Glassman**<sup>4</sup>

<sup>1</sup>Colorado State University; <sup>2</sup>University of California–Irvine; <sup>3</sup>University of California–Davis; <sup>4</sup>University of California–Riverside

**Goals:** Climate change coupled to shifting land-use patterns have increased the frequency and size of wildfires across the globe. These wildfire disturbances deplete soil-microbial biomass and alter the community composition of the soil microbiome, which drives critical terrestrial-biogeochemical cycling across ecosystems. Through this project, the team aims to couple field-derived metagenomic datasets, biogeochemical data via experimental pyrocosms, and ecosystem models to evaluate the impact of wildfire on microbially mediated nitrogen (N) cycling across ecosystems (e.g., Mediterranean grasslands, chaparral shrublands, coniferous forests).

Wildfires are increasing in both frequency and severity with climate change. They reduce soil microbial biomass and alter the community composition of the soil microbiome, selecting for pyrophilous taxa with encoded traits that enable them to thrive in burned soil. The soil microbiome plays a vital role in biogeochemical cycling and ecosystem function and is an important player in terrestrial N cycling, but it is poorly understood how the altered post-fire soilmicrobiome community composition influences microbially mediated soil N cycling and subsequent emissions of greenhouse gasses (GHGs) like nitrous oxide from post-fire ecosystems. Multiomics (i.e., metagenomics and metatranscriptomics) data allows researchers to infer lifestyle traits (i.e., Grimes' C-S-R framework) and function of pyrophilous taxa in post-fire soils. Through this project, the team has compiled an extensive multiomic dataset including 108 metagenomes and 12 metatranscriptomes from fireimpacted Colorado coniferous forests representing different burn severities (low and high severity) and across time (60year chronosequence to 1-year post-fire). These sequencing efforts, totaling nearly 9 Tb of data, have resulted in 1651 metagenome-assembled genomes that span the Actinobac*teria* (*n* = 861 metagenome assembled genomes), Proteobacteria (n = 315), Acidobacteria (n = 115), along with 17 other bacterial phyla. Further, putative pyrophilous taxa from previous studies are represented, including the Actinobacteria Arthrobacter (n = 14) and Blastococcus (n = 11), and Proteobacteria *Massilia* (n = 9). Data from 1 year post fire (Nelson et al. 2022) revealed that pyrophilous traits (e.g., fast growth, heat resistance, ability to use pyrogenic carbon) were critical in the post-fire soil microbiome, with their importance increasing with increased burn severity. Further, the dominance of MAGs exhibiting these traits was coupled to the loss of N-cycling functions, including the absence of evidence for the expression of the bacterial gene catalyzing N fixation (nifH) and loss of both nitrifying taxa (Nitrospira) and genes (amoA and nxrAB) in severely burned surface soils. Further analyses on MAGs derived from longer-term studies (3, 5, and 11 years post fire to 6 decades post fire) and other ecosystems (i.e., California grasslands and chaparral shrublands) will reveal whether these shortterm influences on N cycling are unique to Colorado coniferous forests and if they recede with time following burning. Combined, these datasets will shed light on the impact of wildfire on ecosystem N losses and the emission of GHGs from wildfire-impacted landscapes.

Nelson, R., et al. 2022. "Wildfire-Dependent Changes in Soil Microbiome Diversity and Function," *Nature Microbiology* 7, 1419–30.

**Funding Information:** This work is funded through Department of Energy (DOE) BER No. DE-SC0023127 "Predicting Post-Fire N Cycling Through Traits and Cross-Kingdom Interactions."

### Predicting Post-Fire Nitrogen Cycling through Traits and Cross-Kingdom Interactions

Sydney I. Glassman<sup>1\*</sup>, Steven D. Allison<sup>2</sup>, Joanne B. Emerson<sup>3</sup>, Peter M. Homyak<sup>1</sup> Michael J. Wilkins<sup>4</sup>

<sup>1</sup>University of California—Riverside; <sup>2</sup>University of California—Irvine; <sup>3</sup>University of California—Davis; <sup>4</sup>Colorado State University

Goals: Wildfires are increasing in frequency, size, and severity across the globe. Unlike ecosystem disturbances that primarily impact vegetation, wildfires kill microbes, thereby dramatically altering the composition, function, and abundance of post-fire soil microbiomes, with downstream impacts on soil nitrogen (N) cycling. Despite widespread microbial mortality during fires, post-fire environments can also favor the growth of pyrophilous fire-loving microbes. Pyrophilous microbes have been documented in widespread post-fire habitats, yet their traits and impacts on ecosystem N losses remain largely uncharacterized. Here, researchers focus on how wildfire severity and pyrophilous microbial interactions regulate N cycling and the emission of greenhouse gases (GHG) like nitrous oxide, a powerful greenhouse gas with 300 times the warming potential of carbon dioxide, with implications for long-term ecosystem recovery, regional air quality, and Earth's climate. The overarching project goal is to answer the question: Do conserved genomic traits and cross-kingdom interactions drive post fire N cycling across ecosystems?

Understanding how microbial interactions and their traits govern N cycling is critical to forecasting post-fire soil N dynamics and ecosystem recovery. Using pyrophilous microbiomes as model systems, the team will scale across systems of increasing complexity, from individual genomes to more complex microbiomes to predict the impacts of wildfire disturbance on ecosystem N cycling with the DEcomposition Model of ENzymatic Traits (DEMENT). Across three ecosystems that are experiencing increased fire frequency (i.e., Mediterranean grasslands, chaparral shrublands, montane coniferous forests) the team asks: (1) How do microbial traits change during post-fire succession? (2) How does fire severity influence microbial succession and gene expression of N-cycling functions? (3) How do cross-kingdom interactions change during post-fire succession? (4) How do traits and interactions affect ecosystem N fates and cycling?

To answer these questions, the team will (1) identify putative pyrophilous traits across microbiota and biomes and cross-kingdom interactions among archaea, bacteria, fungi, and viruses that affect N-cycling genes and biogeochemistry using metagenomic datasets and culture-and microcosm-based assays; (2) test and refine the trait predictions by coupling microbiomes and N-cycling genes to N biogeochemistry via experimental pyrocosms to simulate soil heating under controlled and replicable conditions; and (3) scale up microbial traits and interactions to the ecosystem level by integrating the measurements with the trait-based DEMENT model. Microbiological insights assessed via genomics, metagenomics, metatranscriptomics, and viromics will be paired with biogeochemical approaches that combine traditional laboratory assays with isotopic approaches to track the production and consumption of substrates involved in GHG emissions. Experimental approaches and modeling will span microbial domains and diverse fire-impacted ecosystems to assess the generality of the results at broad spatial and temporal scales.

**Funding Information:** This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0023127.

## Impacts of Altered Climate on Microbial Growth and Nutrient Assimilation in an Ombitrophic Peat Bog

Sheryl L. Bell<sup>1\*</sup>, Amy E. Zimmerman<sup>1</sup>, Bram W. G. Stone<sup>1</sup>, Kaitlin R. Rempfert<sup>1</sup>, Victoria Monsaint-Queeney<sup>2</sup>, Michaela Hayer<sup>2</sup>, Steven J. Blazewicz<sup>3</sup>, Kirsten S. Hofmockel<sup>1</sup>, **Bruce A. Hungate<sup>2</sup>** 

<sup>1</sup>Pacific Northwest National Laboratory; <sup>2</sup>Northern Arizona University; <sup>3</sup>Lawrence Livermore National Laboratory

Goals: The work proposed here will integrate genomicsand isotope-enabled measurements of Growth Rate, growth Efficiency, and the stoichiometry of Essential Nutrients during growth (GREEN) omics. The overarching objective is to develop and apply omics approaches to investigate microbial community processes involved in nutrient cycling. The specific objectives of the proposed work are (1) to evaluate the microbial ecology of nutrient uptake, testing hypotheses about nutrient assimilation in response to temperature variation; (2) to evaluate the ecology of nutrient-use efficiency for soil microorganisms within a framework of ecological theory, and (3) to develop new isotope-enabled genomics and transcriptomics techniques that probe the microbial ecology of nutrient dissimilation. This work will push the frontier of isotope-enabled genomics by connecting quantitative stable-isotope probing (SIP) to ecological theory about nutrient assimilation, nutrient-use efficiency, metabolic efficiency, and by applying these tools to understand the basic biology and ecology of soil microorganisms and how they transform nutrients in the environment.

Numerous studies have examined overall shifts in microbial community composition and diversity metrics under increased temperature, while far fewer have considered the importance of elevated carbon dioxide (CO<sub>2</sub>) due to climate change. Standard sequencing methods to profile soil microbiomes cannot determine whether community shifts are driven by increases in select members, losses of others, or masked by large pools of relic DNA. Instead, researchers used quantitative SIP with <sup>18</sup>O-water to estimate the *in situ* growth rates of individual bacterial taxa under long-term elevated CO<sub>2</sub> and across a gradient of warming treatments in a northern Minnesota peat bog at the SPRUCE (Spruce and Peatland Responses Under Changing Environments) experiment, representing a particularly vulnerable terrestrial carbon reservoir.

The microbial communities at SPRUCE have been subjected to experimental conditions for more than four years, and therefore had ample time to acclimate to the treatments. Researchers found that a large proportion of bacterial taxa displayed little to no growth across temperatures under ambient CO<sub>2</sub> concentrations but faster growth under certain temperatures with elevated CO<sub>2</sub>, highlighting a strong interplay between warming and CO<sub>2</sub> concentrations. Growth responses of multiple taxa could be clustered into three response patterns under ambient CO<sub>2</sub>, and just two response patterns under elevated CO<sub>2</sub>. Elevated CO<sub>2</sub> shifted the temperature of maximum growth for the two dominant lineages of the peat microbiome. The temperature of maximum growth among Proteobacteria increased under elevated CO<sub>2</sub>, contrasting the Acidobacteria whose maximum growth temperature decreased. Researchers found support for phylogenetic conservation of growth patterns among Acidobacteria and Proteobacteria at approximately the genus-level under ambient but not elevated CO<sub>2</sub>. Among Proteobacteria, groups with tight plant associations such as Rhizobiales (N-fixing symbionts) exhibited enhanced growth only at elevated temperature and CO<sub>2</sub>, suggesting these microorganisms may benefit from increases in rhizodeposition. These results suggest that certain taxa may be predisposed for growth under altered climate conditions with a disproportionate influence on carbon cycling and peatland feedbacks to climate change.

Northern peatlands are also of interest due to their exceedingly low inorganic nitrogen (N) contents and reliance on organic N. As such researchers have additionally leveraged peat collected from SPRUCE to conduct a laboratory incubation to explore the interactive effects of N source and temperature on nutrient assimilation and growth. Researchers used multiple isotope tracers to characterize either microbial growth (<sup>18</sup>O-water) or N assimilation with three different substrates representing inorganic (<sup>15</sup>N-ammonium and <sup>15</sup>N-nitrate) and organic (<sup>15</sup>N-glutamate) sources. Metabolomics profiles using Fourier transform ion cyclotron resonance (FT-ICR) indicate significant similarities among unique metabolites across the N treatments relative to controls. N-amendment increased the richness of molecules classed as condensed hydrocarbons (less bioavailable) at 25°C while decreasing richness of protein-like (more bioavailable) molecules at 15°C relative to matched unamended controls. Further, the average metabolite pool contained less potential energy at 25°C across all N amendments as well as control soils. Preliminary results reveal a significant impact of both temperature and substrate at the bulk level on respiration and N assimilation into microbial biomass with respiration most strongly impacted by glutamate amendment but N-assimilation into microbial biomass highest under ammonium amendment. Ongoing work to generate taxon-specific growth rates and nutrient-use efficiencies will be used to clarify whether differences at the bulk level are driven by the differential responses of individual taxa, and stable-isotope probing (SIP)-metagenomes and SIP-proteomes will be leveraged to explore altered metabolism and nutrient allocation under the treatments.

Together these experiments reveal how individual microorganisms vary in their acquisition, use, and release of nutrients—attributes that directly impact the rate and fate of environmental nutrient transformations in a globally important ecosystem.

**Funding Information:** This research was supported by a grant from the Department of Energy's Biological Systems Science Division Program in Genomic Science (No. DE-SC0020172, FWP 77475). Research conducted at the Pacific Northwest National Laboratory was conducted under the auspices of U.S. Department of Energy Contract DE-AC05-76RL01830. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

## The Predictive Power of Phylogeny on Growth Rates in Soil Bacterial Communities

Jeth Walkup<sup>1\*</sup> (jgwalkup@mix.wvu.edu), Ember M. Morrissey<sup>1</sup>, Chansotheary Dang<sup>1</sup>, Rebecca L. Mau<sup>2</sup>, Michaela Hayer<sup>2</sup>, Egbert Schwartz<sup>2</sup>, Bram W. G. Stone<sup>3</sup>, Kirsten S. Hofmockel<sup>3</sup>, Benjamin J. Koch<sup>2</sup>, Alicia M. Purcell<sup>2,4</sup>, Jennifer Pett-Ridge<sup>5,6</sup>, Chao Wang<sup>7</sup>, **Bruce A. Hungate<sup>2</sup>** 

<sup>1</sup>West Virginia University; <sup>2</sup>Northern Arizona University; <sup>3</sup>Pacific Northwest National Laboratory; <sup>4</sup>Texas Tech University; <sup>5</sup>Lawrence Livermore National Laboratory; <sup>6</sup>University of California– Merced; <sup>7</sup>Chinese Academy of Sciences

**Goals:** Microorganisms are major engines of the land carbon cycle, responsible for influencing the composition and radiative properties of the atmosphere, and for both creating and consuming soil organic carbon, a resource that provides multiple ecosystem services, and, when lost, exacerbates climate change. This project investigates the interactions within microbial communities and between microbes and their environment that underpin these dual roles of microorganisms in creating and consuming soil carbon. Overarching objectives are to develop and apply omics approaches to investigate microbial-community processes involved in carbon and nutrient cycling, develop community and taxon-specific microbial controls over key biogeochemical processes in terrestrial environments, and test quantitative ecological and biogeochemical principles using omics data. This work aims to facilitate scaling of taxon-specific microbial data to connect the ecology of microorganisms with ecosystem level rates of carbon and nutrient cycling.

Predicting ecosystem function is critical to assess and mitigate the impacts of climate change. Quantitative predictions of microbially mediated ecosystem processes are typically uninformed by microbial biodiversity. Yet new tools allow the measurement of taxon-specific traits within natural microbial communities. There is mounting evidence of a phylogenetic signal in these traits, which may support prediction and microbiome-management frameworks. Researchers investigated phylogeny-based trait prediction using bacterial growth rates from soil communities in Arctic, boreal, temperate, and tropical ecosystems. Here, research shows that phylogeny predicts growth rates of soil bacteria, explaining up to 58% of the variation within an ecosystem. Despite limited overlap in community composition across these ecosystems, shared nodes in the phylogeny and ancestral-trait reconstruction allowed cross-ecosystem predictions, which showed that phylogenetic relationships can explain up to 38% of the variation in growth rates across biomes. Results suggest that shared evolutionary history creates similarity in the relative growth rates of related bacteria in the wild, allowing phylogeny-based predictions to explain a significant amount of the variation in taxon-specific functional traits, within and across ecosystems.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0020172 and DE-SC0016207. Work at LLNL was performed under the U.S. Department of Energy Contract DE-AC52-07NA27344 and Awards SCW1590 and SCW1679.

## Ultrasensitive Protein SIP to Quantify Activity and Substrate Uptake in Microbiomes with Stable Isotopes

**Manuel Kleiner**<sup>1\*</sup> (manuel\_kleiner@ncsu.edu), Angela Kouris<sup>2</sup>, Marlene Violette<sup>1</sup>, Grace D'Angelo<sup>1,3</sup>, Yihua Liu<sup>2,4</sup>, Abigail Korenek<sup>1</sup>, Nikola Tolić<sup>5</sup>, Timo Sachsenberg<sup>6</sup>, Janine McCalder<sup>2</sup>, Mary S. Lipton<sup>5</sup>, Marc Strous<sup>2</sup>

<sup>1</sup>North Carolina State University; <sup>2</sup>University of Calgary; <sup>3</sup>Max Planck Institute for Marine Microbiology; <sup>4</sup>Max Planck Institute for Biology Tübingen; <sup>5</sup>Environmental Molecular Science Laboratory; <sup>6</sup>University of Tübingen

https://sourceforge.net/projects/calis-p/

**Goals:** The project's goal is to use stable isotope probing (SIP) to address the question of how microbes and minerals make necromass that persists.

SIP approaches are a critical tool in microbiome research to determine associations between species and substrates, as well as the activity of species. The application of these approaches ranges from studying microbial communities important for global biogeochemical cycling to hostmicrobiota interactions in the intestinal tract. Current SIP approaches, such as DNA-SIP or nanoscale secondary ion mass spectrometry, allow researchers to analyze incorporation of stable isotopes with high coverage of taxa in a community and at the single-cell level, respectively, however they are limited in terms of sensitivity, resolution, or throughput.

The team has developed an ultrasensitive, high-throughput protein-based SIP approach (protein-SIP), which cuts cost for labeled substrates by 50-99% as compared to other SIP and protein-SIP approaches and thus enables isotope labeling experiments on much larger scales and with higher replication. The approach allows for the determination of isotope incorporation into microbiome members with species-level resolution using standard metaproteomics liquid chromatography-tandem mass spectrometry measurements. At the core of the approach are new algorithms to analyze the data, which have been implemented in an open-source software. Research demonstrates sensitivity, precision, and accuracy using bacterial cultures and mock communities with different labeling schemes. Furthermore, the team benchmarks the approach against two existing protein-SIP approaches and shows that in the low-labeling range, the team's approach is the most sensitive and accurate. Finally, researchers measure translational activity using <sup>18</sup>O heavy water labeling in a 63-species community derived from human fecal samples grown on media simulating two different diets. Activity could be quantified

on average for 27 species per sample, with nine species showing significantly higher activity on a high-protein diet, as compared to a high-fiber diet. Surprisingly, among the species with increased activity on high protein were several *Bacteroides* species known as fiber consumers. Apparently, protein supply is a critical consideration when assessing growth of intestinal microbes on fiber, including fiber-based prebiotics.

In conclusion, research demonstrates the protein-SIP approach allows for the ultrasensitive (0.01% to 10% label) detection of stable isotopes of elements found in proteins using standard metaproteomics data.

Kleiner, M., et al. 2023. "Ultra-Sensitive Isotope Probing to Quantify Activity and Substrate Assimilation in Microbiomes," *Microbiome* **11**, 24.

**Funding Information:** This work was supported by the U.S. Department of Energy under Grant number DE-SC0022996, by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R35GM138362, and the U.S. National Science Foundation grant OIA #1934844. The funders had no role in the design of the study, data generation, analysis, and interpretation of study results.

## Computational Tools for Multiomic Data Standardization and Integration to Represent Whole Microbial Communities

Kostas Konstantinidis<sup>1\*</sup>(kostas@ce.gatech.edu), Caitlin Petro<sup>1</sup>, Katherine Duchesneau<sup>1</sup>, Malak Tfaily<sup>2</sup>, Rachel Wilson<sup>3</sup>, Jeffrey P. Chanton<sup>3</sup>, Christopher W. Schadt<sup>4</sup>, Spencer Roth<sup>4</sup>, **Joel E. Kostka**<sup>1</sup>

<sup>1</sup>Georgia Institute of Technology; <sup>2</sup>University of Arizona; <sup>3</sup>Florida State University; <sup>4</sup>Oak Ridge National Laboratory

**Goals:** The goals of this research are to: (1) standardize methods for detecting the relative abundance of molecular features (e.g., genes, pathways or species) in various omics datasets (e.g., metagenomics, metatranscriptomics, metaproteomics and metabolomics) for use by the scientific community; (2) extend the previously developed dynamic mathematical models for water-based ecosystems by integrating the standardized data from goal 1 and additional omics data, such as metabolomics, as parameters of the model towards identifying microbe-microbe and microbeenvironment interactions within a microbial community; and (3) apply the advanced models to appropriate multiomic data from the DOE's Spruce and Peatland Responses Under Changing Environments (SPRUCE) project to provide insights into the microbial interaction networks that mediate belowground carbon cycling in these peatland soils

as well as how these interaction networks may be altered by climate change drivers (e.g., elevated temperature and CO<sub>2</sub>).

Microbial species, especially in soils, are engaged in incredibly complex interactions based on their physiological responses to the environment and chemical communication via a wide range of molecules in low concentrations. Deciphering the multidimensional causes and consequences of such interactions during environmental transitions is challenging because traditional methods reveal only the numerically dominant members of the community related to the flow of the major carbon and nitrogen sources, or are typically limited to static correlation networks of abundances that cannot encompass well the dynamic environment. A predictive understanding of how the functioning of soil (and other) ecosystems responds to future environmental perturbations is limited by the inability to elucidate the physiological interactions within complex soil microbial communities and the effect of the physicochemical environment on those interactions. Further, the identification of key microbial guilds (i.e., microbial groups of species that exploit the same resource(s) related to carbon turnover) remains essentially elusive.

To address these challenges, researchers have recently developed mathematical models that represent microbe-microbe and microbe-environment interactions within a community and can predict how these interactions change in the future when environmental parameters such as temperature or precipitation change (i.e., the models represent dynamic models of whole microbial communities). The mathematical models are based on the principle of the ecological Lotka-Volterra (LV) differential equations and require time-series omic data (Dam et al. 2016). Applications of these models to available metagenomic data from freshwater lakes led to a number of insights and predictions, some of which are supported by biological evidence. For instance, an interaction cluster was identified in Lake Lanier (Atlanta, GA) that included cyanobacterial primary producers and proteobacterial heterotrophs that live on the exudates of the cyanobacteria (Dam et al. 2020). Notably, ~46% of all species-species interactions in Lake Lanier were negative indicating competition, while others were positive suggesting cooperation. These results contrasted with those for Lake Mendota (Madison, WI), a lake that freezes in the wintertime that indicated a higher level (~66%) of competition, presumably driven by the fact that Lake Lanier experiences much milder weather fluctuations (Dam et al. 2020). Even though some of the findings may appear to be somewhat anticipated based on existing knowledge, it is important to note that the mathematical biologically agnostic approach is able to quantify these effects on population abundance dynamics and interactions, which is essential for forecasting future behavior. Therefore, the modeling framework provides a new strategy for integrating omics data to summarize

the functionality and species-species interactions within natural habitats, while the application of these models to soil data as part of this project represents a novel contribution. Researchers will report on the efforts to adapt these LV models to the soil multiomic data available from the DOE's SPRUCE project.

An essential part of mathematical modeling is the accurate estimation of in situ abundance of molecular features (e.g., genes, pathways or species). However, how to precisely measure the abundance of features in metagenomic or other omic datasets remains challenging because the available methods have not yet been standardized, and it is not clear how the data from different approaches can be compared and interpolated. Estimation of *in situ* abundances is the cornerstone for several additional downstream analyses such as identifying differentially abundant taxa between samples, metabolic modeling, etc. Researchers will present the approaches to standardize the methods for detecting the relative abundance of features in various omics datasets (e.g., metagenomics, metatranscriptomics, metaproteomics and metabolomics) for use by the scientific community. As a representative example, researchers have recently advanced the tool for metagenomic (or metatranscriptomic) read recruitment plotting to provide precise estimates of whole-genome and individual gene abundances, and the extent of intra-population gene-content and sequence diversity (Gerhardt et al. 2021). Further, by analyzing publicly available genome and metagenome data, researchers show that the diversity within species is organized in distinct 99.5% Average Nucleotide Identity (ANI) clusters that can be used to consistently describe genomovars and strains (Rodriguez-R et al. 2022). Using these standards and concepts, researchers will also present the efforts to quantify how individual species and strains within species respond to the temperature and CO<sub>2</sub> treatments applied at the DOE's SPRUCE project.

Dam, P., et al. 2016. "Dynamic Models of the Complex Microbial Metapopulation of Lake" *npj Systems Biology and Applications* **2**, 16007.

Dam, P., et al. 2020. "Model-based Comparisons of the Abundance Dynamics of Bacterial Communities in Two Lakes," *Scientific Reports* **10**, 2423.

Gerhardt, K., et al. 2021. "RecruitPlotEasy: An Advanced Read Recruitment Plot Tool for Assessing Metagenomic Population Abundance and Genetic Diversity," *Frontiers in Bioinformatics* 1, 826701.

Rodriguez-R, L. M., et al. 2022 "A Natural Definition for a Bacterial Strain and Clonal Complex," bioRxiv. DOI:10.1101/2022.06.27. 497766.

## Uncovering the Microbial Networks that Degrade Plant-Derived Phenolic Compounds and Their Role in Peatland Soil Carbon Sequestration: Revisiting the 'Enzyme Latch' Hypothesis

Joel E. Kostka<sup>1\*</sup> (joel.kostka@biology.gatech.edu), Kostas Konstantinidis<sup>1</sup>, Caitlin Petro<sup>1</sup>, Katherine Duchesneau<sup>1</sup>, Malak M. Tfaily<sup>2</sup>, Rachel M. Wilson<sup>3</sup>, Jeffrey P. Chanton<sup>3</sup>, Christopher W. Schadt<sup>4</sup>, Spencer Roth<sup>4</sup>

<sup>1</sup>Georgia Institute of Technology; <sup>2</sup>University of Arizona; <sup>3</sup>Florida State University; <sup>4</sup>Oak Ridge National Laboratory

Goals: The goal is to elucidate the fundamental principles driving physiology and metabolic exchange within microbial interaction networks that regulate the rate-limiting steps in soil organic matter (SOM) degradation, specifically the oxidation of phenolic compounds derived from lignocellulose and lignin-like polymers in carbon-rich peatlands and their role in the preservation of organic matter under anaerobic, water-saturated conditions. The project combines multiomics with advanced analytical chemistry to test the enzyme latch hypothesis and its response to climate change drivers. Field and laboratory investigations will be integrated to construct and calibrate a predictive framework that links specific microbial processes and interactions to the mechanisms driving the rate limiting steps of enzymatic SOM decomposition (phenolic compound oxidation, hydrolysis), SOM persistence, and greenhouse-gas production in peatland soils. The project leverages infrastructure and extensive datasets of DOE's Spruce and Peatland Responses Under Changing Environments (SPRUCE) in Marcell Experimental Forest.

Peatlands represent climate critical regions that cover only 3% of the Earth's land surface but store approximately 1/3of all soil carbon (C). The future role of peatlands in C sequestration remains uncertain and depends on the impact of global change-related perturbations on their C balance. Hypotheses driving the proposed research are: (1) Under flooded anoxic conditions, persistent, plant-derived compounds (lignocellulose and lignin phenols) act as bottlenecks to microbial SOM decomposition by binding and inhibiting microbial hydrolase enzymes (e.g. CAZymes, peptidases); thus their degradation is the rate limiting step in SOM decomposition. (2) Soil moisture content and O<sub>2</sub> availability along with SOM quality largely determine the functional diversity of heterotrophic microbes and metabolic pathways of lignocellulose and lignin degradation, which in turn regulate soil C storage through the enzyme latch. (3) Climate change drivers, which are expected to warm and dry out peatlands, will release the enzyme latch

and accelerate SOM decomposition by enhancing the oxidation of phenolic compounds and concomitantly stimulating hydrolase activity. (4) Conversely, warming induced shifts in plant species composition away from mosses and toward lignin-rich vascular plants (ericaceous shrubs) will act to bolster the enzyme latch, inhibiting microbial decomposition through the accumulation of plant-derived phenolic compounds.

The project leverages an unprecedented time series generated from S1 Bog at the SPRUCE site (MN, U.S.) including shotgun metagenomic profiles (131 metagenomes, 2.4 Tbp of sequences), amplicon sequencing, and physicalchemical-biological data from 2014-2022. 810 dereplicated metagenome-assembled genomes (MAGs) of prokaryotes have been obtained across all metagenome datasets, and short read recruitment against these MAGs reveals that they represent the majority of the sampled microbial communities at all peat depths. Taxonomic diversity is dominated by the Acidobacteria, which comprise half of the 10 most abundant MAGs, with some recovered genomes comprising up to ~15% of the total community. These organisms are known to be metabolically flexible and contain an abundance of genes that encode degradation of plant-derived polysaccharides under both aerobic and anaerobic conditions. Metagenomes were screened for phenol oxidase and peroxidase genes, revealing the metabolic potential for phenolic compound oxidation.

Plant-derived phenolic compounds implicated as inhibitory compounds in the latch mechanism, including sphagnum acid, are persistent in surface soils. These compounds are supplied at the surface by plant litter but appear resistant to microbial decay as evidenced by their accumulation at depths down to 200 cm. Using microcosm experiments researchers quantified the inhibitory effect of soluble phenolics on anaerobic C mineralization and linked this effect to soil organic matter quality and peatland type. By manipulating the concentration of free soluble phenolics with polyvinylpyrrolidone (PVP), a compound that binds and inactivates phenolics, thereby preventing phenolic-enzyme interactions, rates of  $CO_2$  and  $CH_4$  production in soils were shown to be 62% and 54% inhibited by naturally occurring porewater phenolics, respectively.

Phenol oxidase and hydrolase activities were profiled with depth and between whole-ecosystem warming treatments in soils of the SPRUCE enclosures. At the *in situ* pH of 4 and room temperature, all measured enzyme activities (phenol oxidase, b-glucosidase, cellobiohydrolase, b-Nacetylglucosaminidase, acid phosphatase) declined with peat depth. A generational drought occurred in 2021, which provided intriguing evidence for the complex controls of the enzyme latch mechanism. Phenol oxidase and hydrolase activities were inversely correlated with water table elevation, which dropped by 0.11 m during the drought and declined with whole ecosystem warming. Thus, researchers hypothesize that enzyme activity is enhanced by putative oxygenation due to drought and higher evapotranspiration rates in the warming treatments.

The recent metabolomic observations from a range of peatlands indicate that the decomposability of peat SOM, and conversely C sequestration, is directly correlated with carbohydrate content (more reactive C substrates) and inversely correlated with aromatic content (recalcitrant C that resists degradation). Notably, Sphagnum-dominated peat soils are outliers, exhibiting a high proportion of labile substrates (carbohydrate content), but low GHG production rates. A range of site-specific factors are likely to impact enzyme-phenol interactions, including biotic (functional diversity of microbial communities, vegetation) and abiotic (temperature, pH) parameters. While the investigations to date suggest that temperature limits the latch mechanism over other factors, soluble phenolics from Sphagnum mosses appear to be especially effective at limiting decomposition under anoxic conditions.

**Funding Information**: This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0023297.

## Direct Routes for Microbial Carbon Stabilization of Volatile Organic Compounds in Soil

Juliana Gil-Loaiza<sup>1\*</sup> (juligil@arizona.edu), Parker Geffre<sup>1</sup>, Jordan Krechmer<sup>2</sup>, Malak M. Tfaily<sup>1</sup>, Linnea Honeker<sup>1</sup>, J. Rob Roscioli<sup>2</sup>, Joanne Shorter<sup>2</sup>, Elizabeth Lunny<sup>2</sup>, Manjula Canagaratna<sup>2</sup>, Megan Claflin<sup>2</sup>, Joseph Palmo<sup>3</sup>, David Hagan<sup>3</sup>, Eben Cross<sup>3</sup>, Kolby J. Jardine<sup>4</sup>, Romy Chakraborty<sup>4</sup>, Eoin Brodie<sup>4</sup>, **Laura K. Meredith**<sup>1</sup>

<sup>1</sup>University of Arizona; <sup>2</sup>Aerodyne Research, Inc.; <sup>3</sup>QuantAQ, Inc.; <sup>4</sup>Lawrence Berkeley National Laboratory

**Goals:** The overarching project goal is to verify and quantify volatile organic compounds (VOCs) as direct and indirect contributors to soil carbon (C) stabilization within the rhizosphere and beyond through teleconnections and to determine their underpinning ecological and metabolic mechanisms (See Fig. 16, p. 125). The long-term motivation for this project is to transform the current conceptual understanding and predictive capacity of microbial systems and soil C stabilization to include the important roles of volatile compounds. This presentation falls under the objective to determine the direct pathways and contributions of root-released VOCs and VOC transformations by soil microbiomes to soil C cycling and stabilization. Specifically,

in this task, researchers will aim to identify VOC-consuming microbes and traits and identify pathway(s) for VOC-C stabilization in soil pools using soil incubations and time-resolved <sup>13</sup>C-VOC stable isotope labeling.

VOCs are ubiquitous carbon pools in the Earth system but often remain uncharacterized as vectors of soil organic C transformations. Roots, litter, aboveground vegetation, and microbial metabolism are all sources of VOCs; however, little is known about how these omnipresent metabolites (Honeker et al. 2021; Meredith and Tfaily 2022) can contribute to C cycling in soils. In this project, researchers aim to verify and quantify the direct contributions of VOCs to soil C pools and determine their underpinning ecological and metabolic mechanisms. Here, researchers present the efforts to identify and quantify VOC uptake rates in soil, identify VOC-consuming microbes and traits, and determine whether the VOC uptake contributes to soil C pools. In a first experiment, the team exposed soil columns to controlled amounts of three VOCs (isoprene, methyl vinyl ketone, ethanol) below the soil surface to mimic root emissions. By monitoring subsurface VOC concentrations using novel in situ diffusive soil probes (Roscioli et al. 2021; Gil-Loaiza et al. 2022) connected to a proton transfer reaction time of flight mass spectrometer, researchers identified different uptake capacities for the three VOCs. Furthermore, the team discovered that microbial communities appeared to dramatically enhance their isoprene-uptake rates in response to isoprene exposure, revealing that microbial VOC metabolism can acclimate to resource availability. By tracking shifts in community composition and the isoprene monooxygenase gene copy number in these experiments, the aim is to identify the responsible microbes and the metabolic pathways involved. In a second experiment, by characterizing shifts in soil C concentration and composition (Fouriertransform ion cyclotron resonance mass spectrometry) in response to additions of different VOCs (methanol, acetaldehyde, acetone, isoprene, and monoterpene *a*-pinene), researchers will test the hypothesis that VOCs can have a direct impact on soil carbon composition. Researchers will use these results to refine plans for subsequent <sup>13</sup>C isotopelabeling experiments to track VOC-C into soil C pools and microbial groups. These experiments will provide new understanding of the direct impacts of VOCs on soil C and the importance to the conceptual models of soil C.

Gil-Loaiza, J., et al. 2022. "Versatile Soil Gas Concentration and Isotope Monitoring: Optimization and Integration of Novel Soil Gas Probes with Online Trace Gas Detection," *Biogeosciences* **19**(1), 165–85.

Honeker, L. K., et al. 2021. "The Volatilome: A Vital Piece of the Complete Soil Metabolome," *Frontiers in Environmental Science* **23**(9).

Meredith, L. K., and M. M. Tfaily 2022. "Capturing the Microbial Volatilome: An Often Overlooked 'Ome'," *Trends in Microbiology* **30**(7), 622–31.

Roscioli, J. R., et al. 2021. "Soil Gas Probes for Monitoring Trace Gas Messengers of Microbial," *Scientific Reports* **11**(1), 8327.





**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant Nos. DE-SC0023189 and the NSF Signals in the Soil program, grant no. 203419. Part of this work was also performed at Lawrence Berkeley National Laboratory supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Contract No. DE-AC02-05CH11231.

### Root-Mediated Impacts of Plant Volatile Organic Compound Emissions on Soil Carbon

Linnea Honeker<sup>1\*</sup> (linneah@arizona.edu), Holly Andrews<sup>1</sup>, Manjula Canagaratna<sup>2</sup>, Elizabeth Lunny<sup>2</sup>, Juliana Gil-Loaiza<sup>1</sup>, Malak M. Tfaily<sup>1</sup>, Jordan Krechmer<sup>2</sup>, Megan Claflin<sup>2</sup>, J. Rob Roscioli<sup>2</sup>, Joanne Shorter<sup>2</sup>, Duke Pauli<sup>1</sup>, Jeffrey Demieville<sup>1</sup>, Meara Clark<sup>1</sup>, Justin Grigory<sup>1</sup>, Robert Stanley<sup>3</sup>, Dusan Velickovic<sup>3</sup>, David Hoyt<sup>3</sup>, Robert Young<sup>3</sup>, Chaevien S. Clendinen<sup>3</sup>, Kolby J. Jardine<sup>4</sup>, Romy Chakraborty<sup>4</sup>, Margaret Torn<sup>4</sup>, Eoin Brodie<sup>4</sup>, **Laura K. Meredith<sup>1</sup>** 

<sup>1</sup>University of Arizona; <sup>2</sup>Aerodyne Research, Inc.; <sup>3</sup>Pacific Northwest National Laboratory; <sup>4</sup>Lawrence Berkeley National Laboratory

**Goals:** The overarching project goal is to verify and quantify volatile organic compounds (VOCs) as direct and indirect

contributors to soil carbon (C) stabilization within the rhizosphere and beyond through teleconnections in the soil matrix, and to determine their underpinning ecological and metabolic mechanisms (See Fig. 16, this page). The long-term motivation for the project is to transform the current conceptual understanding and predictive capacity of microbial systems and soil C stabilization to include the important roles of volatile compounds. This presentation falls under the objective to determine the direct pathways and contributions of root-released VOCs and VOC transformations by soil microbiomes to soil C cycling and stabilization. Specifically, in this task, researchers will aim to identify VOC-consuming microbes and traits and identify pathway(s) for VOC-C stabilization in soil pools using soil incubations and time-resolved <sup>13</sup>C-VOC stable isotope labeling.

Plants are recognized as the dominant source of VOCs to the atmosphere, where they play critical roles in air quality and climate, yet the parallel impact of plant-derived VOCs on the pedosphere (soil) remains poorly quantified. VOCs released by decomposing litter can contribute to soil C pools including those associated with soil C stabilization, and researchers hypothesize that root VOCs can also contribute to these soil C pools. Furthermore, researchers anticipate that this pathway for soil C stabilization will depend on plant physiological traits, rhizosphere microbes and their activity, and soil environmental factors. While the composition and flux of root-soil emissions of VOCs in the rhizosphere remain poorly quantified due to the lack of developed methods, here the team characterized the belowground release of plant VOCs to the soil system in two distinct ecosystems. Beneath the bioenergy-relevant crop sorghum in a semi-arid agroecosystem at the Maricopa Agriculture Center in Arizona, the team quantified soil VOC concentrations (proton transfer reaction time of flight mass spectrometer interfaced with novel in situ diffusion gas probes; Roscioli et al. 2021; Gil-Loaiza et al. 2022) and root metabolites (nuclear magnetic resonance, NMR) to identify root VOCs and estimate their potential *in situ* uptake rates in soil. In addition, the team leveraged an experimental soil warming treatment in the temperate coniferous Blodgett Experimental Forest in the Sierra Foothills in CA to explore associations between regions of root growth, increased VOC concentrations (NMR), and enhanced soil organic carbon (SOC) stocks. In bulk surface soils, researchers observed increases in both SOC stock and VOC concentrations including ethanol and methanol. Together, these field observations help to begin to define the contributions of root VOCs to soil C stabilization. They also serve as a reference for upcoming experimental tests of the impact of plant traits and environmental conditions on this under-resolved C transformation pathway.

Gil-Loaiza, J., et al. 2022. "Versatile Soil Gas Concentration and Isotope Monitoring: Optimization and Integration of Novel Soil Gas Probes with Online Trace Gas Detection," *Biogeosciences* **19**(1), 165–85.

Roscioli, J. R., et al. 2021. "Soil Gas Probes for Monitoring Trace Gas Messengers of Microbial," *Scientific Reports* **11**(1), 8327.

**Funding Information**: This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant nos. DE-SC0023189 and DE-SC0018459, A portion of this research was performed on a project award 10.46936/fics. proj.2021.60045/60000398 under the FICUS program and used resources at the DOE Environmental Molecular Sciences Laboratory, a DOE Office of Science User Facility sponsored by the BER program and operated under Contract No. DE-AC05-76RL01830. Part of this work was also performed at Lawrence Berkeley National Laboratory supported by the U.S. Department of Energy, Office of Science, Office of BER under Contract No. DE-AC02-05CH11231.

## Fungal-Bacterial Interactions: Bridging Soil Niches in Regulating Carbon and Nitrogen Processes

**Nhu H. Nguyen**<sup>1\*</sup>, Jonathan Deenik<sup>1</sup>, Jennifer Pett-Ridge<sup>2</sup>, Mengting Yuan<sup>3</sup>, Jizhong Zhou<sup>4</sup>, Siyang Jian<sup>4</sup>, Tai Maaz<sup>1</sup>, Katerina Estera-Molina<sup>2,3</sup>

<sup>1</sup>University of Hawai'i–Mānoa; <sup>2</sup>Lawrence Livermore National Laboratory; <sup>3</sup>University of California–Berkeley; <sup>4</sup>University of Oklahoma

Goals: As dominant groups of soil microbes, fungi, and bacteria together drive essential biogeochemical cycles belowground. However, the dynamics, mechanisms, and ecological implications of bacterial-fungal interactions (BFIs) are poorly understood, especially on the community level and under abiotic stress. The goal is to build a quantitative and mechanistic framework to address how BFIs determine the availability and fate of carbon (C) and nitrogen (N) across the complexity of soil niches. The three interrelated objectives each addresses a critical factor determining the behavior of BFIs across soil C source, mineralogy, and water availability: (1) measure how grassland BFIs are shaped by the availability of different C source, and in turn mediate soil C and N mineralization; (2) determine how BFIs may mediate C stabilization and mineralization via aggregation and mineral surface interaction across soils of different mineralogies; and (3) quantify how reduced water availability interplays with C source, C availability, and soil mineralogy in structuring BFIs and BFI-mediated soil processes.

Fungi and bacteria are the two dominant groups of soil organisms that consume, process, and translocate plantderived organic matter and thus are critical to global nutrient cycling. Fungal hyphal networks are important gateways for C and nutrient exchanges between plants and soils, and there is an increasing recognition that such processes are co-mediated by their interactions with bacteria. Yet the understanding of these interactions has generally been correlative, and the mechanisms of these interactions in the context of nutrient cycling are far from understood. The team hypothesize that bacterial-fungal interactions (BFIs) fundamentally determine the outcomes of soil ecosystem function by enabling C and N mineralization, competing for limited nutrients, and contributing to soil organo-mineral interactions and aggregate formation. Through this project, researchers aim to build a quantitative and mechanistic framework of how BFIs can change soil processes, the availability and the fate of C and N across the complexity of soil niches in different soil types and abiotic conditions. Researchers will present the initial concepts of this project through a set of three experiments: (1) Characterize how Mediterranean and tropical grasslands BFIs are shaped by

the availability of different C source, and quantify soil C and N mineralization mediated by the BFIs; (2) Understand whether and how BFIs mediate C stabilization and destabilization in soils of different mineralogies; and (3) Quantify how drought interplays with C source, C availability, and soil mineralogy in influencing BFIs and BFI-mediated soil processes.

Researchers propose three major tasks that scale in complexity, starting with a field experiment where the team will segregate soil niche compartments by using ingrowth cores that sequentially exclude incoming roots and fungi. This allows researchers to trace <sup>13</sup>C in isotopically labeled photosynthate as it moves through these niche compartments. Next, in a field-based mesocosm experiment, researchers will deploy the same ingrowth cores into intact megaliths of different soil types and use <sup>13</sup>C and <sup>15</sup>N tracers to measure how soil mineralogy interplay with microbial processes that influence the incorporation of these tracers into soil aggregates and mineral surfaces. Next, researchers will fine-tune the mechanisms that control BFIs in a laboratory soil microcosm experiment that measures the molecules involved when bacteria and fungi come into direct contact and the functional outcomes of these interactions. These three experiments will be performed under a drought treatment that can help address how microbial interactions can change with environmental conditions. Finally, researchers will integrate the data from these experiments using network analysis and omics-informed, niche-identified ecosystem (MEND) modeling.

The project will address the gap resulting from the oversimplified culture-based BFI studies and the mainly correlative field surveys of fungal-bacterial co-occurrence. Leveraging the above methods and technologies, researchers will be able to identify the mechanisms important to BFIs across broad grassland ecosystems, quantify C and N dynamics, model interactions in natural soil environments, and use this powerful set of data to better predict terrestrial C and N cycles under climate change.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0023106. Work at LLNL is performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

### Using <sup>13</sup>CO<sub>2</sub> to Track Carbon Flow Mediated by Fungal-Bacterial Interactions in Grassland Soils

Mengting Yuan<sup>1\*</sup> (maggieyuan@berkeley.edu), Katerina Estera-Molina<sup>1,2</sup>, Javier A. Ceja-Navarro<sup>3</sup>, Christina Fossum<sup>1</sup>, Anne Kakouridis<sup>1</sup>, Erin Nuccio<sup>2</sup>, Giovana Simão Slanzon<sup>4</sup>, Siyang Jian<sup>5</sup>, Zhifeng Yang<sup>5</sup>, Jizhong Zhou<sup>5</sup>, Jennifer Pett-Ridge<sup>2</sup>, Mary K. Firestone<sup>1</sup>, **Nhu H. Nguyen**<sup>4</sup>

<sup>1</sup>University of California–Berkeley; <sup>2</sup>Lawrence Livermore National Laboratory; <sup>3</sup>Northern Arizona University; <sup>4</sup>University of Hawai'i–Mānoa; <sup>5</sup>University of Oklahoma

Goals: As the dominant groups of soil microbes, fungi, and bacteria together drive essential biogeochemical cycles belowground. However, the dynamics, mechanisms, and ecological implications of bacterial-fungal interactions (BFIs) are not well-understood, especially on the community level and under abiotic stress. The broader goal is to build a quantitative and mechanistic framework to address how BFIs determine the availability and fate of carbon (C) and nitrogen (N) across the complexity of soil niches. The three interrelated objectives each addresses a critical factor determining the behavior of BFIs across soil C source, mineralogy, and water availability: (1) measure how grassland BFIs are shaped by the availability of different C source, and in turn mediate soil C and N mineralization; (2) determine how BFIs may mediate C stabilization and mineralization via aggregation and mineral surface interaction across soils of different mineralogies; and (3) quantify how reduced water availability interplays with C source, C availability, and soil mineralogy in structuring BFIs and BFI-mediated soil processes.

Fungi and bacteria are the two dominant groups of soil organisms that consume, process, and translocate plantderived soil organic matter (SOM) and thus are critical to global biogeochemical cycling. While fungi and bacteria's separate stereotypical processing of SOM are relatively well documented, there is an increasing recognition that niche-bridging fungal network and their interacting bacteria commonly co-mediate the flow and fates of plant-derived C. Less understood is how bacteria and fungi interact, and how these interactions change in response to environmental perturbations to influence the rate of C processes and fates of C in soil. The hyphosphere, soil explored by fungal hyphae, represent a potential hotspot niche of microbial activity that is just starting to be understood.

To characterize the interactions between fungi and bacteria in driving the flow of photosynthetic C belowground, researchers developed the Dynamic Ecosystem Labeling (DEL) system to deliver  $^{13}\mathrm{CO}_2$  to live plants, which

assimilated the <sup>13</sup>C and transferred it belowground via rhizodeposition. In a large-scale field manipulation study on the impacts of reduced precipitation on soil microbial interactions in a Mediterranean grassland, the team conducted four separate labeling events. Each labeling event differed in duration, between 5 to 14 days, and covered different phenological timepoints of the dominant grass Avena spp.: seedling, exponential growth, and peak biomass. During labeling, the DEL system tightly controlled the chamber headspace CO<sub>2</sub> concentration, driving the headspace average  ${}^{13}C$  atom% to 42.9 ± 13.3% and a maximum of 70.3%. A part of the <sup>13</sup>C incorporated into the ecosystem was quickly mineralized and detected as CO<sub>2</sub> over at least four weeks after labeling ended. Incredibly, 23% of the <sup>13</sup>C entered soil was still detectable after 2 years, three quarters of which were found to be associated with soil minerals. Sufficient stable isotope was incorporated into the belowground microbial communities such that labeled C could be followed into the DNA through quantitative Stable Isotope Probing (qSIP). Using amplicon sequencing of 16S and ITS genes coupled with qSIP, the DNA of nearly 150 taxa of bacteria and fungi were found to be significantly enriched with <sup>13</sup>C after 5 days of labeling, indicative of a potential food web that facilitates the bioprocessing and flow of rhizodeposits. Although the fungal and bacterial community composition and co-occurrence networks changed most profoundly over time, reduced precipitation significantly reduced the number of taxa that were <sup>13</sup>C enriched. This suggests that bacteria and fungi under reduced precipitation might have less access to newly fixed plant C. <sup>13</sup>C provides a powerful and sensitive tracer to follow and quantify the flow of C mediated by soil bacteria and fungi.

In a greenhouse experiment, the DEL system was used to label Avena barbata growing in mesocosms that had an air gap to separate the hyphosphere of arbuscular mycorrhizal fungi (AMF) Rhizophagus intraradices from the rhizosphere. The team quantified the amount of C being transported away from the rhizosphere by AMF hyphae crossing the air gap and studied the microbial composition associated with AMF hyphae and their <sup>13</sup>C enrichment. In 6 weeks, over 1% of the total soil C in the hyphal compartment was <sup>13</sup>C labeled, and a quarter of which was found associated with soil minerals. Amplicon sequencing indicated that AMF significantly modified the soil prokaryote community composition but not diversity; nineteen amplicon sequence variants significantly increased in the presence of AMF including Arthrobacter sp., Caulobacter sp., Rhizobium sp., Dongia sp., and Verrucomicrobia. Identification of the <sup>13</sup>C enriched taxa, which could be the primary consumers of <sup>13</sup>C imported via AMF hyphae, is underway.

Incorporating <sup>13</sup>C-informed C pool sizes and microbial activity into the Microbial ENzyme Decomposition (MEND) model improved the prediction of the

decomposition rates for different C pools. The next steps include (1) developing a BFI module for the MEND model; (2) using field-based fungal ingrowth core experiment and simplified soil interactions microcosm (SIM) experiments to derive the essential parameters that indicate the types and strengths of BFIs; and (3) assessing the soil C stability and storage mediated by the BFIs in different soil mineralogies and under reduced water availability.

**Funding Information**: This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0023106. Work at LLNL is performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

#### EARLY CAREER

## Crosstalk: Interkingdom Interactions in the Mycorrhizal Hyphosphere and Ramifications for Soil Carbon Cycling

**Erin Nuccio**<sup>1\*</sup> (nuccio1@llnl.gov), Jeffrey Kimbrel<sup>1</sup>, Megan Kan<sup>1</sup>, Edith Lai<sup>1</sup>, Eric Slessarev<sup>1</sup>, Kyungjin (KJ) Min<sup>2</sup>, Anne Kakouridis<sup>3</sup>, Jessica Wollard<sup>1</sup>, Marissa Lafler<sup>1</sup>, Peter Kim<sup>4</sup>, Trent R. Northen<sup>4</sup>, Karis McFarlane<sup>1</sup>, Mary K. Firestone<sup>3,4</sup>, Vanessa Brisson<sup>1</sup>, Jennifer Pett-Ridge<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory; <sup>2</sup>Seoul National University, South Korea; <sup>3</sup>University of California–Berkeley; <sup>4</sup>Lawrence Berkeley National Laboratory

Goals: Arbuscular mycorrhizal fungi (AMF) are ancient symbionts that form root associations with most plants. AMF play an important role in global nutrient and carbon cycles, and understanding their biology is crucial to predict how carbon is stored and released from soil. This DOE Early Career research investigates the mechanisms that underpin synergistic interactions between AMF and microbes that drive nitrogen and carbon cycling, addressing DOE's mission to understand and predict the roles of microbes in Earth's nutrient cycles. By coupling isotope-enabled technologies with next-generation DNA sequencing techniques, this project investigates soil microbial interactions in situ using natural levels of soil complexity. This work will provide a greater mechanistic understanding needed to determine how mycorrhizal fungi influence organic matter decomposition and will shed light on nutrient cycling processes in terrestrial ecosystems.

**Background:** The arbuscular mycorrhizal association between Glomeromycota fungi and land plants is ancient and widespread; 72% of all land plants form symbiotic associations with AMF. While AMF are obligate symbionts that depend on host plants for carbon (C) and cannot decompose soil organic matter (SOM), AMF can stimulate the decomposition of SOM and dead plant tissue. The team's prior research strongly suggests that AMF partner with their microbiome in the zone surrounding hyphae (or hyphosphere) to encourage decomposition. The molecular mechanisms that underpin interactions between AMF and the microbial community during nitrogen (N) uptake from SOM is a key knowledge gap. Researchers examine AMF-microbial interactions in both reduced complexity microcosms and in the field to assess the impact of AMF on terrestrial C and N cycling processes. In the field, team members assess how a deeply rooted perennial grass impacts soil C stocks and alters the zone of influence of AMF in soil depth profiles relative to shallow-rooted annuals.

Approach: AMF serve important roles in the soil microbial food web by stimulating SOM decomposition and providing plant C to the soil community. To identify the genomes of actively growing bacteria and archaea in the AMF hyphosphere, researchers tracked plant-fixed <sup>13</sup>CO<sub>2</sub> through AMF hyphae into the <sup>13</sup>C-hyphosphere microbiome using highthroughput stable isotope probing (HT-SIP) combined with high-resolution SIP-metagenomics (14 metagenomes per gradient). To separate the hyphosphere-C from the rhizosphere-C, the microcosms contained an airgap between plant and hyphal compartments that excluded roots but permitted fungal hyphae into living soil. SIP showed that the AMF Rhizophagus intraradices and associated metagenome assembled genomes (MAGs) were highly enriched (10-33 atom%<sup>13</sup>C), even though bulk soil enrichment was low (1.8 atom% <sup>13</sup>C). Of the 212 assembled <sup>13</sup>C-hyphosphere MAGs, the taxa that assimilated the most AMF-<sup>13</sup>C were from the phyla Myxococcota, Fibrobacterota, Verrucomicrobiota, and the ammonia-oxidizing archaeon genus Nitrososphaera. The phylogenetic composition and gene content of the highly <sup>13</sup>C-enriched MAGs highlight the potential for cross-kingdom trophic interactions in the AMF hyphosphere, including predation, decomposition of fungal necromass or plant detritus, and archaeal ammonia oxidation (that may utilize ammonium or CO<sub>2</sub> released from the aforementioned processes). In combination with other omics technologies, such as metatranscriptomics or proteomics, these MAGs will provide an important genomic resource for future experiments exploring interactions between AMF and their native microbiome.

To facilitate metabolomics and mechanistic studies of the hyphosphere, researchers have developed a sterile plant-mycorrhizal microcosm (called MycoChip, based off the EcoFAB platform) that researchers can use to interrogate hyphal-microbial interactions *in situ*. The MycoChip is intended to allow both destructive and nondestructive resampling of hyphosphere communities over time, and it is optically clear to permit microscopic investigation. This system has a raised airgap flanked by two 20  $\mu$ m mesh barriers to create a hyphosphere zone isolated from the rhizosphere, which permits hyphae to enter the hyphae chamber but blocks root entry. The raised airgap contains a dam that prevents solute exchange between chambers in either vertical or horizontal positions. In the most recent design, the team created larger rectangular chambers that stand upright and accommodate more experimental soil in both chambers. Researchers are testing these chambers so they can be used in future experimental studies.

Most knowledge about physiology and ecology of AMF (and most soil organisms) has been learned from surface soils that are less than 20 cm deep. In a national field study, researchers assessed how the rhizosphere of a deeprooted perennial bioenergy grass-switchgrass (Panicum *virgatum*)—impacts soil C stocks and alters the zone of influence of AMF and surface soil bacteria in depth profiles. Rhizosphere and bulk samples from paired switchgrass and shallow-rooted fields were collected from 2.5 m deep soil cores across nine field sites in the eastern United States (TX, MS, NC, NY, MI, WI, IL, SD). The team characterized the impact of switchgrass on soil microbial communities (AMF and bacteria), soil organic carbon (SOC), radiocarbon (<sup>14</sup>C), root abundance, and a range of soil physical and chemical properties. Switchgrass standing root biomass was significantly greater than annual standing root biomass. Across sites, radiocarbon and natural abundance <sup>13</sup>C data suggests that switchgrass-derived C was present to a 1 m depth and ~95% of the roots were in the top 1 m. Differences in the SOC stock were highly variable, but the effect of switchgrass on SOC was more consistently positive in southerly sites featuring Alamo. The team used amplicon sequencing to characterize the AMF and bacterial communities throughout the rooting depth profiles using WANDA and 16S primer sets, respectively; this analysis will show if deeply rooted switchgrass extends the habitat of AMF down the soil profile, thus increasing their zone of influence and contribution to subsoil C cycling and weathering processes. The project's C results indicate that standing root biomass may be a significant contributor to belowground C stocks in the first decades following conversion of annual cropland to perennial cover.

**Funding Information:** This research is supported by the U.S. Department of Energy Office of Science, Biological and Environmental Research Genomic Science Program under Early Career award SCW1711, DE-SC0016247 and DE-SC0020163 to UC Berkeley, and LLNL Laboratory Directed Research and Development grant 19-ERD-010. Work conducted at Lawrence Livermore National Laboratory was supported under the auspices of the U.S. DOE under Contract DE-AC52-07NA27344. Work at Lawrence Berkeley National Laboratory is supported by the m-CAFEs Microbial Community Analysis and Functional Evaluation

in Soils (m-CAFEs@lbl.gov) science focus area and was performed under the auspices of the U.S. Department of Energy Contract No. DE- AC02-05CH11231.

## Deciphering Virocell Metabolic Dynamics and Ecosystem Outputs Using a Novel Integration of Machine Learning and Metabolomics

Sumudu Rajakaruna<sup>1\*</sup> (rajakaruna2@arizona.edu), Marion Urvoy<sup>2</sup>, Roya AminiTabrizi<sup>1</sup>, Cristina Howard-Varona<sup>2</sup>, **Matthew B. Sullivan<sup>2</sup>**, **Malak M. Tfaily<sup>1,3</sup>** 

<sup>1</sup>University of Arizona, <sup>2</sup>The Ohio State University, <sup>3</sup>Pacific Northwest National Laboratory

**Goals:** The overarching goal of this project is to establish ecological paradigms for how viruses alter soil microbiomes and nutrient cycles by developing foundational (eco) systems biology approaches for soil viruses. Here, the team used metabolomics to investigate phage-specific metabolic reprogramming in virus-infected cells (virocells) to build critically needed model systems and *in silico* resources and tools that can be extended to new soil model phage-host systems. Researchers used an already established marine phagehost model system as the base to (1) characterize metabolic dynamics of virocell infection, (2) assess the output of virocell metabolic reprogramming on ecosystem function, and finally (3) develop the analytics that will be directly transferable to soil systems.

Microorganisms, including bacteria and viruses, play a vital role in biogeochemical cycles and global ecosystem function. At present, virus contributions are largely assessed through community-scale geochemical measurements or through evolutionary inferences such as identifying horizontal gene transfer. However, viruses also metabolically reprogram their bacterial host cells towards virion synthesis during infection, which effectively makes the infected cells (virocells) ecologically, metabolically, and physiologically different from uninfected bacteria. Here the team investigated the endo- and exometabolomes of an ecologically important marine heterotroph (Cellulophaga baltica strain #18), independently infected by three viruses with different morphologies, genomes and, therefore, infection strategies, to understand phage-specific virocell metabolic reprogramming and ecosystem outputs in a highly resolved infection time course. Through numerous ordination and multivariate statistical analyses, researchers show not only that a virocell's metabolite dynamics are different to that of an uninfected cell, but also that such metabolite dynamics differ temporally and between each virocell in ways that qualitatively associate with virus infection efficiencies, likely impacting ecosystem function. To assess this impact most

comprehensively, researchers are optimizing metabolite annotation—a major limitation in metabolomics—via testing and implementing combined machine learning and deep learning algorithms to obtain probabilistic annotations for unknown metabolites. These efforts have enabled a 24% increase in total annotations so far, which ultimately would significantly improve the biological interpretation of metabolomics results. Given that soil microbes can be infected by viruses at any given time, collectively these findings suggest that viruses can play an important role in regulating present and future carbon cycling.

**Funding Information:** This research is based upon work supported by the U.S. Department of Energy Office of Science, Biological and Environmental Research Program under Award Number DE-SC0020173 and DE-SC0023307 to the Ohio State University.

## Diversity Patterns and Ecological Footprint of RNA Viruses Along a Permafrost Thaw Gradient

Akbar Adjie Pratama<sup>1\*</sup>, Guillermo Dominguez-Huerta<sup>1\*</sup>, Ahmed A. Zayed<sup>1</sup>, James M. Wainaina<sup>1</sup>, Benjamin Bolduc<sup>1</sup>, Funing Tian<sup>1</sup>, Jared Ellenbogen<sup>2</sup>, Jiarong Guo<sup>1</sup>, EMERGE Biology Integration Institute Team<sup>1</sup>, EMERGE Coordinators<sup>1</sup>, Kelly C. Wrighton<sup>2</sup>, **Matthew B. Sullivan<sup>1</sup>** 

<sup>1</sup>The Ohio State University <sup>2</sup>Colorado State University

**Goals:** The overarching goal of this project is to establish ecological paradigms for how viruses alter soil microbiomes and nutrient cycles by developing foundational (eco)systems biology approaches for soil viruses. Specifically, this study aims to establish a fundamental knowledge of RNA viruses in permafrost regions and provide a comprehensive framework for understanding their complex interactions with hosts and the environment. To achieve this, the team will investigate the diversity and distribution of RNA viruses. In addition, to investigate their ecological impact, team members will examine the virus-host dynamics and evolution and identify the occurrence of auxiliary metabolic genes in RNA viruses. Ultimately, this research can significantly contribute to the integration of RNA viruses into the ecosystem and climate-related models.

As global atmospheric temperatures rise, sequestered permafrost soil carbon may be released via microorganismal activity, which can lead to a positive feedback loop effectively speeding up greenhouse gas emissions and climate change. While data are emerging about how prokaryotes and their DNA viruses may respond to permafrost thaw, the role of RNA viruses that infect soil eukaryotes remains less clear even though they may be important for nutrient cycling as demonstrated previously for the global oceans (Zayed et al. 2022; Dominguez-Huerta et al. 2022) and a diversity of forest, mountain, semi-desert, agricultural and sedimentary soils (Chen et al. 2022). Here team members leverage 55 metatranscriptomes (630 gigabases) collected along a permafrost thaw gradient in the model ecosystem Stordalen Mire to identify, quantify, and ecologically contextualize RNA viruses in these soils. Application of analytical approaches optimized for maximal sensitivity and largely automated systematic classification to these data identified 2,651 species-like RNA virus taxa. Though most of these species derive from the five known established phyla (Wolf et al. 2018), including one vOTUs in the recently suggested phylum Taraviricota (Zayed et al. 2022), nearly all represent novel species within these higher taxa and five likely represent a novel class in the phylum Lenarviricota (proposed name, Stomiviricetes). Ecological analyses of the approximately species-level taxa revealed strong habitat specificity as well as depth trends where RNA virus diversity decreased with depth. To assess the ecological footprint of Stordalen Mire RNA viruses, team members predicted hosts and evaluated gene content. This revealed that most (68%) likely infect key nutrient cycling eukaryotes that span multiple levels in the food web (only a few percent RNA viruses were predicted to infect prokaryotes), as well as 96 that carried virus-encoded auxiliary metabolic genes hinting at metabolic reprogramming of miscellaneous metabolic pathways, cellular and molecular processes, such as transport, transcription, countering antiviral defense, and coping to environmental stress. Together these findings provide baseline RNA virus diversity and ecology data and tantalizing hints at ecosystem impacts that establish fundamental knowledge for integrating them into predictive ecological models in the rapidly changing Arctic environment.

Chen, Y., et al. 2022. "RNA Viromes from Terrestrial Sites Across China Expand Environmental Viral Diversity," *Nature Microbiology* 7, 1312–23.

Dominguez-Huerta, G., et al. 2022. "Diversity and Ecological Footprint of Global Ocean RNA Viruses," *Science* **80**(1208), 1202–08.

Wolf, Y. I. et al. 2018. " Origins and Evolution of the Global RNA Virome," *MBio* **9**, 1–31.

Zayed, A. A. et al. 2022. "Cryptic and Abundant Marine Viruses at the Evolutionary Origins of Earth's RNA Virome," *Science* **80**(376), 156–62.

### Genetic Determinants of *Klebsiella* Phage Infection

Marissa R. Gittrich<sup>1\*</sup> (gittrich.1@osu.edu), Courtney M. Sanderson<sup>1</sup>, Cara M. Noel<sup>1</sup>, Jonathan E. Leopold<sup>1</sup>, Erica Babusci<sup>1</sup>, Olivia R. Farinas<sup>1</sup>, Aghiad Abadoul<sup>1</sup>, Steven Grabski<sup>1</sup>, Paul Hyman<sup>2</sup>, Vivek K. Mutalik<sup>3</sup>, **Matthew B. Sullivan<sup>1</sup>** 

<sup>1</sup>The Ohio State University; <sup>2</sup>Ashland University; <sup>3</sup>Lawrence Berkeley National Laboratory

**Goals**: The overarching goal of this project is to establish ecological paradigms for how viruses alter soil microbiomes and nutrient cycles by developing foundational (eco) systems biology approaches for soil viruses. Here the team seeks to understand how phages infect bacteria, specifically, what bacterial genes are required for infection to (1) see if researchers scan predict the bacterial genes required for phage infection based on phage sequences, and (2) use these data to generate models to understand how phages and bacteria interact in an ecological setting.

Novel bacteriophages (phages) are being cataloged at unprecedented rates, and current research broadly credits them with driving nutrient and energy cycling across many of Earth's ecosystems. However, little is known about the bacterial genes required for infection beyond a few model phage-host systems. Such data are critical for modeling phage-host interactions in complex communities. Here, the team mapped bacterial genetic determinants of phage infection using a randomly barcoded, genome-wide loss-offunction transposon mutant library (RB-TnSeq) of a plant growth-promoting rhizobacterium (*Klebsiella sp.* M5a1). This library was individually challenged by 25 diverse, double-stranded DNA phages spanning four known phage families at three multiplicities of infection (0.1, 1, and 10). The genetic screen uncovered a multitude of bacterial factors involved in phage infection, such as genes involved in receptor formation, transcription regulation, electron transport, and genes with unknown functions. When disrupted, some bacterial genes, such as those encoding putative glycosyltransferases involved in LPS biogenesis, conferred resistance to up to 50% of the phages across multiple phage families, potentially due to preventing phage adsorption. Other bacterial genes involved in intracellular functions, such as the electron transport chain and transcriptional regulation, were phage-specific, indicating that such cellular processes are differentially required across the diverse phage set. This supports previous findings in Escherichia coli that genes involved in intracellular functions are phage-specific, while genes encoding for receptors required for phage adsorption are more broadly required across phages. Additionally, the team found that some phages, although highly related, required a

unique set of bacterial genes for phage infection. Together these findings provide a foundation to develop predictive models of phage infection that can be applied to environmental systems.

**Funding Information**: This research is based upon work supported by the U.S. Department of Energy Office of Science, Biological and Environmental Research Program under award Nos. DE-SC0020173 and DE-SC0023307 to The Ohio State University. Marissa R. Gittrich was supported in part by the NIH T32 GM086252.

## From Bulk Organic Matter Profiling to Specific Metabolite Identification: Improving Metabolomics Data Analysis, Annotation, Interpretation, and Integration

Christian Ayala-Ortiz<sup>1\*</sup> (cayalaortiz@arizona.edu), Sumudu Rajakaruna<sup>1</sup>, Yuri E. Corilo<sup>2</sup>, Jordan Rabus<sup>2</sup>, Dalal Alharthi<sup>1</sup>, **Malak M. Tfaily<sup>1</sup>** 

<sup>1</sup>University of Arizona; <sup>2</sup>Environmental Molecular Sciences Laboratory (EMSL); Pacific Northwest National Laboratory (PNNL)

Goals: Over the past decade, advances in different omics technologies such as metagenomics, metatranscriptomics, metaproteomics, and metabolomics have revolutionized biological research by enabling high-throughput monitoring of biological processes at the molecular and organismal level and their responses to environmental perturbation. Metabolomics is a newer and fast-emerging technology in systems biology that aims to profile small compounds within a biological system that are often end products of complex biochemical cascades. Despite its importance, metabolomics has long been overshadowed by other omics. While metabolomics has not always been considered a standard tool in environmental and microbiome science, it can augment the power of genomics, transcriptomics, and proteomics by providing a functional snapshot of all upstream biological processes, thereby filling in gaps left behind by genomics and proteomics. The overarching goal of this project is to develop user-friendly and open-source tools (MetaboDirect and MetaboTandem) to optimize, streamline, and improve current data analysis pipelines for metabolomics data sets from complex and heterogeneous samples. All tools will be available for other researchers and practitioners who can replicate or extend the work.

Metabolites constitute the chemical currency of environments as they are used, transformed, and exchanged by the microorganisms found within a system. Understanding how different perturbations change the metabolome will allow improvement of the knowledge of the mechanisms used by the microbial communities that drive ecosystem functions. However, fully characterizing the metabolome is challenging due to its complexity and heterogeneity, often requiring multiple approaches with different objectives. One approach is the bulk characterization of metabolites or organic compounds in a system through the use Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). While the high-resolving power of this instrument allows for molecular-level characterization of organic matter from diverse environments, it generates hundreds of millions of data points that need to be processed and visualized. In response, the comprehensive, open-source, command-line based tool MetaboDirect was developed based upon years of analytical expertise with diverse sample types. The current version of MetaboDirect was published to be fully compatible with the output of the molecular formula assignation software, Formularity. Efforts are in place to integrate this pipeline with CoreMS, the comprehensive mass spectrometry framework being developed by EMSL at PNNL. Liquid chromatography tandem mass spectrometry (LC-MS/MS) is another approach that can be used for metabolome characterization of complex environmental samples that, unlike FT-ICR MS, is commonly used to get the true identity of the metabolite. However, LC-MS/MS suffers from additional challenges related to data processing and metabolite annotation and identification. Researchers are developing Metabo-Tandem as a package and a shiny app for the easy and fast analysis, annotation, and visualization of LC- MS/MS data that has a special emphasis on providing a comprehensive metabolite annotation that goes beyond the use of in-house databases to include a combination of publicly available databases and *in silico* prediction tools.

While MetaboDirect and MetaboTandem are being developed to address two challenges with metabolomics data (Challenge 1: Easy-to-use software; Challenge 2: Effective data visualization), team members acknowledge that the metabolomics field continues to struggle with two additional challenges (Challenge 3: Metabolite annotation and identification and Challenge 4: Big data). As such, researchers have been working on combining analytical chemistry, computer science (ML/AI), and statistics to develop bioinformatics tools that address these two additional challenges. As such, the team is not currently testing multiple machine learning–based algorithms to putatively identify unannotated metabolites. Accessing the information hiding in the unidentified metabolites will prove to be key to the effective integration of metabolomics with other omics data.

**Funding Information**: This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0021349.

## Gatekeepers of Arctic Carbon Loss: Landscape-Scale Metabolites-to-Ecosystems Profiling to Mechanistically Map Climate Feedbacks

McKenzie Kuhn<sup>1\*</sup>, Virginia Rich<sup>2</sup>, Gene Tyson<sup>3</sup>, Simon McIlroy<sup>3</sup>, Kelly C. Wrighton<sup>4</sup>, Malak M. Tfaily<sup>5</sup>, Scott Saleska<sup>5</sup>, Jeffrey P. Chanton<sup>6</sup>, Rachel M. Wilson<sup>6</sup>, Jinyun Tang<sup>7</sup>, **Ruth Varner**<sup>1</sup>

<sup>1</sup>University of New Hampshire; <sup>2</sup>The Ohio State University; <sup>3</sup>Queensland University of Technology; <sup>4</sup>Colorado State University; <sup>5</sup>University of Arizona; <sup>6</sup>Florida State University; <sup>7</sup>Lawrence Berkeley National Laboratory

The team proposes to identify and characterize microbes and metabolites critical for carbon (C) transformation in high-latitude interconnected terrestrial and aquatic sediment systems, which are undergoing rapid climate change. Discontinuous permafrost in these regions is rapidly thawing, and the potential climate feedbacks are substantial. These landscapes encompass various permafrost thaw stages and lesser-studied lakes, which can be the exit point for a significant fraction of methane (CH<sub>4</sub>) lost post-thaw via ebullition (bubbling) and are projected to increase with warming. At the same time, thaw-initiated succession of plant communities can increase net soil C storage and potentially also plant-derived inhibitory compounds that dampen C processing. Accurate predictive understanding of the net effect of these simultaneous coupled loss, gain, and stabilization processes, under increasing temperatures is an area of urgent study.

The team will leverage a model terrestrial ecosystem with site-specific genomes, metabolite spectra, C gas emissions and isotopes profiles—including lake sediments with anaerobic  $CH_4$  oxidizers present at some of the highest abundances ever observed in a natural system. The team will compare integrated substrate-microbiome-emissions in the two habitats (recently thawed terrestrial fens and lake sediments) which together dominate climate feedbacks (via >90% of landscape  $CH_4$  emissions), quantify rates and controls, and distill insights into models to more accurately predict C cycling and climate feedbacks.

## Construction of a Microbial Methane Observatory Reveals Metabolic Dynamics of Freshwater Wetland Microbiomes

Mikayla A. Borton<sup>1</sup>, Angela Oliverio<sup>1,2\*</sup>, Jared Ellenbogen<sup>1</sup>, Brooke Stemple<sup>1</sup>, Rory Flynn<sup>1</sup>, Bridget McGivern<sup>1</sup>, David Hoyt<sup>3</sup>, Jorge Villa<sup>4</sup>, Christopher S. Miller<sup>5</sup>, Gil Bohrer<sup>6</sup>, Timothy Morin<sup>7</sup>, **Kelly Wrighton**<sup>1</sup>

<sup>1</sup>Colorado State University; <sup>2</sup>Syracuse University; <sup>3</sup>Pacific Northwest National Laboratory; <sup>4</sup>University of Louisiana; <sup>5</sup>University of Colorado; <sup>6</sup>The Ohio State University; <sup>7</sup>State University of New York College of Environmental Science and Forestry

Goals: This project interrogates microbial contributions to carbon cycling in soils from a freshwater, coastal wetland adjacent to Lake Erie, OH. This site was selected as it has the highest annual methane fluxes within the AmeriFlux network. To profile the microbial contributions to methane release the team instrumented the wetland with in situ porewater and greenhouse gas (GHG) flux (i.e., carbon dioxide and methane) measurements and flux tower for site-wide flux measurements. A first goal of this project required construction of a microbial genome-resolved database generated to interrogate microbial physiological contributions to wetland GHG fluxes. Secondly, the team developed computational application in the annotation software Distilled Refined Annotation of Metabolism (DRAM) to profile the microbial traits in this genomic database. Lastly, the team mapped metatranscriptomes collected from temporal sampling and highly spatially resolved sampling to resolve expressed microbial community metabolic traits over years, within seasons, and along centimeter depth to meter land-coverage gradients. Leveraging the multiomics data in conjunction with geochemical, metabolomic, and GHG measurements, researchers provide unprecedented insight into how hydrological perturbations impact methane flux from a coastal freshwater wetland.

Freshwater wetlands contribute over a third of global methane emissions and store 30% of global soil carbon, which may become increasingly available for microbial degradation into GHGs. Despite their global climatic significance, it remains difficult to establish a mechanistic understanding of the microbial controls on soil biogeochemical processes, limiting the ability to predict GHG emissions. Here, researchers built a comprehensive catalog of 17,333 metagenome-assembled genomes representing 2,502 dereplicated genomes spanning 72 phyla from freshwater wetlands, of which 57% represent novel lineages with no genomic representation. The team then coupled its wetlands genome database with 133 genome-resolved metatranscriptomes and highly resolved compositional and geochemical profiling of over 700 samples to identify the dominant axes of environmental variation shaping the composition and transcription of microbial genomes in the wetlands system. Researchers further delineate the contributions of major microbial lineages to biogeochemically relevant processes including methanogenesis, sulfate reduction, and denitrification.

Although soil methanogens and methanotrophs have been cultivated for decades, the project's genome recovery approach resulted in 88 methanogen genomes representing all methanogenic pathways. These include three novel families, 17 novel genera, and for some taxa, the first genomes identified in a wetland environment-illuminating the phylogenetic and metabolic diversity harbored in terrestrial ecosystems. First, the team used nuclear magnetic resonance-metabolites to survey the distribution of methanogenic substrates across the wetland. With flooding, researchers observed increases in methanogen substrates of acetate and methanol only in the surface soils. Concomitant with increased availability of these substrates, the team observed gene activity from methanogens utilizing acetate and methanol in these surfaces increased 4.9 and 1.6-fold respectively with flooding. This expanded substrate availability, and concentrations likely contributed to the nearly 5-fold increase in methane production reported in the surface soils with flooding. Surprisingly, while researchers expected the shift to anoxia with flooding would decrease aerobic methanotrophy in surface soils, gene expression data indicated aerobic methanotrophs metabolized methane using very low oxygen concentrations. Team members also demonstrated the first gene expression data evidence for nitrate-enabled methanotrophy in soil systems. Given that methanol utilizing methanogens and anaerobic methanotrophs are currently not accounted for in climate models, the findings show important data add increased realism to biogeochemical models of terrestrial methane emissions.

Using co-expression network analysis, the team revealed coordinated, active microbial neighborhoods that are localized to depth but stable across months, seasons, and years. These microbial guilds are predictive of *in situ* GHG concentrations. Using integrated metabolite and metatranscriptome data researchers reconstructed the metabolic circuitry explaining a substantial fraction of variability in these GHGs. Findings link biogeochemical shifts to the genome transcription of specific microbial lineages and processes and begin to establish an integrative framework for leveraging high-dimensional multiomics data towards process-based models of the ecology and biogeochemistry of freshwater wetlands.

**Funding Information**: This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant No. DE-SC0023084. This program is supported by the U.S. Department of Energy,

Office of Science, through the Genomic Science Program, Biological and Environmental Research Program, under FWP ERKP123.

## Putting Microorganisms on the Map: Continental-Scale Context for Thousands of Newly Sampled Microbial Genomes from North American Wetlands

Emily Bechtold<sup>1</sup>, Jared Ellenbogen<sup>1</sup>, Mikayla A. Borton<sup>1</sup>, Bridget McGivern<sup>1</sup>, Djennyfer Ferreira<sup>1</sup>, David Hoyt<sup>2</sup>, Elisha Wood-Charlson<sup>3</sup>, William Riley<sup>4</sup>, Jorge Villa<sup>5</sup>, Christopher S. Henry<sup>3</sup>, Gil Bohrer<sup>6</sup>, Michael J. Wilkins<sup>1</sup>, **Kelly C. Wrighton<sup>1</sup>** 

<sup>1</sup>Colorado State University; <sup>2</sup>Pacific Northwest National Laboratory; <sup>3</sup>Argonne National Laboratory; <sup>4</sup>Lawrence Berkeley National Laboratory; <sup>5</sup>University of Louisiana; <sup>6</sup>The Ohio State University

Goals: Despite their vital roles in transforming nutrients and controlling greenhouse gas (GHG) fluxes in wetlands, microbial knowledge is often limited to taxonomic identity alone and rarely includes cross-site comparisons. The team proposes to address this knowledge gap using coordinated, reproducible field measurements collected across a wetland-methane continuum from at least 25 wetlands in the continental United States. The overarching project goal is to decode the unifying microbial properties governing soil carbon decomposition and methane fluxes, both within and among freshwater wetlands. This project tests the overarching hypothesis that microbial genomic attributes are conserved across high methane-emitting samples or wetlands, such that some level of biological representation into models will enhance predictions of soil methane fluxes. First, the team will use a cross-wetland approach to define the microbial membership, physiology, and interactions directly contributing to wetland methane production. Next, the team will uncover the microbial decomposition network features that classify high methane emitting wetlands. Using this information, the team will test the genomic resolution needed to make robust predictions of methane fluxes across regional and global models. These integrated field, laboratory, and modeling approaches will identify the biotic attributes conserved across methane-emitting wetlands, such that some level of biological representation into models will enhance predictions of soil methane fluxes.

Today wetland contributions to the global methane budget are estimated from ecosystem scale models. These models exclude representation of soil microbial metabolism or are based on incomplete or outdated knowledge on the physiological controllers on soil methane metabolism. Instead, these models use abiotic (e.g., temperature) and
indirect biotic (e.g., gross primary production) variables to approximate the environmental states enabling soil methane flux. However, years of observations from more than 40 freshwater wetlands showed that these variables only partially predicted annual methane fluxes. The deviation of the predictions from observations indicates predictions derived only from abiotic variables incompletely represented methane flux, especially for the highest-emitting wetlands. Team members posit that knowledge of methane-cycling microorganisms and their physiological networks will enhance freshwater wetland model predictions. In this research, the team will identify the microbial processes impacting wetland methane fluxes and evaluate their biogeographic conservation. Researchers will distill this content into an ecosystem model with the goal of closing the gap between measured and predicted methane fluxes from wetlands.

In the first year of this project, the team leveraged publicly available 16S amplicon data to begin addressing the biogeographical conservation of microbial community composition and metabolic function across wetlands. Researchers cataloged the microbial community membership from 1,118 wetland samples collected from nine geographically dispersed wetlands. Samples screened with 16S rRNA were also contextualized with soil chemistry and flux data, yet many sites also included genome-resolved metagenomics, transcriptomics, and metabolomics. The analysis shows that marshes, fens, and bogs have distinct microbial communities such that wetland type—more than geography, climate, or soil features drives microbial community composition. Despite sitewide differences at the community level, researchers did observe six genera of methanogens (e.g., Methanoregula, Methanoth*rix*) and four genera of methanotrophs (e.g., *Methylobacter*) were core across wetland samples. Notably, the three highest methane emitting wetlands, all marshes, shared methanogen and methanotroph membership and distribution patterns. In fact, the dominance of Methanoregula is a strong predictor of methane flux across wetlands. Researchers also show that wetlands with annual methane emission that deviate the most from temperature-based predictions had the highest Methanoregula relative abundance. Together this preliminary cross-wetland data illuminates the conservation

of microorganisms across high methane-emitting wetlands, narrowing the diverse soil carbon cycling community to a most-wanted list of fruitful targets for genomic and metabolic network efforts, yielding process-based knowledge needed for biologically aware models.

To begin to illuminate the metabolic features of carbon decomposition in high methane emitting wetlands, the team created the second version of the Genome Resolved Open Watershed (GROW2) database. This public, genomic resource contains the identity and distribution of 26,000 unique microbial genomes from wetlands, including over 500 methanogen and methanotroph genomes. This spatial sampling scheme coupled to a breadth of ecological dimensions (e.g., wetland type, methane emission rate, land use) will enable us to systematically identify the microorganisms and metabolic networks associated with high methane-emitting wetlands. As a preliminary approach, the team performed site-specific, co-occurrence analysis to uncover the network of microorganisms coordinated to methanogens across each site. Linking the 16S rRNA data to the project's GROW2 genomes, team members developed metabolic profiles for these methanogen connected taxa. Researchers show that obligate fermenters (e.g., many syntrophs) have the highest connectivity to the most abundant methanogens. Additionally, the highest methane-emitting wetlands had the least amount of methanogen connections, suggesting streamlined metabolic circuits may contribute to enhanced methane production across wetland soils. Ultimately, GROW2 is a living road map, articulating the power of microbiome science to decode microbial organismal and metabolic patterns at scales necessary for ingestion into predictive modeling frameworks. GROW2 is publicly available on KBase, engendering collaborative enterprises with the goal to advance a new era of climate-driven research in wetlands.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0023084. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP ERKP123.

## **SCIENCE FOCUS AREA**

## Bacterial-Fungal Interactions and Their Role in Soil Functioning

## An Online Public Resource for Bacterial-Fungal Interaction Research

Aaron Robinson, Julia Kelliher, Buck Hanson, Reid Longley, Demosthenes Morales, **Patrick Chain**\* (pchain@lanl.gov)

Los Alamos National Laboratory (LANL)

https://genomicscience.energy.gov/lanlbfi/

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

**Goals:** As part of the LANL science focus area (SFA) on Bacterial-Fungal Interactions (BFIs), researchers are developing novel bioinformatic and experimental tools and resources for the identification and characterization of BFIs that occur within complex natural microbiomes. The theoretical framework of this project is built upon gaining a more comprehensive understanding of how bacteria and fungi sense, respond to, and co-evolve with one another using multiomics-based interrogations. A more complete understanding of the molecular mechanisms underlying BFIs allows interrogation of how the dynamics of these relationships are altered in the context of environmental change (e.g., nutrient availability, temperature). Through these studies, the team hopes to gain a predictive understanding of how these interactions impact microbiome function and how they may be altered to steer the function of soil ecosystems and increase resilience to climate change and other environmental perturbations.

The burgeoning field of BFI research is quickly gaining interest in broader fields such as microbial ecology and evolution. This is at least partially due to the fact that current knowledge on BFIs suggests that interactions between these two kingdoms are quite common (Robinson et al. 2021). Furthermore, it has been demonstrated that BFIs can have direct impacts on the ecological functions performed by participating bacterial and fungal partners, suggesting that BFIs play an important role in the ecology and evolution of environmental microbiomes (Deveau et al. 2018; Pierce et al. 2021). However, it is currently very challenging to assess the state of the field, particularly with respect to what bacterial and fungal taxa have been previously reported as participants in BFIs. This is due to nonstandardized methods and format in which BFIs are reported in literature, as different descriptors (BFIs, endofungal, symbiotic, etc.) are often used, and some interactions are only reported in tables and/or figures. To address this problem, the BFI SFA has performed a comprehensive search to establish a database containing current knowledge on BFIs and which bacterial and fungal taxa participate in them. This database has been integrated into the BFI Research Portal (https:// sfa-bfi.edgebioinformatics.org/search), which can be queried using specific taxa to identify if any known BFIs involving that taxa have been previously reported in the literature. The database is a dynamic resource that will be updated as new BFI descriptions are published and allows users to submit any descriptions of BFIs not presently represented in the database after internal review. Additionally, the team is continually analyzing nonpublished fungal sequencing data from the National Center for Biotechnology Information Sequence Read Archive to find potential bacterial associates/interactions. Interactive visual outputs allow users to expand upon their initial queries to gain insights as to the diversity of BFIs related to their taxa or lineages of interest. This centralized resource will continually be developed to allow researchers in the field the ability to conduct BFI relevant analyses using custom bioinformatic workflows and enable the BFI database to be integrated into any analyses performed within the portal.

Deveau, A., et al. 2018. "Bacterial–Fungal Interactions: Ecology, Mechanisms and Challenges," *FEMS Microbiology Reviews* **42**(3), 335–52.

Pierce, E. C., et al. 2021. "Bacterial–Fungal Interactions Revealed by Genome-Wide Analysis of Bacterial Mutant Fitness," *Nature Microbiology* **6**(1), 87–102.

Robinson, A. J., et al. 2021. "Widespread Bacterial Diversity Within the Bacteriome of Fungi," *Communications Biology* **4**(1), 1168.

**Funding Information:** This research was supported by an SFA Grant from the U.S. Department of Energy (DOE), Biological and Environmental Research (BER) Program, Biological Systems Science Division (BSSD) under grant No. LANLF59T.

## Expanding Knowledge of Bacterial-Fungal Interactions in Environmental Microbiomes

Aaron Robinson\* (arobin@lanl.gov), Julia Kelliher, Buck Hanson, Reid Longley, Demosthenes Morales, **Patrick Chain** 

Los Alamos National Laboratory (LANL)

https://genomicscience.energy.gov/lanlbfi/

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

**Goals:** As part of the LANL science focus area (SFA) on Bacterial-Fungal Interactions (BFIs), researchers are developing novel bioinformatic and experimental tools and resources for the identification and characterization of BFIs that occur within complex natural microbiomes. The theoretical framework of this project is built upon gaining a more comprehensive understanding of how bacteria and fungi sense, respond to, and co-evolve with one another using multiomics-based interrogations. A more complete understanding of the molecular mechanisms underlying BFIs allows interrogation of how the dynamics of these relationships are altered in the context of environmental change (e.g., nutrient availability, temperature). Through these studies, the team hopes to gain a predictive understanding of how these interactions impact microbiome function and how they may be altered to steer the function of soil ecosystems and increase resilience to climate change and other environmental perturbations.

Diverse members of the bacterial and fungal kingdoms often co-dominate environmental microbiomes. Over the past decade, it has become clear that members from these two kingdoms frequently interact (Robinson et al. 2021). However, many gaps in knowledge, resources, and data remain in the field of BFIs. The BFI SFA has performed a number of investigations which have provided increased knowledge on the diversity of these interactions, how bacteria and fungi co-evolve, the molecular mechanisms that drive BFIs, and other important areas within the field. These investigations have involved taxonomically diverse bacteria and fungi, including a mix of model organisms which provide high tractability and resources for experimental work, and nonmodel organisms from dryland environments. This mix has provided fundamental knowledge on the mechanisms employed by bacteria and fungi to sense and respond to one another using model BFIs, while enabling continued investigations into the diversity of these mechanisms in natural systems through the use of nonmodel BFIs. Furthermore, this approach allows the assessment of how BFI may be altered over evolutionary time or as a result of changes in their environment.

Robinson, A. J., et al. 2021. "Widespread Bacterial Diversity Within the Bacteriome of Fungi," *Communications Biology* **4**(1), 1168.

**Funding Information**: This research was supported by an SFA Grant from the U.S. Department of Energy (DOE), Biological and Environmental Research (BER) Program, Biological Systems Science Division (BSSD) under grant No. LANLF59T.

## **SCIENCE FOCUS AREA**

ENIGMA: Ecosystems and Networks Integrated with Genes and Molecular Assemblies

## ENIGMA Environmental Atlas: An Integrated Approach to Linking Microbial Genotype to Phenotype in a Dynamic Subsurface Ecosystem

Romy Chakraborty<sup>1\*</sup> (rchakraborty@lbl.gov), Adam M. Deutschbauer<sup>2</sup>, Brandon C. Enalls<sup>1</sup>, Jennifer V. Kuehl<sup>2</sup>, Markus de Raad<sup>2</sup>, Bradley W. Biggs<sup>2</sup>, Hira Lesea<sup>2</sup>, Lauren M. Lui<sup>2</sup>, Valentine V. Trotter<sup>2</sup>, John-Marc Chandonia<sup>2</sup>, Torben Nielsen<sup>2</sup>, Hans K. Carlson<sup>2</sup>, Mingfei Chen<sup>1</sup>, Peter J. Walian<sup>1</sup>, Trent R. Northen<sup>2</sup>, Adam P. Arkin<sup>2</sup>, **Paul D. Adams<sup>1</sup>** 

<sup>1</sup>Lawrence Berkeley National Laboratory (LBNL); <sup>2</sup>University of California–Berkeley

**Goals:** The goal of the Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) science focus area (SFA) is to develop theoretical, technological, and scientific approaches to gain a predictive and mechanistic understanding of the biotic and abiotic factors that constrain microbial communities' assembly and activity in dynamic environments. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA uses a systems biology approach to integrate and develop laboratory, field, and computational methods.

Despite decades of research, there are still significant gaps in fundamental understanding of microbes, their interactions to form communities, and their relationships with the environment. Towards this, this project is constructing an 'ENIGMA Environmental Atlas,' a valuable resource that enables mapping genotype to phenotype for a significant number of diverse subsurface microbes at the research field site, the Oak Ridge Reservation Field Research Center. This Atlas includes a growing collection of over 2200 microbial isolates representing 36 orders and 895 unique strains from the site. Genome sequencing of 750 isolates to date has revealed both macro and microdiversity. High-resolution electron microscopy images reveal unique morphotypes and features. Researchers have established genetic toolkits and genome-wide mutant libraries in 25 diverse isolates to date and are using these resources to annotate genes of unknown function and interrogate physiological responses to environmental stressors.

The team has combined several statistical analyses using field environmental and sequencing-based metadata to identify high-priority targets for deeper isolation and characterization effort based on abundance, community correlation, and other metrics. The team is developing diverse assays to investigate the physiology of these 'most wanted' microbes including high-throughput carbon utilization, metal toxicity thresholds, biofilm formation, and exometabolomic profiling. Exometabolomic profiling of 135 isolates confirmed that substrate use is phylogenetically conserved. Together, these measurements enable understanding the interaction of microbes with each other and determining if field measurements of co-occurrence coincide with positive interactions among isolates, helping to progress the ability to predict community function from metagenomic and amplicon sequence variant data.

The team presents progress thus far on the development of this unique community-usable platform and highlights several instances where the Atlas can be used to better understand the complexities that govern microbial community structure and function in the environment.

**Funding Information**: This material by ENIGMA SFA Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program under contract No. DE-AC02-05CH11231.

## ENIGMA Long Read Sequencing and Assembly for Microbial Genomes: KBase Integration for Assembly and LISA Workshop

Lauren M. Lui<sup>1\*</sup> (Imlui@lbl.gov), Torben Nielsen<sup>1</sup>, John-Marc Chandonia<sup>1</sup>, **Paul D. Adams<sup>1,2</sup>** 

<sup>1</sup>Lawrence Berkeley National Laboratory (LBNL); <sup>2</sup>University of California–Berkeley

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

needed to infer and parameterize models of microbial system dynamics.

The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) science focus area (SFA) has spent time developing pipelines for sequencing and assembly of long read data from microbial isolates and metagenomes to help achieve the goal of casual microbial ecology. Researchers have developed the capability to isolate diverse organisms, extract the high-molecular weight DNA needed for single-molecule long-read sequencing, and perform the sequencing using Oxford Nanopore Technologies MinION and PromethION sequencers. To characterize the microbial diversity and activity at the Oak Ridge Reservation at Oak Ridge National Laboratory, ENIGMA anticipates isolating thousands of bacteria and archaea, as well as generating spatiotemporal series of fully resolved enrichments and metagenomes from the site. These sequencing projects assist the goals of linking genotype to phenotype and understanding the temporal, dynamic, and complex factors influencing microbial community structure and activity at the research site. ENIGMA uses isolates to help link genotype to phenotype by analyzing genomes in conjunction with transposon mutant libraries, metabolomics, and growth-condition data. High-quality genomes are essential for these types of experiments and ENIGMA science.

Researchers are currently adding new functionality to DOE Systems Biology KnowledgeBase (KBase) by implementing tools for using long-read data for assembly of isolates and methylation detection. By developing workflows within KBase, these tools will be more broadly available across the ENIGMA SFA and to other scientists, especially for scientists that do not specialize in computational methods. These apps and workflows will enable ENIGMA, as well as other DOE SFAs and microbiologists to (1) address scientific questions that would otherwise be infeasible with isolate assemblies using only short reads, (2) track provenance of data and methods used for assembly, and (3) share assemblies across the SFA for collaborations. By providing this new functionality in KBase, a foundation will be provided for further extensions in KBase to support developments in long-read technology. Currently, the genome assembler Unicycler has been released as a KBase app, and Flye is in development. Filtlong, a read quality tool, and Polypolish, an assembly polishing tool, are in beta.

In Summer 2023, the team will hold the Long-read Isolate Sequencing and Assembly (LISA) workshop at LBNL. This workshop will teach participants how to go from a microbial isolate to sequencing to assembly of a genome. The workshop will have a wet lab session to learn how to extract high-molecular-weight DNA, make nanopore libraries, and run a MinION sequencer. A separate computational session will be held on base calling of nanopore data, genome assembly using long read sequencing data, and genome annotation using KBase. Scientists from computational, modeling, and bench backgrounds are encouraged to attend both sessions. This workshop will be designed to accommodate learners from these diverse technical backgrounds.

**Funding Information**: This material by ENIGMA SFA Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program under contract No. DE-AC02-05CH11231.

## The Use of Synthetic Communities Reveals Disturbance of Process Partitioning Among Denitrifying Microbes Leads to Increased Nitrous Oxide Emissions

Jacob J. Valenzuela<sup>1\*</sup> (jvalenzu@isbscience.org), James Wilson<sup>1</sup>, Alex Carr<sup>1</sup>, Serdar Turkarslan<sup>1</sup>, Sara Altenburg<sup>2</sup>, Heidi Smith<sup>2</sup>, Anne Otwell<sup>3</sup>, Kristopher Hunt<sup>3</sup>, Jennifer Goff<sup>4</sup>, Farris Poole<sup>4</sup>, Xiaoxuan Ge<sup>4</sup>, Michael Thorgersen<sup>4</sup>, Peter J. Walian<sup>5</sup>, Vivek K. Mutalik<sup>5</sup>, Adam M. Deutschbauer<sup>5</sup>, Trent R. Northen<sup>5</sup>, Michael W. Adams<sup>4</sup>, Romy Chakraborty<sup>5</sup>, Damien A. Elias<sup>5</sup>, David A. Stahl<sup>3</sup>, Matthew W. Fields<sup>2</sup>, Nitin S. Baliga<sup>1,2</sup>, Adam P. Arkin<sup>5,6</sup>, **Paul D. Adams<sup>5,6</sup>** 

<sup>1</sup>Institute for Systems Biology; <sup>2</sup>Montana State University; <sup>3</sup>University of Washington; <sup>4</sup>University of Georgia; <sup>5</sup>Lawrence Berkeley National Laboratory; <sup>6</sup>University of California–Berkeley

#### http://enigma.lbl.gov

Goals: The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) science focus area (SFA) uses a systems biology approach to understand the interaction between microbial communities and the ecosystems they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods. Thus, ENIGMA has been organized into several campaigns involving multiple institutes with varying expertise. An overarching goal of the Environmental Simulations and Modelling Campaign is to simulate, model, and predict the mechanistic underpinnings of field-observed phenomena. This includes characterizing the process partitioning of nitrous oxide  $(N_2O)$  emissions in varying ecological contexts (pH, metal availability, oxygen, and phage attack) using field isolates assembled into synthetic communities (SynComs).

The Field Research Center (FRC) at Oak Ridge, Tenn., has some of the highest subsurface nitrate concentrations (>10g/L) ever recorded. This pool of subsurface nitrate, a remnant of legacy activities, can end up as the greenhouse

gas N<sub>2</sub>O via incomplete denitrification or as nitrogen gas  $(N_2)$  when completely denitrified. Based on spatiotemporal field surveys of biogeochemistry, hydrology, metagenomics, and activity measurements, the ENIGMA team is formulating mechanistic hypotheses to explain ecologically important phenomena, such as the emissions of increased N<sub>2</sub>O emissions in wells that transiently transition from a neutral to an acidic pH. Here, using the N<sub>2</sub>O emission phenomenon, researchers explicate how such phenomena are being dissected through laboratory studies using synthetic communities assembled from field isolates of relevant organisms. The ultimate goal is to characterize the mechanism(s) responsible for each field phenomenon and test generalizability of findings back at the field site.

Analysis of FRC field isolates with denitrifying capabilities showed that more than half of the isolates were missing at least one step in the denitrification pathway. Researchers hypothesized that multiple microbes with complementary enzymatic steps likely work together in communities to complete denitrification through the exchange of nitrogen intermediates. Further, the team hypothesized that different abiotic and biotic factors that inhibit specific enzymatic reactions would likely decouple complete denitrification, contributing to significant levels of N2O emissions at the FRC. To test this hypothesis, the team established a synthetic community (SynCom) of two field isolates-Rhodanobacter sp. R12 and Acidovorax sp. 3H11—which together can perform complete denitrification but cannot independently. Therefore, a cross-campaign initiative was generated to elucidate different mechanisms of abiotic control, including pH shifts, microaerobic environments, and metal availability. Using time-course experiments, researchers determined that a shift in pH from neutral pH 7 to pH 6 or an increase in nickel (Ni) concentrations was enough to decouple the complete denitrification process of the SynCom by different mechanisms. Strikingly, both perturbations resulted in significant increases in N<sub>2</sub>O emissions. Transcriptome analysis of the SynCom at differing pH or Ni conditions suggests dynamic changes in community composition and physiological states. Current experiments are focused on shifts in pH at varying carbon/nitrogen ratios, oxygen, and even phage attack that may result in the decoupling of denitrification partitioning. Transmission electron microscope imaging suggests very different morphologies between the two field isolates that may play an essential role in carbon, nitrogen, and phosphorous fluxes between the organisms. Together, all these efforts are leading to the construction of a context-specific gene regulatory network that can be used to predict how environmental fluctuations at the field site will impact emissions of N<sub>2</sub>O.

**Funding Information**: This material by ENIGMA SFA Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program under contract No. DE-AC02-05CH11231.

## Three-Dimensional High Spatial Resolution Simulation for Groundwater Flow and Nitrogen Transport Under Rainfall Perturbations in the Subsurface of Area 3

Jinwoo Im<sup>1\*</sup> (jinwooim@lbl.gov), Andrew Putt<sup>2</sup>, Kathleen F. Walker<sup>2</sup>, Dominique C. Joyner<sup>2</sup>, James Marquis<sup>3</sup>, Lauren M. Lui<sup>1</sup>, Dipankar Dwivedi<sup>1</sup>, Alex Carr<sup>4</sup>, Yupeng Fan<sup>5</sup>, Jennifer Goff<sup>6</sup>, Kristopher Hunt<sup>7</sup>, Jonathan Michael<sup>5</sup>, Farris Poole<sup>6</sup>, Yajiao Wang<sup>5</sup>, Michael W. Adams<sup>6</sup>, Nitin S. Baliga<sup>4</sup>, David A. Stahl<sup>7</sup>, Jizhong Zhou<sup>5</sup>, Matthew W. Fields<sup>3</sup>, Terry C. Hazen<sup>2,8</sup>, Michelle E. Newcomer<sup>1</sup>, Adam P. Arkin<sup>1,9</sup>, **Paul D. Adams<sup>1,9</sup>** 

<sup>1</sup>Lawrence Berkeley National Laboratory (LBNL); <sup>2</sup>University of Tennessee–Knoxville; <sup>3</sup>Montana State University; <sup>4</sup>Institute for Systems Biology; <sup>5</sup>University of Oklahoma; <sup>6</sup>University of Georgia; <sup>7</sup>University of Washington; <sup>8</sup>Oak Ridge National Laboratory (ORNL); <sup>9</sup>University of California–Berkeley

#### http://enigma.lbl.gov

**Goals:** The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) science focus area (SFA) 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.

Uranium and nitrate contaminant transport in Area 3 near the S3 ponds at ORNL are investigated through 3D fieldscale modeling and simulation. This project leverages a recently acquired Cone Penetration Testing (CPT) dataset which provides the hydraulic conductivity field of Area 3 with a high spatial resolution as input data to the numerical subsurface model. The CPT data shed light on local heterogeneity of subsurface materials, significantly decreasing the uncertainty due to limited sampling points. By further using 27 well survey data collected by the ENIGMA SFA (e.g., meteorological, hydrological, microbial, and geochemical datasets), a hydrogeological model is built on PFLOTRAN and runs on the high-performance computing system at National Energy Research Scientific Computing Center. Generalized stoichiometries are used for biogeochemical reactions related to nitrogen cycling. Computational results of the 3D field-scale simulation show: (1) different flow and transport regimes depending on subsurface materials, (2) impacts of rainfall events on nitrous oxide emission,

(3) influential controls of flow conditions through sensitivity analysis enabling a full treatment of the ModEx approach to designing and implementing the Subsurface Observatory (SSO). The results help researchers understand nitrogen cycling in Area 3 and determine the location of the ENIGMA SSO site. Furthermore, the results will be compared to omics-informed modeling and simulation as planned in the Framework for Integrated, Conceptual, and Systematic Microbial Ecology (Lui et al. 2021). Lui, L. M., et al. 2021. "Mechanism Across Scales: A Holistic Modeling Framework Integrating Laboratory and Field Studies for Microbial Ecology," *Frontiers in Microbiology* **12**, 642422. DOI:10.3389/ fmicb.2021.642422.

**Funding Information:** This material by ENIGMA SFA Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program under contract No. DE-AC02-05CH11231.

## **SCIENCE FOCUS AREA**

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

## Examine Plant-Microbe-Soil Interactions Using Fabricated Ecosystems Across Scales

Hsiao-Han Lin<sup>1\*</sup> (hsiaohanlin@lbl.gov), Albina Khasanova<sup>1</sup>, Shwetha M. Acharya<sup>1</sup>, Mon O. Yee<sup>1</sup>, Spencer Diamond<sup>2</sup>, Peter F. Andeer<sup>1</sup>, Nameera F. Baig<sup>1</sup>, Omolara T. Aladesanmi<sup>1</sup>, Catharine A. Adams<sup>2</sup>, Marta Torres Béjar<sup>1</sup>, Catharine A. Adams<sup>2</sup>, Trenton K. Owen<sup>1</sup>, Hans K. Carlson<sup>1</sup>, Nick Downs<sup>1</sup>, Jillian Banfield<sup>2</sup>, N. Louise Glass<sup>2</sup>, Romy Chakraborty<sup>1</sup>, Kateryna Zhalnina<sup>1</sup>, Jenny C. Mortimer<sup>1, 3</sup>, Adam M. Deutschbauer<sup>1</sup>, **Trent R. Northen<sup>1\*</sup>**(TRNorthen@lbl.gov)

<sup>1</sup>Lawrence Berkeley National Laboratory; <sup>2</sup>University of California– Berkeley; <sup>3</sup>University of Adelaide, Australia

#### http://mCAFEs.lbl.gov

Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design. To do this, advanced fabricated ecosystems are used in combination with gene-editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions in situaddressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies-harnessing beneficial microbiomes to support sustainable bioenergy and improving understanding of nutrient cycling in the rhizosphere.

Studying plant-microbe-soil interactions including those mediated by root glycans is challenging due to the complexity and variability found in natural ecosystems. Fabricated ecosystems offer an opportunity to recapitulate aspects of these systems in reduced complexity and controlled laboratory settings. However, it is important to benchmark their performance against established systems and identify specific protocols for these studies. Here, researchers compared the colonization and persistence of a field-derived microbiome in colonizing the model grass Brachypodium distachyon when grown in sterile devices called EcoFABs as compared to conventional containers like pots and tubes (Acharya et al. 2023). Comparable plant growth and microbial community composition was obtained between conventional containers and the EcoFAB. The team also observe a distinct microbiome profile for the rhizosphere (root tip or root base) and the bulk soil. Researchers then used a synthetic community (SynCom) to study how different growth media and inoculation methods affect microbial community assembly (Coker et al. 2022). The results showed that sample types (sand, rhizosphere, and root) but not the inoculation method significantly affects microbial community assembly. Next, the team studied how plant cell wall composition and root exudates affect microbial community assembly in the rhizosphere by using a lignin biosynthetic B. distachyon transgenic line that has an altered root cell wall composition.

Researchers observed a higher *Rhizobium* colonization in the transgenic line than in the wild type line Bd21-3 suggesting a possible connection between root aromatics and root colonization. To explore this in more detail, the team performed metabolomic analysis of a diverse collection of *Brachypodium* accessions to identify lines with elevated aromatic compounds in the exudates. Researchers then used a meter-scale fabricated ecosystem to investigate SynCom colonization between two lines with dramatically different aromatic exudate compositions and again observed altered rhizosphere community compositions. Together, these results highlight the potential for using fabricated ecosystems to study the molecular ecology of plant-microbe interactions.

Acharya, S.M., et al. 2023. "Fine Scale Sampling Reveals Spatial Heterogeneity of Rhizosphere Microbiome in Young *Brachypodium* Plants," *BioRxiv*.

Coker, J., et al. 2022. "A Reproducible and Tunable Synthetic Soil Microbial Community Provides New Insights into Microbial Ecology," *mSystems*, e0095122.

**Funding Information**: This material by m-CAFEs Microbial Community Analysis and Functional Evaluation in Soils, (m-CAFEs@lbl.gov) a science focus area led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under contract No. DE-AC02-05CH11231.

### Phage Engineering for Targeted Editing of Microbial Communities

Benjamin A. Adler<sup>1\*</sup> (ba14@berkeley.edu), Tomas Hessler<sup>1</sup>, Avery Roberts<sup>2</sup>, Matthew Nethery<sup>2</sup>, John Beckley<sup>2</sup>, Brady F. Cress<sup>1</sup>, Arushi Lahiri<sup>1</sup>, Jillian Banfield<sup>2</sup>, Rodolphe Barrangou<sup>2</sup>, Jennifer A. Doudna<sup>1</sup>, Adam M. Deutschbauer<sup>3</sup>, **Trent R. Northen<sup>3\*</sup>** 

<sup>1</sup>University of California–Berkeley; <sup>2</sup>North Carolina State University; <sup>3</sup>Lawrence Berkeley National Laboratory

#### http://mCAFEs.lbl.gov

Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design. To do this, advanced fabricated ecosystems are used in combination with gene-editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions in situaddressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving the understanding of nutrient cycling in the rhizosphere.

Bacteriophages are estimated to be the most abundant biological entities on Earth, outnumbering bacteria by 10 to 1.

Owing to their natural ecological abundance, genetic diversity, and ability to transduce DNA, they represent attractive gene delivery vehicles to edit microbial communities in situ. However, the ability to broadly edit phages themselves has been limited by a diversity of mechanisms for phages to protect DNA genomes. In order to edit the diversity of phages, we establish a generalizable editing for phage genetic manipulation based off RNA-guided, RNA-targeting endonuclease, LbuCas13a. Researchers find LbuCas13a to be a remarkably potent phage inhibitor, suggesting that phage RNA is generally vulnerable during viral infection. When challenged against Escherichia coli phage phylogeny, researchers find no apparent phage-encoded limits to LbuCas13a antiviral activity. Further, researchers highlight how leveraging this potent antiphage activity can be used to flexibly edit diverse phages with edits as small as a single codon or as large as multigene deletions. Researchers further discuss opportunities for engineered phages to edit microbial within fabricated ecosystems using phage-derived base editing technology and novel phages infecting members of synthetic microbial consortia. The ability to robustly edit bacteriophages will not only lead to a deeper understanding of phage genetic diversity but also facilitate meaningful genetic changes to microbial communities.

Adler, B. A., et al. 2022. "Broad-Spectrum CRISPR-Cas13a Enables Efficient Phage Genome Editing," *Nature Microbiology* **7**, 1967–79.

Nethery, M. A., et al. 2022. "CRISPR-Based Engineering of Phages for *in situ* Bacterial Base Editing," *Proceedings of the National Academy of Sciences* **119**, e2206744119.

**Funding Information**: This material by m-CAFEs Microbial Community Analysis and Functional Evaluation in Soils (m-CAFEs@lbl.gov), a science focus area led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under contract No. DE-AC02-05CH11231.

## **SCIENCE FOCUS AREA**

Phenotypic Response of the Soil Microbiome to Environmental Perturbations

### Impact of Soil Viruses on Microbial Compositions and Functions

Ruonan Wu<sup>\*1</sup> (ruonan.wu@pnnl.gov), Clyde A. Smith<sup>2</sup>, Garry W. Buchko<sup>1</sup>, Ian K. Blaby<sup>3</sup>, Nikos Kyrpides<sup>3</sup>, Yasuo Yoshikuni<sup>3</sup>, John R. Cort<sup>1</sup>, Michelle R. Davison<sup>1</sup>, William C. Nelson<sup>1</sup>, Yuqian Gao<sup>1</sup>, Kristin E. Burnum-Johnson<sup>1</sup>, Montana Smith<sup>1</sup>, Mary S. Lipton<sup>1</sup>, Ryan McClure<sup>1</sup>, Jason E. McDermott<sup>1</sup>, Janet K. Jansson<sup>1</sup> **Kirsten S. Hofmockel<sup>1</sup>** 

<sup>1</sup>Pacific Northwest National Laboratory; <sup>2</sup>Stanford University; <sup>3</sup>D0E Joint Genome Institute

#### https://www.pnnl.gov/projects/soil-microbiome/research

Goals: PNNL's Phenotypic Response of Soil Microbiomes science focus area aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. Researchers perform multiscale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi, viruses, and plants that regulate community functions throughout the soil profile. Data are used to parametrize individual- and population-based models for predicting interspecies and interkingdom interactions. Predictions are tested in laboratory and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose and sequester organic carbon and enables prediction of how biochemical reaction networks shift in response to changing moisture regimes.

Soil is known to harbor diverse and abundant viruses, but most soil viruses are uncharacterized. The ecological impacts of soil viruses and their responses to climate change remain understudied. To address these knowledge gaps, researchers launched a cross-scale study, from viral genes of interest to viral communities in soil microcosms and field experiments. Viruses carry auxiliary metabolic genes (AMGs) that potentially contribute to soil metabolic processes while tuning the host machinery towards their own replication (Jansson and Wu 2022). The team, therefore, first focused on AMGs as viral genes of interest. In collaboration with the Joint Genome Institute (JGI), the Environmental Molecular Sciences Laboratory (EMSL) and the Stanford Synchrotron Radiation Lightsource (SSRL), researchers verified the activity of a chitosanase encoded by a soil viral AMG and determined the first protein structure within the all-chitosanase family GH75 at ultra-high resolution (<0.9 Å; Wu et al. 2022). Co-crystal structures with site-directed mutants and chitohexaose further elucidated the catalytic mechanism of the viral chitosanase. This study provides more molecular evidence that soil viruses may aid their hosts in organic carbon decomposition. To quantify the metabolic contribution of soil viruses at the community level, researchers next investigated viral population dynamics under the impact of environmental perturbation.

Using three contrasting field experiments, researchers tested the viral response to changes in soil moisture, studied viral communities in a range of grassland soils with different historical precipitation patterns (Wu et al. 2021), and then generated hypotheses to test in a soil incubation experiment (Wu et al. 2021). In collaboration with the National Energy Research Scientific Computing Center (NERSC), researchers assembled the deeply sequenced soil metagenomes (>1 Tb each; Nelson et al. 2020) and recovered a total of 2,631 viral contigs including 14 complete circular viral genomes (Wu et al. 2021). Researchers found that soil with a lower historical moisture content harbored significantly higher viral diversity and abundance, while displaying less evidence of virus-host interactions, suggesting a predominance of lysogenic viruses in drier soils. The detection of AMGs involved in 18 metabolic pathways further supports the finding of viral contributions to carbon metabolism in soil (Wu et al. 2022). Researchers then selected the grassland soil exposed to an intermediate historical precipitation and either experimentally wetted the soil to saturation or airdried the soil to represent experimental wet and dry treatments, respectively (Wu et al. 2021). Researchers observed a lower overall level of transcription in drier soil but across more diverse DNA viruses. A higher fraction of noncoding RNAs and more transcripts of lysogenic markers (i.e., integrases and excisionases) were detected in drier soil, further supporting a higher prevalence of lysogenic viruses in drier soils as shown in the field study.

To demonstrate the direct viral impact on soil microbiome of changing soil moisture, researchers applied high-throughput chromosomal confirmation capture (Hi-C) metagenomics to capture and identify viruses that were infecting hosts at the time of sampling and metatranscriptomics to detect the transcriptional activity of the host-associated viruses (Wu et al., submitted). Although the host-associated viruses accounted for only 5.3% to 15.0% of the total viral sequence abundance, they shared similar patterns previously detected in the whole viral communities (Wu et al. 2021; Wu et al. 2021). The host-associated viruses in wetter soils had higher transcriptional levels and were inversely correlated with abundances of their hosts (p < 0.05). The richness (number of different types of virus) of the host-associated viruses and the average viral copies per host (VPH), however, were higher in drier soils. These results suggest that viral infections were mostly lytic under wet conditions while more prevalent and lysogenic under dry conditions. The hosts infected by soil viruses were found to be central in community co-occurrence networks, highlighting the impact of viral infections on soil microbiome structure. This study is the first to target the detection of host-associated viruses in soil and reveals the impact of soil viruses on microbial composition with changing soil moisture.

Future work aims to bridge the findings across scales by leveraging the modeling capabilities (e.g., mechanistic modeling to simulate viral predation in porous systems) to test and generate a more comprehensive and transferable understanding of soil viruses. The cross-scale framework will continue to provide new information of the influence of changing soil moisture on viruses and their potential ecological impacts on soil microbiomes.

Jansson, J. K, and R. Wu. 2022. "Soil Viral Diversity, Ecology and Climate Change," *Nature Reviews Microbiology*, 1–16.

Nelson, W. C., et al. 2020. "Terabase Metagenome Sequencing of Grassland Soil Microbiomes," *Microbiology Resource Announcements* **9**(32).

Wu, R., et al. 2022. "Structural Characterization of a Soil Viral Auxiliary Metabolic Gene Product–a Functional Chitosanase," *Nature Communications* **13**(1), 1–14.

Wu, R., et al. 2021. "DNA Viral Diversity, Abundance, and Functional Potential Vary Across Grassland Soils with a Range of Historical Moisture Regimes," *mBio* **12**(6), e02595-21.

Wu, R., et al. 2021. "Moisture Modulates Soil Reservoirs of Active DNA and RNA Viruses," *Communications Biology* **4**(1), 1–11.

Wu, R., et al. Submitted. "High-Throughput Chromosomal Confirmation Capture (Hi-C) Metagenome Sequencing Reveals Soil Moisture Impacts on Phage-Host Interactions."

**Funding Information:** PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science program, Biological and Environmental Research Program, under FWP 70880.

### OMEGGA: A Computationally Efficient Omics-Guided Global Gapfilling Algorithm for Phenotype-Consistent Metabolic Network Reconstruction

Hyun-Seob Song<sup>1\*</sup> (hsong5@unl.edu), Firnaaz Ahamed<sup>1</sup>, David M. L. Brown Jr.<sup>2</sup>, Christopher S. Henry<sup>3</sup>, Janaka N. Edirisinghe<sup>3</sup>, Aimee K. Kessell<sup>1</sup>, William C. Nelson<sup>4</sup>, Jason E. McDermott<sup>4</sup>, **Kirsten S. Hofmockel**<sup>4</sup>

<sup>1</sup>University of Nebraska-Lincoln; <sup>2</sup>Pacific Northwest National Laboratory (PNNL); <sup>3</sup>Argonne National Laboratory; <sup>4</sup>Pacific Northwest National Laboratory

https://www.pnnl.gov/projects/soil-microbiome/research

Goals: PNNL's Phenotypic Response of Soil Microbiomes science focus area (SFA) aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. Researchers perform multiscale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition using chitin as a model compound. Integrated experiments address spatial and interkingdom interactions among bacteria, fungi, viruses and plants that regulate community functions throughout the soil profile. Data are used to parametrize individual- and population-based models for predicting interspecies and interkingdom interactions. Predictions are tested in laboratory and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

Genome-scale metabolic networks are a valuable tool for gaining a mechanistic understanding of microbial metabolism and predicting trophic interactions within microbial communities. Metabolic network models are constructed through three key steps-draft metabolic model construction based on genome annotations, filling knowledge gaps in biochemical pathways by adding missing reactions (gapfilling), and further refinement and curation. The DOE Systems Biology Knowledgebase (KBase) automates this process by providing a suite of computational apps and modules (https://docs.kbase.us/apps/analysis/ metabolic-modeling). Draft networks constructed based on genome annotations alone do not contain all of the key reactions necessary for robust predictions, and therefore fail to predict biomass production/cell growth as experimentally observed. Gapfilling (i.e., identifying and adding those missing reactions) is a critical next step for enhancing the predictive power of metabolic networks. Current gapfilling

algorithms seek a minimal number of reactions following the parsimonious approach and repeats this process in a sequential manner for a given set of phenotypic growth data. However, the reactions added as such are not always biologically relevant, causing the model predictions to be inconsistent.

To address these issues, researchers designed a new gapfilling algorithm OMEGGA (OMics-Enabled Global GApfilling). As indicated by its name, OMEGGA uses diverse data sources (such as amplicon, transcriptomic, proteomic, and metabolomic data) to simultaneously fit a draft metabolic model to all available phenotype data. In this work, researchers demonstrate the two major capabilities of OMEGGA: global and omics-guided gapfilling.

For global (or simultaneous) gapfilling, researchers developed a linear programming (LP)-based algorithm to identify a minimal set of reactions meeting all experimentally observed growth conditions, without iterative fitting. The LP-based algorithm shows far superior performance compared to existing mixed integer linear programming (MILP)-based algorithms as demonstrated through a case study using Escherichia coli. While the computational burden builds up exponentially as the number of media conditions increases, the actual computational time was indeed acceptable, indicating that the algorithm can be flexibly extended to more complex datasets by leveraging higher performance computational power in KBase. Importantly, the clever design of LP-based algorithm allows researchers to use non-proprietary LP solvers, avoiding any potential licensing issues. In parallel, researchers also developed an algorithm that incorporates (1) gene annotations from multiple complementary pipelines (e.g., RAST, Prokka, Koala, DeepEC), and (2) additional omics data (e.g., transcriptomic profiles). In the case study of *E. coli*, the gapfilling solutions showed much stronger genomic and experimental consistencies than the typical parsimonious gapfilling. The inclusion of biologically relevant reactions is critical to avoid false positives, which traditionally requires manual curation.

Researchers will extend the test cases to include nonmodel organisms by using condition-specific multiomics and phenotype data from the Model Soil Consortia-2 (MSC-2) and associated isolated organisms developed through PNNL's Soil Microbiome SFA. Researchers are working with the KBase team to incorporate the OMEGGA algorithm into KBase. The team will build an external application library for gapfilling and incorporate that into existing KBase apps that can leverage the multiomics data to derive more biologically relevant and realistic gapfilling solutions. To improve quality and supportability of the software, testing and documentation will also be incorporated into automated processes. The new optimization algorithms and KBase apps greatly facilitate the construction of high-quality metabolic networks by simultaneously incorporating molecular and phenotypic observations, eliminating the need for time-consuming, manual troubleshooting.

**Funding Information**: PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science program, Biological and Environmental Research Program, under FWP 70880 and FWP 78749.

### Removal of Primary Nutrient Degraders Reduces Growth and Modifies Functional Pathways of Soil Microbial Communities with Genomic Redundancy

Ryan McClure<sup>1\*</sup> (Ryan.Mcclure@pnnl.gov), Marci Garcia<sup>1</sup>, Sneha Couvillion<sup>1</sup>, Amy E. Zimmerman<sup>1</sup>, Yuliya Farris<sup>1</sup>, **Kirsten S. Hofmockel<sup>1,2</sup>** 

<sup>1</sup>Pacific Northwest National Laboratory (PNNL); <sup>2</sup>Iowa State University https://www.pnnl.gov/projects/soil-microbiome/research

Goals: PNNL's Phenotypic Response of Soil Microbiomes science focus area aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. Researchers perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi, viruses, and plants that regulate community functions throughout the soil profile. Data are used to parametrize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in laboratory and field experiments to reveal individual and community microbial phenotypes. Knowledge gained provides fundamental understanding of how soil microbes interact to decompose and sequester organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

Many ecosystem functions related to plant growth or carbon and nutrient cycling and stabilization rely on microbial metabolism. As a result, microbial communities are major drivers of carbon use efficiency (CUE). Furthermore, ecological theory indicates the importance of keystone community members that may carry out critical aspects of community functioning or interact with many other community members positively or negatively. Therefore, a quantitative assessment of species-specific responses within a community context is required to understand how microorganisms and keystone species within a soil community interact to support collective growth and community function related to carbon cycling.

To investigate how individual members of a microbial community contribute to decomposition, community growth, and CUE, researchers used a model substrate, chitin, and a Model Soil Consortium, MSC-2 (McClure et al. 2022). While MSC-2 can grow using chitin as the sole carbon source, the individual functions and metabolic contributions of the constituent species remain unknown. To quantify specific roles within this model community, researchers carried out experiments leaving out members of MSC-2 to test the implications for community biomass yields, CO2 production, proteomic and lipidomic profiles, and extracellular metabolites. Researchers chose two members to iteratively leave out: Streptomyces, as it is predicted via gene expression analysis to be a major chitin degrader in the community, and Rhodococcus as it is predicted via species co-abundance analysis to interact with several other members (McClure et al. 2020). The experiments revealed that when MSC-2 lacked Streptomyces, growth and respiration of the community was severely reduced, even though other members of MSC-2 can degrade chitin. Removal of Streptomyces also led to changes in abundance for several other species compared to the complete MSC-2 community, pointing to a comprehensive shifting of the community structure when important members are removed. In addition, while the absence of Streptomyces led to differences in microbial taxonomy compared to the complete MSC-2 community there were only minor changes compared to the community's starting point, indicating minimal growth and activity when this keystone species is removed. In contrast, while the absence of Rhodococcus also led to taxonomic changes compared to the complete MSC-2 community, removal of this keystone species had little effect on community growth and respiration. A further multiomic analysis of communities lacking Streptomyces showed that without this member the proteomic profile of the community was distinct from the complete MSC-2. Proteins from Sinorhizobium and Ensifer were the most abundant in a community lacking Streptomyces while proteins from Ensifer and Sphingopyxis were more prevalent in the complete MSC-2. Major differences were

also seen between the lipidomic profiles of MSC-2 with and without *Streptomyces* with the breakdown of triglycerides more prevalent in complete MSC-2 communities. As complete MSC-2 communities grow far better than those lacking *Streptomyces* this triglyceride breakdown may be a response to chitin being fully metabolized, forcing members to rely on other energy sources at the later timepoints of the experiments. Interestingly, polar metabolite abundances in the culture supernatants were relatively similar between communities with and without *Streptomyces*.

These results show that when keystone chitin-degrading members are removed, other members, even those with the ability to degrade chitin, do not fill the same metabolic niche to promote community growth. In addition, highly connected members may be removed with similar or even increased levels of community growth and respiration. This suggests that removal of keystone members can have positive or negative effects on overall community growth, an outcome driven not only by identity of the keystone member but the magnitude and type of the interactions it has with other members. The findings are critical to a better understanding of soil microbiology, specifically in how communities maintain activity when biotic or abiotic factors lead to changes in biodiversity in soil systems.

McClure, R., et al. 2022. "Interaction Networks Are Driven by Community-Responsive Phenotypes in a Chitin-Degrading Consortium of Soil," *Msystems* 7(5), e00372-22.

McClure, R, et al. 2020. "Development and Analysis of a Stable, Reduced Complexity Model Soil Microbiome," *Frontiers in Microbiology* **11**, 1987.

**Funding Information**: PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science program, Biological and Environmental Research Program,, under FWP 70880. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL.

## **SCIENCE FOCUS AREA**

## **Plant-Microbe Interfaces**

# Plant-Microbe Interfaces: Capturing and Interpreting the Role of *Populus'* Microbiome

Dale A. Pelletier\* (pelletierda@ornl.gov), Paul E. Abraham, Manasa R. Appidi, Leah Burdick, Dana L. Carper, Alyssa A. Carrell, Robert L. Hettich, Sara S. Jawdy, Dawn M. Klingeman, Jennifer Morrell-Falvey, Bryan T. Piatkowski, David J. Weston, Jia Wang, **Mitchel J. Doktycz** 

Oak Ridge National Laboratory

http://pmiweb.ornl.gov/

**Goals**: The goal of the Plant-Microbe Interfaces (PMI) science focus area (SFA) is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. Populus and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, the project focuses on (1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic *Populus*-microbe interfaces, (2) defining the chemical environment and molecular signals that influence community structure and function, and (3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.

Microbial communities play an integral role in the health and survival of their plant hosts. In ongoing efforts in the PMI SFA, researchers are capturing members of Populus' microbiome to understand basic concepts of plant and environmental selection. Representative bacterial strains from environmental samples of *Populus* roots have been isolated using a direct plating approach and compared to amplicon-based sequencing analysis of root samples (Carper et al. 2021). The resulting culture collection contains 3,211 unique isolates representing 10 classes, 18 orders, 45 families, and 120 genera from six phyla, based on 16S rRNA gene sequence analysis. The collection represents a significant fraction of the natural community of plant-associated bacteria as determined by phylogenetic analysis. Additionally, a representative set of 553 strains have had their genomes sequenced to facilitate functional analyses. This

culture collection allows for the exploration of microbial community function and an understanding of basic concepts of plant and environmental dependent selection.

The team is employing this collection to understand the mechanisms of microbial adaptation to Populus' root endosphere and rhizosphere. Microbial diversity of the endosphere is low compared to rhizosphere, indicating high selectivity of this compartment for specific taxa and microbial adaptation of functions needed to compete in this unique environment. Using Populus-bacterial model systems with communities of specific Variovorax strains, researchers successfully demonstrated that they could identify bacterial strains, genes, and associated functions potentially required for fitness in *Populus'* root rhizosphere or endosphere niches. L-fucose metabolism, glycoside hydrolases, pili/fimbriae production, and exopolysaccharide production were identified as important bacterial traits associated with efficient endosphere colonization. The team found the enrichment of genes in the L-fucose metabolic pathway intriguing, as metabolism of cell surface fucose residues has been implicated in enrichment of beneficial mammalian gut bacteria and suppression of pathogens. Additionally, L-fucose biosynthesis and fucosylation of cell surface macromolecules have been demonstrated to play a role in plant immune response. Researchers constructed gene deletion mutants of L-fuconolactonase in the L-fucose metabolic pathway via homologous recombination in Variovorax and confirmed this strain is defective in L-fucose growth and in Populus root colonization, relative to the wild-type strain.

In other applications of the microbial collection, the team has constructed microbial communities to elucidate organizational principles of community formation. Using genome-defined strains, systematic experiments, and computational modeling, researchers are identifying potential metabolic exchanges among species and gaining mechanistic insights into community structure. Co-culture and serial transfer experiments performed in defined media identified emergent, stable microbial communities (Wang et al. 2021; Shrestha et al. 2021). Using a complex medium environment, the effects of different initial inoculum ratios, up to three orders of magnitude, on community structure were investigated. The final compositions of the mixed communities with various starting compositions indicate that community structure is not dependent on the initial inoculum ratio. Modeling and omics analysis provide mechanistic insights into the emergence of community structure and indicate competitive relationships among the persistent organisms. These findings enlighten understanding of bacterial community formation and may guide efforts to manage rhizosphere bacterial communities. Collectively, these diverse applications of cultured representatives of Populus' microbial community are facilitating understanding of how Populus selects microbial partners and how its microbiome is structured.

Carper, D. L., et al. 2021. "Cultivating the Bacterial Microbiota of *Populus* Roots," *mSystems* **6**(3), e01306-20. DOI:10.1128/ mSystems.01306-20.

Shrestha, H. K., et al. 2021. "Metaproteomics Reveals Insights into Microbial Structure, Interactions, and Dynamic Regulation in Defined Communities as They Respond to Environmental Disturbance," *BMC Microbiology* **21**, 308. DOI:10.1186/ s12866-021-02370-4.

Wang, J., et al. 2021. "Formation, Characterization and Modeling of Emergent Synthetic Microbial Communities," *Computational and Structural Biotechnology Journal* **19**, 1917–27. DOI:10.1016/j. csbj.2021.03.034.

**Funding Information**: Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. The Plant-Microbe Interfaces SFA is supported by the U.S. Department of Energy, Office of Science, through the Genomic Science program, Biological and Environmental Research (BER) Program under FWP ERKP730.

### Plant-Microbe Interfaces: Molecular Insights into the Mutualistic Symbiosis Between *Populus* and Plant Growth–Promoting Bacteria

Bryan T. Piatkowski<sup>1\*</sup> (piatkowskibt@ornl.gov), Dana L. Carper<sup>1</sup>, Alyssa A. Carrell<sup>1</sup>, Caroline Harwood<sup>2</sup>, Sara S. Jawdy<sup>1</sup>, Jennifer Morrell-Falvey<sup>1</sup>, Yasuhiro Oda<sup>2</sup>, Dale A. Pelletier<sup>1</sup>, Amy Schaefer<sup>2</sup>, Amber Webb<sup>1</sup>, David J. Weston<sup>1</sup>, Xiaohan Yang<sup>1</sup>, **Mitchel J. Doktycz<sup>1</sup>** 

<sup>1</sup>Oak Ridge National Laboratory (ORNL); <sup>2</sup>University of Washington

#### https://pmiweb.ornl.gov/

**Goals**: The goal of the Plant-Microbe Interfaces (PMI) science focus area (SFA) is to characterize and interpret the physical, molecular, and chemical interfaces between plants and microbes and determine their functional roles in biological and environmental systems. Populus and its associated microbial community serve as the experimental system for understanding the dynamic exchange of energy, information, and materials across this interface and its expression as functional properties at diverse spatial and temporal scales. To achieve this goal, the product focuses on (1) defining the bidirectional progression of molecular and cellular events involved in selecting and maintaining specific, mutualistic Populus-microbe interfaces, (2) defining the chemical environment and molecular signals that influence community structure and function, and (3) understanding the dynamic relationship and extrinsic stressors that shape microbiome composition and affect host performance.

Plants have been co-evolving with microbes since their emergence onto land some half a billion years ago, yet

predictive understanding of how mutualistic symbioses between these diverse groups of organisms are selected and maintained is still in its infancy. Recently, considerable attention has been focused on plant growth-promoting bacteria (PGPB) for their application to agriculture, bioprotection, and phytoremediation. Such PGPB are known to stimulate plant growth through enhancing nutrient acquisition or modulating hormone levels in select crop species, but little is known about their potential impacts on woody perennials like poplar that are relevant to sustainable biofuel production. To address this knowledge gap, researchers first isolated and sequenced bacteria from the roots of field grown Populus (Blair et al. 2018; Carper et al. 2021). This culture collection represents over 3,200 unique bacterial isolates, and full genome sequences are available for over 550 of these isolates. Using comparative genomics, the team identified potential PGPB isolates that have the enzymatic machinery required for nitrogen fixation. Results showed that one of these isolates, Rahnella sp. OV588, catalyzes acetylene reduction in vitro and that this activity is dependent on the nitrogenase (nifH) enzyme. In co-culture experiments, Rahnella efficiently colonized the endosphere of axenic poplar plants and catalyzed acetylene reduction *in planta* under nitrogen-limiting conditions. Bacterial inoculation significantly increased root biomass, suggesting that the plant host can benefit from treatment with PGPB. To understand the molecular mechanisms involved in the establishment of this plant-microbe symbiosis, researchers quantified host transcriptomic response over a time course. Gene expression was affected by both time and microbial treatment, with the largest treatment effect observed at 24 hours post-inoculation. Genes induced by PGPB treatment were enriched for biological processes, including root morphogenesis, hormone metabolism, and sugar signaling, but also show signatures of an induced systemic immune response. Findings elucidate the mechanisms by which mutualistic relationships are established between poplar and members of its microbiome and highlight possible strategies to improve biofuel feedstock production in marginal habitats.

Blair, P. M., et al. 2018. "Exploration of the Biosynthetic Potential of the *Populus* Microbiome," *mSystems* **3**(5), e00045-18. DOI:10.1128/mSystems.00045-18.

Carper, D. L., et al. 2021. "Cultivating the Bacterial Microbiota of *Populus* Roots," *mSystems* **6**(3), e01306-20. DOI:10.1128/ mSystems.01306-20.

**Funding Information:** Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract No. DE-AC05-00OR22725. The Plant-Microbe Interfaces SFA is sponsored by the Genomic Science program, U.S Department of Energy, Office of Science, Biological and Environmental Research (BER) Program under FWP ERKP730.

## **SCIENCE FOCUS AREA**

## Microbes Persist: Systems Biology of the Soil Microbiome

## Functional Succession of Growing Soil Microorganisms and Virus-Driven Mortality Following Rewetting in a California Grassland Soil

Steven J. Blazewicz<sup>1\*</sup> (blazewicz1@llnl.gov), Alexa Nicolas<sup>2</sup>, Ella T. Sieradzki<sup>2</sup>, Peter Chuckran<sup>2</sup>, Katerina Estera-Molina<sup>1,2</sup>, Rohan Sachdeva<sup>2</sup>, **Mary K.** Firestone<sup>2</sup>, Jillian Banfield<sup>2,3</sup>, Jennifer Pett-Ridge<sup>1,2</sup>

<sup>1</sup>Lawrence Livermore National Laboratory; <sup>2</sup>University of California– Berkeley; <sup>3</sup>Lawrence Berkeley National Laboratory

#### https://sc-programs.llnl.gov/soil-microbiome

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. Researchers hypothesize that microbial cellular chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence and are characterizing this via stable isotope probing (SIP) of genomeresolved metagenomes and viromes. The team focuses on soil moisture as a master controller of microbial activity and mortality since altered precipitation regimes are predicted across the temperate United States. This science focus area'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.

Rewetting of soil stimulates a succession of microbial growth and mortality, a process that could potentially become more frequent as climate change in semi-arid zones is predicted to lead to fewer rain events, potentially allowing for soil drydown between events. It has been hypothesized that certain microbial traits, such as degradation of carbohydrates and acquisition of nitrogen, underlie this succession and confer advantages for growth as both the soil microbial community and available resources change over time. Researchers presume that the initial burst of mortality following wet-up is driven by osmotic lysis due to the rapid change in osmotic pressure, while continued mortality after the first few hours is driven by viruses or other biological factors. Researchers also hypothesized that the summer dry down would drive phages to integrate into host chromosomes and that wet-up of dry soil might serve as an environmental inducer of temperate phages.

To determine the mechanisms driving microbial growth and mortality during wet-up, researchers performed a wet-up experiment using soils that had been previously <sup>13</sup>CO<sub>2</sub>-labeled and maintained under one of two precipitation regimes: the historical average precipitation (100%) and a 50% water reduction. Following the annual summer dry period, soils were collected and incubated with multiple isotopic treatments. Heavy water (<sup>18</sup>O-H<sub>2</sub>O) additions were used to specifically target the growing portion of the microbiome and virome. Samples were harvested at six times following rewetting (0, 3, 24, 48, 72, and 168 hr) for DNA–quantitative stable isotope probing (qSIP), metagenomics, metatranscriptomics, viromics, and CO<sub>2</sub> production.

While total soil respiration did not vary between soils exposed to 100% versus 50% precipitation, respiration of new (labeled) rhizodeposits was higher in the 100% soils, implying functional differences between precipitation groups. This result was supported by large differences in taxon-specific responses for bacterial growth and mortality and differential abundance of traits found in growing (<sup>18</sup>O-labeled) microorganisms for the two precipitation treatments. Differential abundance of traits also revealed a large difference in the functional potential of the microbiome at the end of the dry season. Surprisingly this legacy effect disappeared after one week, indicating functional capacity converged regardless of prior conditions. Temporal changes were observed in the abundance of genes coding for carbohydrate active enzymes in growing organisms implying that substrate availability varied with time. Genes coding for synthesis and export of EPS were more abundant in growing organisms as compared to the total community in alignment with the hypothesis that this function could provide a fitness advantage during both the dry-down and wet-up.

In comparison to temporal abundance patterns in microorganisms, viruses displayed higher spatial heterogeneity in addition to temporal community changes. Quantitative isotope tracing, time-resolved metagenomics and viromic analyses indicated that dry soil held a diverse but low biomass reservoir of virions, of which only a subset thrived following wet-up. Viral richness decreased by 50% within 24 hours post wet-up, while viral biomass increased at least four-fold within one week. Counter to recent hypotheses suggesting temperate viruses predominate in soil, the team's evidence indicates that wet-up is dominated by viruses in lytic cycles. Researchers estimate that viruses drive a measurable and continuous rate of cell lysis with up to 46% of microbial death driven by viral lysis one week following wet-up. Results show that viruses contribute to a significant portion of soil microbial biomass turnover and the widely reported  $CO_2$  efflux following wet-up of seasonally dry soils.

In summary, the team observed temporal changes in growing microbial and viral communities following wet-up that were underpinned by succession of organic carbon degradation capabilities as well as by lytic infection by viruses.

**Funding Information**: This research is based upon work of the LLNL Microbes Persist Soil Microbiome science focus area supported by the U.S. Department of Energy Office of Science, Biological and Environmental Research Program Genomic Science program under Award No. SCW1632 to the Lawrence Livermore National Laboratory with a subcontract to the University of California, Berkeley. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

## Microbes Persist: Towards Quantitative Theory–Based Predictions of Soil Microbial Fitness, Interaction, and Function in KBase

Jeffrey Kimbrel<sup>1\*</sup> (kimbrel1@llnl.gov), Ulas Karaoz<sup>3</sup>, Gianna Marschmann<sup>3</sup>, Benjamin J. Koch<sup>2</sup>, Steven J. Blazewicz<sup>1</sup>, Bruce A. Hungate<sup>2</sup>, Eoin Brodie<sup>3</sup>, Jennifer Pett-Ridge<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory; <sup>2</sup>Northern Arizona University; <sup>3</sup>Lawrence Berkeley National Laboratory

#### https://sc-programs.llnl.gov/soil-microbiome

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. The team hypothesizes that microbial cellular chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and team members are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. Researchers are focusing on soil moisture as a master controller of microbial activity and mortality, since altered precipitation regimes are predicted across the temperate United States. This science focus area's (SFA) 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.

This SFA has pioneered methods that quantify element fluxes with taxonomic resolution and has proposed to

integrate these into KBase. In particular, quantitative stable isotope probing (qSIP) allows us to evaluate in situ activity of individual taxa in complex communities by adding isotope tracers such as <sup>18</sup>O-enriched heavy water or <sup>13</sup>C-enriched compounds. Researchers have refactored a computational workflow that accepts both amplicon or metagenomic sequence SIP input and calculates atom fraction excess (enrichment) as well as growth and mortality rates for individual amplicon sequence variants (ASVs) and genomes assembled from metagenomes (MAGs and viral OTUs). Experiments using <sup>18</sup>O-H<sub>2</sub>O labeling and qSIP provide critical information on organism growth rates and mortality in situ. The analytical pipelines the team is developing within KBase establish a standard qSIP analytical workflow and a qSIP database suitable for robust cross-site comparisons and for model benchmarking. The workflow will enable uniform bioinformatics and calculations of qSIP data (e.g., a uniform approach to density shift calculations) within the existing quantitative insights into microbial ecology (QIIME) platform, and the database will facilitate robust comparisons across experiments. Integration within KBase will support analyses that compare traits of organisms with their performance in nature across environments.

The qSIP pipeline is fully integrated with a genomes-totraits workflow (microTrait) and compatible with a dynamic energy budget-based trait-based model (DEBmicroTrait). With microTrait and DEBmicroTrait, team members have developed and tested a computational workflow to (1) infer ecologically relevant traits from microbial genomes, (2) systematically reduce the high-dimensionality of genome-level microbial trait data by inferring functional guilds (sets of organisms performing the same ecological function irrespective of their phylogenetic origin), (3) quantify within-guild trait variance and capture trait linkages in trait-based models, (4) explore trait-based simulations under different scenarios with varying levels of microbial community and environmental complexity, and (5) benchmark emergent model substrate utilization (digested as chemical abundance data) and qSIP-derived growth and mortality rates (from qSIP database).

Ongoing work to combine both the qSIP and DEBmicro-Trait tools within KBase will provide a strong foundation for researchers who wish to use quantitative *in situ* measurements of microbial ecophysiology and population dynamics to benchmark models and build a predictive understanding of biological processes controlling material fluxes in complex environments.

**Funding Information:** This research is based upon work of the LLNL Microbes Persist Soil Microbiome SFA supported by the U.S. Department of Energy Office of Science, Biological and Environmental Research Program Genomic Science program under Award No. SCW1632 to the Lawrence Livermore National Laboratory, and subcontracts to Northern Arizona University and Lawrence Berkeley National Laboratory. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

## The Path from Root Input to Mineral-Associated Soil Carbon is Dictated by Habitat-Specific Microbial Traits

Noah W. Sokol<sup>1</sup>, Megan M. Foley (mmf289@nau.edu)<sup>2\*</sup>, Amrita Battacharyya<sup>3,4</sup>, Nicole DiDonato<sup>5</sup>, Katerina Estera-Molina<sup>1,6</sup>, Alex Greenlon<sup>6</sup>, Jeffrey Kimbrel<sup>1</sup>, Jose Liquet<sup>1</sup>, Marissa Lafler<sup>1</sup>, Maxwell Marple<sup>1</sup>, Peter S. Nico<sup>3</sup>, Ljiljana Pasa-Tolic<sup>5</sup>, Eric Slessarev<sup>1</sup>, **Mary K. Firestone**<sup>6</sup>, **Bruce A. Hungate<sup>2</sup>**, **Steven J. Blazewicz<sup>1</sup>**, **Jennifer Pett-Ridge**<sup>1,7</sup>

<sup>1</sup>Lawrence Livermore National Laboratory; <sup>2</sup>Northern Arizona University; <sup>3</sup>Lawrence Berkeley National Laboratory; <sup>4</sup>University of San Francisco; <sup>5</sup>Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory; <sup>6</sup>University of California– Berkeley; <sup>7</sup>University of California–Merced

#### https://sc-programs.llnl.gov/soil-microbiome

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. The team hypothesizes that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and researchers are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. The team is focusing on soil moisture as a master controller of microbial activity and mortality since altered precipitation regimes are predicted across the temperate United States. This science focus area's ultimate goal is to determine how microbial soil ecophysiology, population dynamics, and microbe-mineralorganic matter interactions regulate the persistence of microbial residues under changing moisture regimes.

Soil microorganisms influence the global carbon balance by transforming plant inputs into mineral-associated organic matter (MAOM), but which microbial traits control mineral-associated SOC storage is widely debated. While current theory and biogeochemical models have settled on microbial carbon-use efficiency and growth rate as positive predictors of mineral-associated SOC accrual, empirical tests are sparse and show contradictory observations. To investigate the relationship between different microbial traits and MAOM, researchers conducted a 12- week  $^{13}$ C tracer study to track the movement of rhizodeposits and root detritus into microbial communities and SOM pools under moisture-replete (15 ± 4.2%) or water-limited

 $(8 \pm 2\%)$  conditions. Using a continuous <sup>13</sup>CO<sub>2</sub>-labeling growth chamber system, researchers grew the annual grass *Avena barbata* for 12 weeks and measured formation of <sup>13</sup>C-MAOM from either <sup>13</sup>C-enriched rhizodeposition or decomposing <sup>13</sup>C-enriched root detritus. The team also measured active microbial community composition (via <sup>13</sup>C-quantitative stable isotope probing; qSIP) a suite of microbial traits including carbon-use efficiency, growth rate, and turnover (via the <sup>18</sup>O-H2O method), extracellular enzyme activity, bulk <sup>13</sup>C-extracellular polymeric substances (EPS), and total microbial biomass carbon (<sup>13</sup>C-MBC), as well as chemical composition of MAOM via <sup>13</sup>C-nuclear magnetic resonance (NMR) and Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS).

In the microbial habitat around living roots (rhizosphere), the activity of bacterial-dominated communities with fast growth, high biomass, and high production of extracellular polymeric substances were positively associated with the accrual of <sup>13</sup>C-mineral-associated SOC under normal moisture conditions. However, under drought, the rhizosphere and the microbial habitat around decaying roots (detritusphere) had more fungal-dominated communities positively associated with <sup>13</sup>C-mineral associated SOC with slower growth, lower carbon-use efficiency, and higher exoenzyme activity. <sup>13</sup>C-qSIP revealed that bacterial taxa from the families Bacillaceae, Bradyrhizobiaceae, and Comamonadaceae were particularly active in the rhizosphere, whereas filamentous fungi (families Ceratostomataceae, Lasiosphaeriaceae, and Pleosporaceae) were dominant decomposers in the detritusphere. FTICR-MS and <sup>13</sup>C-NMR indicated that rhizosphere MAOM has a higher O/C ratio than the detritusphere as well as having a greater amount of lipids and carbohydrates, whereas the detritusphere had a greater abundance of lignin-like compounds. Together, this suggests a more microbial-processed signature of MAOM in the rhizosphere and a more plant-derived signature in the detritusphere.

Overall, these findings emphasize that microbial traits linked with SOC storage vary with soil habitat and moisture conditions—a fact that emerging SOC models should explicitly reflect, since living versus decaying root ratios and moisture regimes will shift under a changing climate.

**Funding Information**: This research is based upon work of the LLNL Microbes Persist Soil Microbiome science focus area supported by the U.S. Department of Energy Office of Science, Biological and Environmental Research Program Genomic Science program under Award No. SCW1632 to the Lawrence Livermore National Laboratory, and subcontracts to the Northern Arizona University and the Pacific Northwest National Laboratory. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.

## OTHER

## The Twin Ecosystems Project

## The Twin Ecosystems Project: A New Capability for Field and Laboratory Ecosystems Coupled by Sensor Networks and Autonomous Controls

Sheryl L. Bell<sup>1</sup>, Elle M. Barnes<sup>2,3</sup>, Jonathan J. Silberg<sup>4</sup>, Caroline Masiello<sup>4</sup>, Jiwoo Kim<sup>4</sup>, Li Chieh Lu<sup>4</sup>, Peter F. Andeer<sup>2</sup>, Peter Zwart<sup>2</sup>, Marcus Noack<sup>2</sup>, Daniela Ushizima<sup>2</sup>, James Sethian<sup>2,5</sup>, Kirsten S. Hofmockel<sup>1</sup>, Susannah G. Tringe<sup>2,3\*</sup> (sgtringe@lbl.gov), **Trent R. Northen<sup>2, 3</sup>** 

<sup>1</sup>Pacific Northwest National Laboratory (PNNL), <sup>2</sup>Lawrence Berkeley National Laboratory (LBNL), <sup>3</sup>DOE Joint Genome Institute, <sup>4</sup>Rice University, <sup>5</sup>University of California–Berkeley

**Goals**: The goal of the TWIN ecosystem project (TWINS) is to pilot self-driving lab, positron emission tomography (PET), microbial biosensors, and laboratory and field twin ecosystems to gain insights into above- and belowground plant dynamics and interactions. Autonomous experiments are used to gain novel insights into grass responses to nutrient stress, and PET is being used to collect hot spots for omics analyses. These efforts are being used to study compositional changes in root exudates and rhizosphere communities following harvest. Here the field twin defines the climate conditions for the lab twin—providing powerful environmental controls and measurements, which are essentially not possible in the field.

This project has integrated computer vision software and autonomous experimental design software (gpCAM) developed by the Center for Advanced Mathematics for Energy Research Applications (CAMERA) with an automated experimental system for performing fabricated ecosystem experiments (the EcoBOT). Three rounds of experiments have now been performed to map the nutritional landscape of the JGI flagship model grass, Brachypodium distachyon, and the hyperspectral signatures of plant combinations of nutrient stresses. The resulting model will be a valuable tool in interpreting ongoing remote multispectral image data from a field site in Prosser, Washington. At this field site, researchers have leveraged an existing field experiment to define climate conditions for a controlled laboratory twin that replicates field conditions for plants that have been transplanted from the field into large-scale mesocosm environments (EcoPOD, the lab twin).

The laboratory twin enables detailed control and characterization of the composition and dynamics of microbes and exudates under baseline conditions and in response to perturbation. In a pilot experiment, TWINS is investigating how plant biomass harvest alters the soil microbial community structure and function in response to tall wheatgrass (Thinopyrum ponticum) exudates. Samples were collected from both the field and the EcoPOD under twinned environmental conditions in July 2022, three days prior to harvest of the plant's aboveground biomass and again 3 days after the harvest. Bulk and rhizosphere soils were isolated and extracted for nucleic acids and polar metabolites. Amplicon sequencing of the 16S V3–V4 region was used to investigate microbial community responses to plant harvest. Root, rhizosphere, and bulk soil samples have been analyzed using liquid chromatography-tandem mass spectrometry including both reverse phase and hydrophilic liquid interaction chromatography. Untargeted metabolomic analysis has been used to compare thousands of unique chemical features to identify changes in root and rhizosphere metabolites following harvest and to compare these findings between the lab and field twins.

Engineered microorganisms could potentially be used to monitor in situ processes in fabricated ecosystems. To investigate whether information can be transmitted across a soil using a rare volatile metabolite researchers evaluated whether one member of the Model Soil Consortium-2 (MSC-2; Variovorax) can be programmed to produce a unique methyl halide signal. By expressing a methyl halide transferase in Variovorax, researchers showed that this microbe can synthesize a signal that is more than 100-fold higher than the basal level produced by other soil microbes in MSC-2 including Dyadobacter, Ensifer, Rhodococcus, and Streptomyces. The signal generated by Variovorax could be read out directly using gas chromatography or using a Methylorubrum biosensor that produces a fluorescent output. Using soil habitats ranging in size from 1 to 50 grams of soil researchers found that methyl halide cell-cell signaling could be achieved under environmentally relevant water holding conditions. Synthetic methyl halide signaling is expected to simplify fundamental studies of gene expression in hard-to-image materials containing microbiomes, and it should be useful for programming soil consortia to convert information sensed in subterranean settings into overt aboveground visual or gas signals in fabricated ecosystems.

**Funding Information:** The team gratefully acknowledges funding from the U.S. Department of Energy (DOE) Office of Science, Biological and Environmental Research Program, Genomic Science Program under contract No. DE-AC02-05CH11231. Work at PNNL and LBNL was performed under FWPs 78814 and FP00013570, respectively. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.



## **UNIVERSITY PROJECTS**

## Scalable Computational Tools for Inference of Protein Annotation and Metabolic Models in Microbial Communities

Saeedeh Davoudi<sup>1\*</sup> (saeedeh.davoudi@ucdenver.edu), Janaka N. Edirisinghe<sup>2\*</sup> (janakae@anl.gov), Mikayla A. Borton<sup>3</sup> (mikayla.borton@pnnl.gov), Michael Shaffer<sup>3</sup>, Zahmeeth Sakkaff<sup>2</sup>, Filipe Liu<sup>2</sup>, Rory Flynn<sup>3</sup>, Lucia S. Guatney<sup>1,4</sup>, Derick Singleton<sup>1</sup>, James Stegen<sup>5</sup>, Byron C. Crump<sup>6</sup>, James J. Davis<sup>2</sup>, Farnoush Banaei-Kashani<sup>1</sup>, Kelly C. Wrighton<sup>3</sup>, Christopher S. Henry<sup>2</sup>, **Christopher S. Miller<sup>1</sup>** 

<sup>1</sup>University of Colorado–Denver; <sup>2</sup>Argonne National Laboratory; <sup>3</sup>Colorado State University; <sup>4</sup>University of Colorado–Anschutz Medical Campus; <sup>5</sup>Pacific Northwest National Laboratory; <sup>6</sup>Oregon State University

https://www.kbase.us

#### https://github.com/WrightonLabCSU/DRAM

Goals: High-throughput omics technologies have made the assembly of microbial genomes recovered from the environment routine. Computational inference of the protein products encoded by these genomes, and the associated biochemical functions, should allow for the accurate prediction and modeling of microbial metabolism, organismal interactions, and ecosystem processes. However, a lack of scalable, probabilistic protein annotation tools limits the full potential of metabolic modeling. The approach to inference of improved models relies on developing new computational tools in three main areas: (1) improved protein annotations, (2) iterative cycles of gap-filling metabolic models with improved protein annotations and informing probabilistic protein annotations based on metabolic models, and (3) integrating improved protein annotations with community-level flux balance metabolic models. Researchers aim to make these tools broadly accessible via the DOE Systems Biology Knowledgebase (KBase; Arkin et al. 2018).

In the past year researchers continued to improve the genome annotation and modeling capabilities emerging from this project. Researchers have achieved this via advancement of traditional homology-based approaches, and advancement of new approaches leveraging genomescale metabolic models and approaches from the field of natural language processing.

The DRAM (Shaffer et al. 2020) app in KBase was significantly improved, enhancing reliability, usability, and overall quality of output annotations, all within the KBase framework that allows for interoperability of these annotations with other annotation and modeling apps. This work culminated in a new publication highlighting the utility of DRAM within KBase (Shaffer et al. submitted). The team also extended DRAM to be able to predict microbial traits (e.g., nitrate reducer, aerobe, fermenter) from protein annotations. These traits were developed and validated (below) via extensive expert curation.

Beyond DRAM, researchers also deployed a new tool called GLM4EC in which researchers trained and fine-tuned a modification of Generalized Language Models (GLMs; also known as Large Language Models) based on ProteinBERT (Brandes et al. 2022) to the task of annotation of microbial proteins with Enzyme Commission (E.C.) numbers. This model was trained to learn sequence embeddings and annotation classification from a subset of UniRefKB annotated with E.C. numbers. On held-out test sets, the model predicts E.C. numbers with high precision and recall, regardless of input sequence length. Researchers are now testing new models that have additional global features beyond E.C. numbers available for pretraining and exploring the utility of alternative model architectures. Integration of GLM4EC into KBase, along with the improved version of DRAM and other existing annotation pipelines, provides multiple hypotheses for the function of gene products within genomes and metagenome assembled genomes (MAGs) in KBase, all of which can be integrated or explored in a common, interoperable framework with other KBase tools.

To aid in determining which of the potentially alternative functions a gene product actually performs, researchers developed machine learning classifiers to predict growth phenotypes based on multiple functional annotations. These classifiers are now loaded into KBase, along with apps enabling them to be applied to predict phenotypes for any KBase genome or MAG. Further, in collaboration with the Hoffmockel science focus area (SFA), researchers have integrated apps in KBase that reconcile metabolic models with predicted phenotypes, using the alternative functions proposed by the new annotation apps to associate gene candidates with gapfilled reactions. This leads to dramatic improvements in model accuracy from 56% to 72% on average. This also leads to numerous corrected annotations across all genomes, particularly in MAGs where many genes and associated functions are often missing due to incomplete assemblies. Combined with the recent enhancements to the ModelSEED pipeline to improve energy biosynthesis prediction in methanogens (a key target for this project), researchers now have models with greatly improved accuracy to predict both core and periphery metabolism.

With all these components in place, researchers are applying this improved pipeline to build models for over 2,000 MAGs loaded into KBase from the Genome Resolved Open Watersheds (GROW) project. The MAG models from this analysis now have many more reactions and annotations, particularly for poorly annotated clades. Researchers are now assembling these MAG models into compartmentalized community metabolic models for each of the 178 GROW samples. Researchers are loading hand-curated microbial traits (in part aided by DRAM inference) and corresponding trophic interaction networks into KBase as phenotypes, enabling the gapfilling of community models to replicate these hypothesized expert-curated trophic webs. The degree of gapfilling required to replicate trophic webs provides valuable feedback for identification of potential errors in these webs. The resulting models can also be tested against a growing collection of metatranscriptome data gathered from these samples, measuring agreement between reaction flux and associated gene expression.

Arkin, A. P., et al. 2018. "KBase: The United States Department of Energy Systems Biology Knowledgebase," *Nature Biotechnology* **36**, 566–69.

Brandes N, et al. 2022. "ProteinBERT: A Universal Deep-Learning Model of Protein Sequence and Function," *Bioinformatics* **38**(8), 2102–10.

Shaffer, M., et al. 2020. "DRAM for Distilling Microbial Metabolism to Automate the Curation of Microbiome Function," *Nucleic Acids Research* **48**, 8883–900.

Shaffer, M., et al. Submitted. "kb\_DRAM: Distilled Genome Annotations and Metabolic Modeling in KBase," *Bioinformatics* **39**(4), btad110. DOI:10.1093/bioinformatics/btad110.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0021350, and the DOE Joint Genome Institute Community Science Program. KBase was supported by Award Nos. DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

## HypoRiPPAtlas: An Atlas of Hypothetical Natural Products for Mass Spectrometry Database Search

Yi-Yuan Lee<sup>1,2</sup>, Mustafa Guler<sup>1\*</sup> (mguler@andrew.cmu. edu), Desnor N. Chigumba<sup>3</sup>, Shen Wang<sup>1</sup>, Neel Mitta<sup>1</sup>, Cameron Miller<sup>1</sup>, Benjamin Krummenacher<sup>1</sup>, Haodong Liu<sup>1</sup>, Liu Cao<sup>1</sup>, Aditya Kannan<sup>1</sup>, Keshav Narayan<sup>1</sup>, Samuel T. Slocum<sup>4</sup>, Bryan L. Roth<sup>4</sup>, Alexey Gurevich<sup>5</sup>, Bahar Behsaz<sup>1</sup>, Roland D. Kersten<sup>3</sup>, **Hosein Mohimani<sup>1</sup>** 

<sup>1</sup>Carnegie Mellon University; <sup>2</sup>Cornell University; <sup>3</sup>University of Michigan; <sup>4</sup>University of North Carolina; <sup>5</sup>St. Petersburg State University

#### https://hyporippatlas.npanalysis.org/

Recent analysis of hundreds of thousands of public microbial genomes has resulted in the discovery of over a million biosynthetic gene clusters (BGCs; Hadjithomas et al. 2016; Blin et al. 2016; Kautsar et al. 2021). Gene-to-molecule approaches are therefore urgently needed for microbial and plant natural product (NP) discovery in light of rapidly growing microbial and plant genetic resources. Currently, the NPs for the majority of BGCs remain unknown. Global natural product social (GNPS) molecular networking infrastructure harbors billions of mass spectra of NPs with unknown structures and biosynthetic genes. In order to bridge the gap between large-scale genome mining and mass-spectral datasets for NP discovery, researchers developed HypoRiPPAtlas, an atlas of hypothetical NP structures, which can be readily used for *in silico* database search of tandem mass spectra.

HypoRiPPAtlas is constructed by mining the genomes of 22,671 microbial strains from the RefSeq database using seq2ripp, a novel machine learning tool for prediction of ribosomally synthesized and post-translationally modified peptides (RiPPs). Seq2ripp outperforms currently existing RiPP mining tools in identification of known MiBIG RiPPs from genomic inputs. Searching the hypothetical molecules from the Atlas against 46 mass spectral datasets from GNPS resulted in the discovery of numerous RiPPs, including two novel lassopeptides and one lanthipeptide from Streptomyces sp. NRRL B-2660, WC-3904 and WC-3560. Moreover, seq2ripp discovered ten plant RiPPs including elaeagnin, a member of a new BURP-domainderived RiPP class with a novel post-translational modification (PTM) from silverberry (Elaeagnus pungens). By addressing the fundamental challenge of predicting structures from NP biosynthetic genes, the HypoRiPPAtlas approach has therefore the potential to close the gap between biosynthetic genes and their natural products in genomic NP discovery, which could be extended to other NP classes in the future by implementing corresponding biosynthetic logic.

HypoRiPPAtlas and the seq2ripp pipeline are both publicly available at hyporippatlas.npanalysis.org. Users can examine hypothetical RiPPs mined from publicly available genomes and upload their own paired genomic and mass spectral datasets to launch custom seq2ripp runs.

Blin, K., et al. 2016. "The AntiSMASH Database, a Comprehensive Database of Microbial Secondary Metabolite Biosynthetic Gene Clusters," *Nucleic Acids Research* **45**, D555–9.

Hadjithomas, M. et al. 2016. "IMG-ABC: New Features for Bacterial Secondary Metabolism Analysis and Targeted Biosynthetic Gene Cluster Discovery in Thousands of Microbial Genomes," *Nucleic Acids Research* **45**, D560–5.

Kautsar, S. A., et al. 2021. "BiG-SLiCE: A Highly Scalable Tool Maps the Diversity of 1.2 Million Biosynthetic Gene Clusters," *GigaScience* **10**(1), giaa154.

**Funding Information:** This research was supported by U.S. Department of Energy award DE-SC0021340.

### Predicting and Modeling Protein-Protein Complexes at Large-Scale with Deep Learning

Mu Gao<sup>1\*</sup> (mu.gao@gatech.edu), Davi Nakajima An<sup>1</sup>, Jerry M. Parks<sup>2</sup>, **Jeffrey Skolnick**<sup>1</sup>

#### <sup>1</sup>Georgia Institute of Technology; <sup>2</sup>Oak Ridge National Laboratory

Goals: With the advances in next-generation sequencing technologies, the number of sequenced genomes is growing exponentially. This has resulted in a bottleneck for the translation of sequence information into functional hypotheses about each gene. Current gene annotation technologies are primarily based on evolutionary inference by sequence comparison; however, many proteins in a proteome remain uncharacterized. To address this challenge, this collaborative team is developing a suite of novel high-performancecomputing (HPC), deep-learning methods that predict protein structures and interactions at unprecedented accuracy, making use of the Summit supercomputer at the DOE leadership computing facility at the Oak Ridge National Laboratory. The combination of deep learning, HPC, and structural-based analysis will help to understand molecular mechanisms of protein functions, and enable rapid, accurate prediction of gene function on a genomic scale, such as novel protein-protein interactions important to life.

One key observation of proteins in a living cell is that they usually interact with each other to carry out their biological functions. The identification and characterization of these protein-protein interactions are therefore critical to understanding life. Very recently, researchers proposed a deep learning-based approach for the identification of protein-protein interactions. The approach, AF2Complex, is built on the success of AlphaFold 2. AF2Complex extends the idea of structure modeling of a single protein sequence to a complex made of multiple sequences and further predicts protein-protein interactions by using the confidence of its structural modeling. While AF2Complex have been successfully benchmarked in multiple tests including 7,000 protein pairs from the bacteria E. coli, it is important to demonstrate its usefulness by applying it to address some real-world problems. For this purpose, team members investigated the pathway leading to the folding and assembly of outer membrane proteins (OMP) in *E. coli* as a proof-of-concept illustration of approach. OMPs serve an essential functional role such as nutrients exchange with their living environment. The making of these barrel-like OMP proteins is an elaborate process starting within the cytoplasm, where they are first manufactured by ribosomes. Coming out of the ribosomes are nascent, still largely unfolded peptide chains that must subsequently cross the inner membrane, travel through the periplasmic space, and finally land at their destination: the outer membrane. To ensure a successful journey, many other



**Fig. 17.** Super protein complexes in the outer membrane biogenesis pathway in *E. coli* identified and modeled by screening the envelopome of *E. coli* with AF2Complex. [Courtesy Mu Gao]

proteins provide vital help by forming functional protein complexes. However, they are challenging for experimental characterization because many of them are membrane proteins, and the interactions are often transient. By applying the AF2Complex workflow established at Summit to several key proteins in the OMP biogenesis pathway, researchers have identified their functional partners within the top 1% ranking of ~1,500 proteins screened for PPIs per query. Thanks to high confidence structures underlying the top predictions, one can understand many experimental phenomena, particularly in vivo site-directed photo cross-linking data. For example, cross-linked products from the translocon SecYEG or the β-barrel assembly machine (BAM) supercomplexes may be explained by direct physical interactions revealed in predicted structures (see Fig. 17, this page). An unexpected, biologically important interaction has been identified between the enzyme DsbA and chaperon PpiD, which is associated with the SecYEG translocon. Moreover, previously speculated conformations are captured for SurA and BepA. Most importantly, these revealing atomic structures of various supercomplexes suggest mechanistic hypotheses for various steps of the OMP biogenesis pathways.

Gao, M., et al. 2022. "AF2Complex Predicts Direct Physical Interactions in Multimeric Proteins," *Nature Communications* 13, 1744. https://www.nature.com/articles/s41467-022-29394-2.

Gao, M., et al. 2022. "Deep Learning-Driven Insights into Super Protein Complexes for Outer Membrane Protein Biogenesis in Bacteria," *eLife* **11**, e82885. https://elifesciences.org/articles/82885.

**Funding Information:** This research was supported by the DOE Office of Science, Biological and Environmental Research (BER) Program, grant No. DE-SC0021303, and the Advanced Scientific Computing Research (ASCR) Leadership Computing Challenge (ALCC) program. The research used resources provided by the Leadership Computing Facility at Oak Ridge National Laboratory, National Energy Research Scientific Computing Center at Berkeley, and the Partnership for an Advanced Computing Environment (PACE) at Georgia Institute of Technology.

## Developing a National Virtual Biosecurity for Bioenergy Crops Center

Paul Freimuth\* (Freimuth@bnl.gov), Kerstin Kleese van Dam, John Shanklin, Alistair Rogers, Robert McGraw, Qun Liu, Shantenu Jha, **Martin A. Schoonen** 

#### **Brookhaven National Laboratory**

Goals: The goal of this 18-month pilot project is to develop a roadmap for a new U.S. Department of Energy (DOE) Office of Science Biological and Environmental Science capability to address biothreats to bioenergy crops. The main deliverable of this effort is a roadmap toward a new National Virtual Biosecurity for Bioenergy Crop Center (NVBBCC) based on community input as well as the experiences of conducting a limited study on anthracnose, a disease affecting Sorghum, one of DOE's leading biofuels feedstocks. A mature NVBBCC is envisioned to be a distributed, virtual center with multiple DOE laboratories at its core to maximize the use of unique facilities and expertise across the DOE complex. NVBBCC will support community-driven plant pathology research as well as broader BER-relevant plant biology research. The new NVBBCC capability could also support responses to biothreats to unmanaged ecosystems and techniques, workflows, and infrastructure could be readily pivoted to a wider range of biosecurity challenges.

Rationale: The development of resilient and sustainable bioenergy crops is an important part of U.S. government strategy to transition to a net-zero economy. An important consideration in developing the U.S. bioeconomy is the biosecurity of crops grown for bioenergy production. The most likely biosecurity threats to bioenergy crops are either known pests or pathogens that emerge in new areas, possibly driven by climate change or new pests or pathogens that are genetically related to known ones. A robust biosecurity capability optimized to respond rapidly to biothreats to bioenergy crops requires an integrated and versatile platform that delivers rapid detection and targeted sampling, propagation prediction, and timely characterization of the interaction between a pest or pathogen and the bioenergy crop. These capabilities are needed to underpin the development of controls and solutions. Here the team reports on a new pilot study funded by DOE to develop a roadmap for a National Virtual Biosecurity for Bioenergy Crops Center (NVBBCC) organized around four interconnected modules: detection and sampling, biomolecular characterization, assessment, and mitigation.

**Approach:** The team will use a series of community planning meetings and experimental work on a known disease

in Sorghum to develop a roadmap for the development of NVBBCC. The roadmap planning meetings, conducted within the first nine months, will identify partnerships within and outside DOE necessary to establish the full capability required for an end-to-end biosecurity platform and develop a network of experts and facilities that the center can draw upon when faced with a biothreat. A study on a fungal disease, anthracnose, which affects Sorghum, a leading energy crop, will be used to develop material, experimental, and data workflows as well as guide future investments. Anthracnose is caused by Colletotrichum sublineola and can lead to significant reductions by up to 67% in crop yield (Stutts and Vermerris 2020). The team has established the capability to work with the disease to study pathogen-host interactions. Biomolecular imaging capabilities at Brookhaven National Laboratory will be used to advance understanding of the interaction of *C. sublineola* with Sorghum.

Informed by lessons learned from DOE's National Virtual Biotechnology Laboratory (NVBL), the team will develop a dedicated computing platform to support the NVBBCC pilot study. Its backbone is an integrated and flexible computational science software and hardware system to support persistent data storage; advanced AI/ML-enabled data analysis; data fusion; data sharing; and near-real-time visualizations of the geographic localization of disease as well as computational simulations and predictions. The pilot study intends to stand up an initial prototype of a computational platform that can be scaled up if needed. A separate contribution to this meeting will focus on that component of the pilot study.

**Results:** The presentation will be a status review of the pilot study and a preliminary report from two community planning meetings that will have been held by mid-April. One planning meeting, held in February, will be focused on biomolecular characterization of biothreats to bioenergy crops. A second one will focus on atmospheric dispersion pathways of diseases relevant to bioenergy crops.

Stutts, L. R. and W. Vermerris, 2020. "Elucidating Anthracnose Resistance Mechanisms in Sorghum—A Review," *Phytopathology* **110**(12), 1863–76. DOI:10.1094/PHYTO-04-20-0132-RVW.

**Funding Information:** Brookhaven National Laboratory is managed by Brookhaven Science Associates, LLC for the U.S. Department of Energy under contract No. DE-SC0012704. This program is supported by the U.S. Department of Energy, Office of Science, through the Biological Systems Science Division in the Biological and Environmental Research Program, under FWP BI207.



## **DOE JOINT GENOME INSTITUTE**

# Quantify the Impact of Your Data with DOE JGI's Genome Citation Service

Neil Byers, Chris Beecroft, Charles Parker, Kjiersten Fagnan\* (kmfagnan@lbl.gov)

Joint Genome Institute

http://jgi.doe.gov

**Goal:** Development of a new resource that deepens the community's understanding of the impact of data reuse.

Steady increases in sequencing capacity, combined with rapid accumulation of publications and associated resources, have increased the complexity of maintaining associations between literature and genomic data. Accumulated errors and omissions in the literature and databases compound the difficulty of the task. Automated approaches to maintaining and confirming associations among these resources have become necessary.

Here researchers present the U.S. Department of Energy (DOE) Joint Genome Institute's (JGI) Genome Citation Service (GCS), which discovers literature that incorporates genome data whether or not the source of the data was properly attributed by the authors. This service provides a number of advantages over manual curation including consistent coverage of public resources, automatic updating of genome project metadata, and augmentation of genome project metadata through documentation of previously unrecognized uses by the scientific community. The service significantly reduces labor costs associated with manual literature review while improving the quality, accuracy, and consistency of genome metadata maintained by the DOE JGI.

The DOE JGI seeks to deepen its understanding of the impact of its user community's science by connecting its data products to publications. The GCS facilitates this understanding, improves credit attribution for data generators, and can encourage data sharing by allowing scientists to see how reuse amplifies the impact of their original studies. Doing so supports JGI's commitment to findable, accessible, interoperable, and reusable (FAIR) data practices and allows JGI to meet its obligations as a DOE Office of Science Public Reusable Research (SC PuRe) data resource.

The GCS increases the number of known publications that incorporate JGI data products and, as a publicly available resource, the GCS enables researchers to better understand their impact. Researchers seek feedback from the Genomic Science program (GSP) community on the usability of this resource.

**Funding Information**: This research was conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231.

### RNA Phages: Under-Estimated Players in Soil Ecosystems?

Clement Coclet<sup>1,2</sup>, Maureen Berg<sup>1</sup>, Roniya Thapa Magar<sup>2</sup>, Shi Wang<sup>2</sup>, Patrick O. Sorensen<sup>2</sup>, Ulas Karaoz<sup>2</sup>, Vivek K. Mutalik<sup>2</sup>, Eoin Brodie<sup>2,3</sup>, Emiley A. Eloe-Fadrosh<sup>1,2</sup>, **Simon Roux<sup>1,2\*</sup>** (sroux@lbl.gov)

<sup>1</sup>DOE Joint Genome Institute; <sup>2</sup>, Lawrence Berkeley National Laboratory; <sup>3</sup>University of California–Berkeley

**Goals**: The overarching goals of this project are to establish an analytical and experimental framework for comprehensive characterization of viral-driven alteration of microbial metabolisms in soil. The specific results presented here focus on the unexpected diversity of RNA phages detected in soil microbiomes, reveal their specific activity patterns and likely impact on bacterial lysis rate in a model soil ecosystem, and highlight several potential avenues to further characterize these soil RNA phages and their impact on microbiome processes.

Bacteriophages are now recognized as key regulators of microbial communities and processes in virtually all ecosystems, from the human gut to the global oceans. The overwhelming majority of phages described and studied so far, either via cultivation or metagenomics approaches, are double-stranded DNA phages with head-tail virion morphology, i.e., "tailed phages" from the *Caudoviricetes* class. In contrast, RNA-based bacteriophages are rarely reported or isolated and are typically not considered as important components of environmental microbiomes.

Here, the team combined large-scale data mining and time-series analyses to highlight the unsuspected diversity and potential ecological importance of RNA phages in soils. As part of a global survey of RNA viruses across more than 3,500 metatranscriptomes, researchers identified more than 80,000 potential RNA phages, primarily related to the known leviviruses but also including at least eight proposed new families and genera across several phyla (Neri et al. 2022). This global survey also identified a clear enrichment for leviviruses in wastewater, soil, and rhizosphere samples, suggesting these ecosystems are the primary reservoirs of novel RNA phage diversity. Through the specific analysis of a multiomics time series from East River (Colorado) watershed soils, researchers now demonstrate that RNA phage populations in these soils are highly diverse and follow similar activity levels and patterns as dsDNA phages. In particular, team members have observed an increase in RNA phage activity throughout the plant growing season, concomitant with an expected increase in microbial activity. Given the absence of known and predicted lysogenic cycles for RNA phages, they may contribute substantially to the overall bacterial cell lysis and nutrient cycling in the East River watershed, although their exact host range and infection dynamics remains to be characterized. Finally, based on these new RNA phages genomes, researchers identified novel protein families potentially associated with alternative mechanisms for host cell attachment and lysis. Ongoing computational and experimental characterization of these new attachment and lysis proteins should provide further insights in the potential host range of these RNA phages and possibly reveal new molecules of biotechnological interests such as new single-gene lysis proteins.

Taken together with other recent surveys and studies (Callanan et al. 2018), these results suggest that RNA phages should be more broadly considered and included in viral ecology studies, especially in soils; and they represent promising sources of novel genes and molecules for biotechnological applications.

Callanan, J., et al. 2018. "RNA Phage Biology in a Metagenomic Era," Viruses **10**, 386. DOI:10.3390/v10070386. Neri, U., et al. 2022. "Expansion of the Global RNA Virome Reveals Diverse Clades of Bacteriophages," *Cell* **185**, 4023–37, e18. DOI:10.1016/j.cell.2022.08.023.

Funding Information: This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Early Career Research Program awarded under UC-DOE Prime Contract DE-AC02-05CH11231. The work conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. This work was supported in part by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231. A portion of this research was performed under the Facilities Integrating Collaborations for User Science (FICUS) program and used resources at the DOE Joint Genome Institute and the Environmental Molecular Sciences Laboratory (grid.436923.9), which are DOE Office of Science user facilities. Both facilities are sponsored by the Biological and Environmental Research program and operated under Contract Nos. DE-AC02-05CH11231 (JGI) and DE-AC05-76RL01830 (EMSL). This material is based upon work supported as part of the Watershed Function science focus area funded by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award No. DE-AC02-05CH11231.

## KBASE: DOE SYSTEMS BIOLOGY KNOWLEDGEBASE

## **KBase Science and Infrastructure Updates**

Benjamin Allen<sup>3</sup>, Jason Baumohl<sup>1</sup>, Kathleen Beilsmith<sup>2</sup>, David Dakota Blair<sup>4</sup>, John-Marc Chandonia<sup>1</sup>, Dylan Chivian<sup>1</sup>, Zachary Crockett<sup>3</sup>, Ellen G. Dow<sup>1</sup>, Meghan Drake<sup>3</sup>, Janaka N. Edirisinghe<sup>2</sup>, José P. Faria<sup>2</sup>, Jason Fillman<sup>1</sup>, Tianhao Gu<sup>2</sup>, A. J. Ireland<sup>1</sup>, Marcin P. Joachimiak<sup>1</sup>, Sean Jungbluth<sup>1</sup>, Roy Kamimura<sup>1</sup>, Keith Keller<sup>1</sup>, Dan Klos<sup>2</sup>, Miriam Land<sup>3</sup>, Filipe Liu<sup>2</sup>, Erik Pearson<sup>1</sup>, Gavin Price<sup>1</sup>, Priya Ranjan<sup>3</sup>, William Riehl<sup>1</sup>, Boris Sadkhin<sup>2</sup>, Samuel Seaver<sup>2</sup>, Alan Seleman<sup>2</sup>, Gwyneth Terry<sup>1</sup>, Sumin Wang<sup>1</sup>, Pamela Weisenhorn<sup>2</sup>, Ziming Yang<sup>4</sup>, Shinjae Yoo<sup>4</sup>, Qizhi Zhang<sup>2</sup>; Shane Canon<sup>1</sup> (scanon@lbl.gov), Paramvir S. Dehal<sup>1</sup>, Elisha Wood-Charlson<sup>1\*</sup> (elishawc@lbl.gov), Robert Cottingham<sup>3</sup>, Christopher S. Henry<sup>2</sup>, **Adam P. Arkin<sup>1</sup>** 

<sup>1</sup>Lawrence Berkeley National Laboratory; <sup>2</sup>Argonne National Laboratory; <sup>3</sup>Oak Ridge National Laboratory; <sup>4</sup>Brookhaven National Laboratory

#### https://www.kbase.us

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

**Science Updates.** Science focus areas (SFAs) and university collaborators have been testing and releasing new data and functionality in KBase, especially around improving genome quality, functional prediction of microbial communities, and making data and tools accessible to everyone.

The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) SFA is integrating longread sequencing and isolate polishing tools into KBase and is collaborating with KBase to host a training workshop on laboratory and bioinformatics methods that support the generation of high-quality isolate genomes. Next, a major collaborative development area includes model-driven phenotype prediction and mechanistic analysis within KBase. This pipeline starts with a wide range of new tools designed to predict potential functions from protein sequence. New genome annotation pipelines, Distilling and Refining Annotation of Metabolism (DRAM) and Snekmer, expand annotation to new specialty areas of metabolism and offer alternative function hypotheses for difficult to annotate genes. Improvements to the KBase infrastructure were made to support multiple alternative theoretical annotations for genes developed with the Systems Biology Approach to Interactions and Resource Allocation in Bioenergy-Relevant Microbial Communities SFA. Growth phenotype data offers a means of discerning which of these alternative annotations is correct, but this data is commonly not available for many genomes. The KBase Knowledge Engine (KE) team addressed this by developing machine learning-based tools to predict phenotypes based on genome annotations. Following that, researchers determine which combinations of gene annotations lead to the best agreement with predicted phenotype. The Phenotypic Response of the Soil Microbiome to Environmental Perturbations SFA developed an algorithm for automatically fitting metabolic models to predicted (and observed) phenotype data. Further validation of proposed annotations can be done using protein structure-based evidence, which is now supported by a collection of KBase tools that import protein structure data for KBase proteins from the Research Collaboratory for Structural Bioinformatics Protein Data Bank. Finally, all of these model-driven workflows are enhanced by significant improvements to the ModelSEED metabolic model reconstruction and analysis tools in KBase. Together, these tools seamlessly interoperate, offering greatly enhanced understanding of genome metabolism with more accurate and quantitative energy metabolism and improving phenotype prediction accuracy from 56% on average for draft models to 72% accurate.

This pipeline is being applied to a growing collection of high-quality datasets loaded into KBase from collaborators. For example, the Genome Resolved Open Watershed (GROW) project contains 178 metagenomes, 50 metatranscriptomes, and 2,093 metagenome-assembled genomes that are available and linked to rich sample metadata in KBase. Another example is the Plant-Microbe Interfaces SFA, which is working on KBase apps to simplify the isolate selection process for constructed community experiments, adding >550 isolate genomes to KBase, and supporting integration of high-value datasets (Biolog data, BacDive database).

**Infrastructure Updates.** KBase has also undergone some infrastructure improvements over the past year. KBase has continued to dramatically improve bulk-upload support, which now enables upload of large datasets using a spread-sheet to specify object names, filenames, and metadata.

A central goal of KBase is to put users' data in context of all data in the platform, allowing users to quickly find and prioritize all relevant data. To further this goal, the project is introducing "Collections," high quality curated data sets and an interface for rapidly matching and sub selecting data sets based on their relationship to a user's data. These collections will initially focus on high quality, highly relevant datasets from the DOE community and will illuminate relationships between user data and these collections to enable new insights and drive further analysis.

KBase is also working to ensure community contributions are tracked and credited appropriately. KBase assigns Digital Object Identifiers (DOIs) for Narrative workflows that have been made static and documented for publication. This connects KBase research products to the broader publishing infrastructure. KBase will soon enable researchers to have their KBase DOIs visible as part of their individual ORCID record, and KBase will soon be able to report data reuse numbers to DataCite for all Narratives that have received a DOI for publication.

Finally, in collaboration with the DOE Joint Genome Institute through a co-development effort, KBase and Integrated Microbial Genomes (IMG) have generated a mapping of nonredundant protein sequences to UniRef 100 clusters which enables IMG users to identify and link directly to identical sequences in KBase. These platforms continue to work together to improve data connections, which will guarantee that data ownership and embargo periods are honored, even after data has been transferred between platforms.

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

### Learning and Training with KBase

Ellen G. Dow<sup>1</sup> (egdow@lbl.gov), Benjamin Allen<sup>3\*</sup> (allenbh@ornl.gov), Jason Baumohl<sup>1</sup>, Kathleen Beilsmith<sup>2</sup>, David Dakota Blair<sup>4</sup>, John-Marc Chandonia<sup>1</sup>, Dylan Chivian<sup>1</sup>, Zachary Crockett<sup>3</sup>, Meghan Drake<sup>3</sup>, Janaka N. Edirisinghe<sup>2</sup>, José P. Faria<sup>2</sup>, Jason Fillman<sup>1</sup>, Tianhao Gu<sup>2</sup>, A. J. Ireland<sup>1</sup>, Marcin P. Joachimiak<sup>1</sup>, Sean Jungbluth<sup>1</sup>, Roy Kamimura<sup>1</sup>, Keith Keller<sup>1</sup>, Dan Klos<sup>2</sup>, Miriam Land<sup>3</sup>, Filipe Lui<sup>2</sup>, Erik Pearson<sup>1</sup>, Gavin Price<sup>1</sup>, Priya Ranjan<sup>3</sup>, William Riehl<sup>1</sup>, Boris Sadkhin<sup>2</sup>, Samuel Seaver<sup>2</sup>, Alan Seleman<sup>2</sup>, Gwyneth Terry<sup>1</sup>, Sumin Wang<sup>1</sup>, Pamela Weisenhorn<sup>2</sup>, Ziming Yang<sup>4</sup>, Shinjae Yoo<sup>4</sup>, Qizhi Zhang<sup>2</sup>; Shane Canon<sup>1</sup>, Paramvir S. Dehal<sup>1</sup>, Elisha Wood-Charlson<sup>1\*</sup>, Robert Cottingham<sup>3</sup>, Christopher S. Henry<sup>2</sup>, **Adam P. Arkin**<sup>1</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory; <sup>2</sup>Argonne National Laboratory; <sup>3</sup>Oak Ridge National Laboratory; <sup>4</sup>Brookhaven National Laboratory

#### http://kbase.us

**Goals**: The Department of Energy Systems Biology Knowledgebase (KBase) is a knowledge creation and discovery

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

The KBase user interface (UI) enables instructors to work with students to conduct hands-on data science research and analysis without the need for programming skills or computational resources. The KBase team works with instructors and researchers of varying skill and career levels to ensure the transfer of domain knowledge is accompanied by an understanding of bioinformatic tools and techniques. KBase supports learning and training through the KBase Educators Community and by hosting workshops and webinars on community-focused topics. The variety of programming targets different cross-sections of the BER research community, with the overall goal to improve and expand the next generation of data analysis using KBase.

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

Educators from community colleges, primarily undergraduate institutions, and doctoral research institutions make up the KBase Educators Community. There is also representation of diverse student populations from Minority-Serving Institutions, including Hispanic-Serving, Asian American and Native American Pacific Islander-Serving, Alaskan Native-Serving, Native Hawaiian-Serving, Native American-Serving Nontribal, Historically Black Colleges or Universities, and Predominantly Black Institutions. Program growth and expansion will continue, as the community identifies additional areas that are important to support educators and their students.

Come see how KBase can support Promoting Inclusive and Equitable Research (PIER) Plans and develop connections with new collaborators for Reaching a New Energy Sciences Workforce (RENEW) and Funding for Accelerated, Inclusive Research (FAIR). **Outreach.** The KBase team hosts outreach and training events to support research groups, educators, and collaborators advance their research. Training events include workshops and webinars to demonstrate use of the platform and showcase popular workflows. Workshops are used to reach specific institutions to facilitate collaboration with researchers and students. Webinars reach a broader audience through web-based training to introduce new features, showcase workflows, and host speakers including KBase staff, community developers, and subject matter experts.

Webinars are posted on the KBase YouTube channel (https://www.youtube.com/DOEKBase) for anyone to view after the event and often include public Narratives for users to test new tools and workflows.

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

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

## The KBase Knowledge Engine: Ecosystem Classification Prototype

Paramvir S. Dehal<sup>1\*</sup> (psdehal@lbl.gov), Marcin P. Joachimiak<sup>1</sup>, Ziming Yang<sup>4</sup>, William Riehl<sup>1</sup>, Sean Jungbluth<sup>1</sup>, Meghan Drake<sup>3</sup>, Shane Canon<sup>1</sup>, Dylan Chivian<sup>1</sup>, Filipe Lui<sup>2</sup>, Christopher Neely<sup>1</sup>, Priya Ranjan<sup>3</sup>, Shinjae Yoo<sup>4</sup>, Benjamin Allen<sup>3</sup>, Jason Baumohl<sup>1</sup>, Kathleen Beilsmith<sup>2</sup>, David Dakota Blair<sup>4</sup>, John-Marc Chandonia<sup>1</sup>, Zachary Crockett<sup>3</sup>, Ellen G. Dow<sup>1</sup>, Janaka N. Edirisinghe<sup>2</sup>, José P. Faria<sup>2</sup>, Jason Fillman<sup>1</sup>, Tianhao Gu<sup>2</sup>, A. J. Ireland<sup>1</sup>, Roy Kamimura<sup>1</sup>, Keith Keller<sup>1</sup>, Dan Klos<sup>2</sup>, Miriam Land<sup>3</sup>, Erik Pearson<sup>1</sup>, Gavin Price<sup>1</sup>, Boris Sadkhin<sup>2</sup>, Samuel Seaver<sup>2</sup>, Alan Seleman<sup>2</sup>, Gwyneth Terry<sup>1</sup>, Sumin Wang<sup>1</sup>, Pamela Weisenhorn<sup>2</sup>, Qizhi Zhang<sup>2</sup>, Elisha Wood-Charlson<sup>1</sup>, Robert Cottingham<sup>3</sup>, Christopher S. Henry<sup>2</sup>, **Adam P. Arkin<sup>1</sup>** 

<sup>1</sup>Lawrence Berkeley National Laboratory; <sup>2</sup>Argonne National Laboratory; <sup>3</sup>Oak Ridge National Laboratory; <sup>4</sup>Brookhaven National Laboratory

#### https://kbase.us

**Goals:** One of the primary goals of the Department of Energy Systems Biology Knowledgebase (KBase) is the generation and application of biological knowledge from analytical results. To that end, the KBase Knowledge Engine (KE) will leverage existing and novel machine learning and bioinformatics tools to build up such knowledge from the growing body of results from analysis done using KBase and made publicly available. Ultimately, the project seeks to predict key taxa, functions, ecosystem features, and their interactions. To accomplish this, researchers begin by developing (1) classifiers for identification of key determinants of ecosystems; (2) phenotype and trait predictors; and (3) robust pangenomes and their relationships across the microbial tree of life.

Microbial life is a critical component of Earth's ecosystems, and the taxonomic and functional information from environmental genomics can provide insights into microbial roles in the environment. However, comparing this data across metagenomes can be challenging, and furthermore, abundance differences may not reflect important functional differences between environments. As a first KBase KE prototype, researchers aimed to use machine learning to: (1) build and evaluate robust ecosystem classification models using standardized data from ~32,000 metagenomes; and (2) identify important classification features and how they relate to understanding of environments on Earth.

Relying on standardized data from the European Bioinformatics Institute MGnify resource, researchers constructed feature tables associating metagenome sample environment labels with Gene Ontology (GO) term abundance, InterPro (IPR) domain, and predicted taxonomy profiles. Through a series of reusable data preparation and cleaning techniques, input data was generated for reliable model training. Hyperparameter tuning was performed on top multiclass classification methods and model performance was assessed with cross-validation. With a permutation analysis to extract feature importance from the top models, researchers obtained features important for classification and used these to construct trees and networks relating different ecosystems. Using relationships from the environmental classification as well as sample ecosystem outliers, the team interpreted model errors, including misclassifications as hypernyms and hyponyms, and were able to account for most model errors suggesting future improvements through better incorporation of classification semantics into model training. Researchers also identified a series of model predictions, which directly suggest sample relabeling, for example providing more specific terms for samples labeled as Environmental: Aquatic: Marine. Results provide a highperformance metagenome ecosystem classification model and enable model interpretability to learn important ecosystem indicator functions as well as ecosystem and function relationships.

**Funding Information:** This work is supported as part of the Biological and Environmental Research (BER) Program's Genomic Science program. The DOE Systems Biology

Knowledgebase (KBase) is funded by the U.S. Department of Energy, Office of Science, BER Program under Award Nos. DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.

## NLP for Synthetic Biology: Providing Generalizable Literature Mining Through KBase

Shinjae Yoo<sup>1\*</sup> (sjyoo@bnl.gov), Carlos X. Soto<sup>1</sup>, Gilchan Park<sup>1</sup>, Christopher Neely<sup>2</sup>, Vivek K. Mutalik<sup>2</sup>, **Paramvir S. Dehal**<sup>2</sup>

<sup>1</sup>Brookhaven National Laboratory (BNL); <sup>2</sup>Lawrence Berkeley National Laboratory

Goals: The scientific literature contains many decades of research results, which may inform the identification and subsequent engineering of microbial targets for novel applications, yet this knowledge remains largely inaccessible to current researchers due to the scale of the literature and the limitations of current manual information extraction practices using literature search and laborious manual curation processes used in String DB and AraNet. This project will produce a reusable proof-of-concept demonstration applying state-of-the-art natural language processing (NLP) techniques within the DOE Systems Biology Knowledgebase (KBase) framework to automatically extract organism traits from the literature for synthetic biology research. This work seeks to address important knowledge gaps in this field while simultaneously providing a meaningful staging ground to expose new NLP tools to the KBase community and to gather feedback on their efficacy and use. The team will leverage NLP and data collection methods that have been previously developed and successfully applied in isolated settings to accomplish this effort, work with existing KBase tools and functionality, and produce outreach material to communicate and disseminate this work and its user-facing outcomes.

Earth is facing some serious biological resource problems: a scarcity of renewable energy, lack of novel remedies for endemic infectious diseases, water pollution, shortages of arable soil and the resultant food crises, and the degradation of ecosystems to name a few of the most pressing. The project posits that the ability to domesticate and genetically engineer nonmodel microorganisms from relevant niches would help assess new potential solutions to many of these life-threatening global challenges. Though there have been technological innovations happening at rapid pace in addressing many of these challenges, the information for each potential new model organism is distributed throughout the literature and inaccessible to many practitioners, making unnecessarily difficult every new synthetic biology, bioenergy, and bioproduct project in a nonmodel organism. This lack of organized information not only limits machine-readable approaches, it also makes it difficult to assess the scope of work, identify knowledge gaps, and offer suggestions for investment to overcome technological barriers. For example, after decades of development in the field of synthetic biology, it is still challenging to identify suitable microbial targets for specific applications, conditions, and genetic tools necessary for the cultivation and engineering of nonmodel microorganisms. A generalized literature mining tool that keeps track of new technologies and genetic tools important for biotechnology practitioners would be invaluable. This tool will enable discovery of information gaps and opportunities that are buried within the vast literature. For example, the tool should be able to find information such as if the organism of choice is appropriate for domestication in the lab and which genetic tools exist for the organism with more ease than searching hundreds of primary literature sources, thereby saving time, effort, and money for the DOE-funded project. By building this prototype literature mining service into KBase, the team will be able to surface issues with KBase platform integration (including how the integration should be done), to identify the scoping and scaling needs, and to explore how to best address the needs of users.

Recently, improved techniques from the field of NLP have made it possible to analyze text at unprecedented scales (e.g., millions of documents), while extracting meaningful contextual information in ways not previously possible. These techniques are highly suitable to help address the above-mentioned knowledge gap. Here, the team will apply NLP to the biological literature to extract organism traits (see Fig. 18, p. 166) and deposit this mined knowledge in a useful form, supporting the creation of automated, curated centralized systems essential for growing and engineering of DOE-relevant microorganisms. While the BNL NLP techniques have extremely compelling applications, their value to BER researchers has been limited by access and dissemination of their results. KBase integration will bring the knowledge captured by the literature to a much wider audience.

**Funding Information**: This material is based upon work supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) Program, under Award No. DE-AC02-05CH11231 (LBL) and DE-SC-0012704 (BNL).



**Fig. 18. Overview of the Project.** The team proposes to use best natural language processing (NLP) and machine learning models to process and learn from literature data about growth characteristics, conditions, traits (e.g., antibiotic and stress tolerance, fitness traits), and available *in silico* models and genetic tools to engineer DOE-BER centric microorganisms. [Courtesy Vivek Mutalik]

## Probabilistic Annotation and Ensemble Metabolic Modeling in KBase

**Patrik D'haeseleer<sup>1\*</sup>** (dhaeseleer2@llnl.gov), Jeffrey Kimbrel<sup>1</sup>, Ali Navid<sup>1</sup>, Christopher S. Henry<sup>2</sup>, Rhona Stuart<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory (LLNL); <sup>2</sup>Argonne National Laboratory

https://www.kbase.us/research/stuart-sfa/

**Goals:** Functional annotation tools such as Rapid Annotation using Subsystem Technology (RAST) or Kyoto Encyclopedia of Genes and Genomes (KEGG) don't always agree on how to best leverage them for metabolic modeling. This project is developing tools for the DOE Systems Biology Knowledgebase (KBase) to give users a principled way to weigh multiple sources of functional annotation against each other, enable better metabolic modeling of hard-to-annotate organisms and pathways, allow analysis of uncertainty in the resulting models network structure or behavior, and provide an infrastructure on which to build more sophisticated machine learning techniques in KBase.

The  $\mu$ Biospheres science focus area (SFA) at LLNL investigates metabolic interactions in bioenergy-relevant microbial communities. A critical part of this research is development of genome-scale models of metabolism, which requires well-annotated genomes. By combining annotations from multiple sources, researchers can achieve a more complete metabolic network reconstruction, greatly reducing the effort required to curate quality metabolic models (Griesemer et al. 2018). In previous work, researchers developed a set of KBase apps to import, compare, and merge functional annotations from a wide range of different functional annotation tools into KBase for metabolic modeling to achieve significantly improved metabolic models. These apps have proven to be very useful and are currently in daily use in this SFA and several other research groups using KBase.

It is quite common for functional annotation tools to disagree on the function that should be assigned to certain genes, and this uncertainty can have significant consequences on the resulting metabolic networks and the behavior they predict for the organism. The team is now developing a set of tools to deal with these disagreements in a more systematic manner: by calculating the likelihood of metabolic reactions given the annotations from various sources, and then carrying those reaction likelihoods through into the modeling results. Researchers have modified the existing import app to support importing annotation scores and evidence codes, such as reaction probabilities, log likelihoods, Basic Local Alignment Search Tool, or hidden Markov model scores.

Researchers can use a Naive Bayes approach to estimate the probability of each reaction assigned to a gene, given the annotations from a range of different annotation tools. For this the team first needs to evaluate the reliability—False Positive and False Negative rates—for each of the major annotation tools (currently, RAST, Prokka, Distilled and Refined Annotation of Metabolism, and KEGG), by running them on a reference dataset consisting of 15,000 enzymes in Swiss-Prot that have experimental evidence codes. This rigorous validation effort has also led to some significant improvements in the ModelSEED biochemistry database, beyond the well-curated set of template reactions that are normally used by KBase's metabolic modeling engine.

Enzymes in Swiss-Prot have historically been annotated using Enzyme Commission (EC) numbers, which are far from ideal when needing to map to unique metabolic reactions for modeling. Some ECs are overly generic, forcing omission altogether, or to instantiate them as multiple unique reactions. The team is building support into the Ontology application programming interface (that translates from EC numbers and other annotation vocabularies to ModelSEED reactions) to filter out unbalanced, overly generic, or otherwise unsuitable reactions for metabolic modeling. In the longer term, the team may use the new Rhea reaction identifiers that are being curated into the Swiss-Prot database as the reference dataset, which should provide for a much more direct mapping to ModelSEED reactions.

Once reaction probabilities are associated with all the genes in a genome, researchers can then sample from those probabilities to create an ensemble of metabolic models (Medlock, Moutinho, and Papin 2020). Each of these models can then be analyzed using the existing gapfilling and modeling tools (including support for the next generation of modeling tools that the KBase team is developing), eventually resulting in an ensemble of Flux Balance Analysis solutions, reflecting the uncertainty in the underlying enzyme annotations. The team will develop a set of analysis tools to study this ensemble of solutions, using clustering, averaging, analysis of alternative pathway solutions, etc. This will result in higher quality metabolic network reconstruction, but also in much greater insight in the sources of uncertainty in the network, enabling prioritization of how to most efficiently reduce that uncertainty by additional manual curation or experimental data.

This work will provide the SFA and other KBase users a principled way to weight annotation sources against each other, enable better metabolic modeling of hard-to-annotate organisms and pathways, allow analysis of uncertainty in the resulting models network structure or behavior, and provide an infrastructure on which to build more sophisticated machine learning techniques in KBase.

Griesemer, M., et al. 2018. "Combining Multiple Functional Annotation Tools Increases Coverage of Metabolic Annotation," *BMC Genomics* **19**(1), 948.

Medlock, G. L., T. J. Moutinho, and J. A. Papin. 2020. "Medusa: Software to Build and Analyze Ensembles of Genome-Scale Metabolic Network Reconstructions," *PLoS Computational Biology* **16**(4), e1007847.

**Funding Information:** 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 Genomic Science program of the Biological and Environmental Research (BER) Program under the LLNL  $\mu Biospheres$  SFA, FWP SCW1039.

## Crosslinking The Department of Energy Systems Biology Knowledgebase (DOE-KBase) and the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) to Support Protein Function Discovery

Qizhi Zhang<sup>1</sup>, Claudia Lerma-Ortiz<sup>1</sup>, Dennis Piehl<sup>2</sup>, Brinda Vallat<sup>2</sup>, Shuchismita Dutta<sup>2</sup>, Janaka N. Edirisinghe<sup>1</sup>, Samuel Seaver<sup>1</sup>, **Stephen K. Burley<sup>2</sup>**, **Christopher S. Henry<sup>1\*</sup>** (chenry@anl.gov)

<sup>1</sup>Argonne National Laboratory; <sup>2</sup>RCSB Protein Data Bank, Rutgers, The State University of New Jersey *https://kbase.us/n/127880/210/ https://kbase.us/n/130357/26/* 

https://www.rcsb.org/

Goals: Systems Biology Knowledgebase (DOE-KBase) and the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) offer synergistic functionality to investigate and engineer proteins. The collection of systems biology data and tools in KBase enables scientists to analyze their datasets in the context of public data and share their findings. RCSB PDB provides access to more than 200K experimentally determined, rigorously validated, and expertly biocurated 3D structures of proteins and nucleic acids within the open PDB archive. The RCSB. org web portal provides a variety of tools for searching, analyzing, and visualizing 3D biostructure data together with annotations from ~50 public resources. RCSB.org now supports parallel delivery of more than 1M computed structure models from AlphaFold DB and RoseTTAFold. The objective of this project was to lay the foundation for interoperation of these two resources, streamlining the ability of users to leverage structural biology data and workflows provided by RCSB PDB within the KBase platform.

Many projects currently funded by DOE BER aim to mechanistically understand a wide range of complex biological systems with the ultimate goal of supporting the rational manipulation, prediction, and design of these systems. The large fraction of proteins with unknown or incompletely characterized function is one of the greatest impediments to this goal. Structural biology is central to resolving and understanding protein function, particularly with the advent of the AlphaFold2 and RoseTTAFold algorithms for rapidly predicting new computed structure models (CSMs) of proteins with accuracies comparable to that of low-resolution experimental structures. Yet, structural biology approaches are greatly amplified when combined with systems biology data and tools. Toward this end, the KBase and RCSB PDB teams collaborated to develop a series of applications within the KBase platform that leverage the powerful capabilities of RCSB.org data delivery and API services for integrating PDB data into systems biology and structural biology workflows across KBase and RCSB PDB.

Researchers demonstrate these newly developed workflows with two exemplar scenarios. In the first scenario, researchers identify genes encoded by the bacterium Micrococcus *luteus* that are responsible for producing proteins capable of degrading pyridine, a toxic compound found in coal tar. This workflow demonstrates combined use of transcriptomics, mechanistic modeling, and chemoinformatics to propose candidate genes for a novel biochemical pathway for pyridine degradation. Researchers then apply the new KBase-RCSB PDB pipeline to: (1) rapidly search the PDB for experimental structures that are homologous to the candidate gene products; (2) seek experimental structures of proteins co-crystallized with pyridine to identify pyridine binding sites; and (3) import and view AlphaFold2generated structures for the candidate gene products, comparing each predicted structure with the closest experimental structures represented in the PDB archive. Researchers then conducted structure motif searches at RCSB.org to further characterize the pyridine binding site and perform structure comparisons of the AlphaFold predictions with the collection of experimental PDB structures and CSMs available at RCSB.org. Ultimately, binding site analyses aided in the performance of docking simulations in KBase to refine the gene candidates for the novel pathway.

In the second scenario, researchers exemplify concerted use of KBase and RCSB PDB to discover the unknown pyrimidine reductase (EC 1.1.1.193), a key enzyme in the Riboflavin (vitamin B) biosynthesis in Arabidopsis thaliana. Using annotation, modeling, and gapfilling tools in KBase, the team confirmed that the gene encoding pyrimidine reductase was not identified in the Arabidopsis annotation selected for this scenario. Researchers applied two newly developed KBase-RSCB PDB interface apps. The PDB-Import PDB Metadata into KBase Genome app enabled the team to query PDB for experimental structures that match sequences of any of the gene products in the entire Arabidopsis genome. Doing so exposed two Arabidopsis gene products with significant similarity to multiple experimental structures of microbial proteins in PDB currently annotated as pyrimidine reductases. The team used the Query RCSB Databases for Protein Structures app to import and view these structures, as well as offer links to views of these structures in RCSB. org. Then used a capability within RCSB.org to compare the experimental structures of interest with Arabidopsis Alpha-Fold2 CSMs now available on RCSB.org. The team used the

RCSB.org pairwise structure alignment tool to determine that AlphaFold2 CSMs of both *Arabidopsis* candidate gene products aligned to distinct portions of a microbial pyrimidine reductase structures housed in the PDB. Although both *Arabidopsis* proteins are structurally similar, a detailed 3D analysis revealed that AT3G47390 lacks essential zinc-binding residues within its putative deaminase domain. Taken together with the observation that AT4G20960 had already been identified as the deaminase led to the hypothesis that AT3G47390 is the pyrimidine reductase (which was confirmed in publications).

KBase Narratives and RCSB tools applied in each of these scenarios are described in detail in a publicly available training workshop on YouTube: https://www.youtube. com/watchv=vs\_UyhhtSFk&list=PLHib7JgKNUUf8Z-8jSK57FsJrms94w paL0&index=1Z

**Funding Information:** Argonne National Laboratory is managed by UChicago Argonne, LLC for the U.S. Department of Energy under contract No. DE-AC02-06CH11357. This program is supported by the U.S. Department of Energy, Office of Science, through the Genomic Science program, Biological and Environmental Research Program, under FWP PRJ34888. RCSB PDB Core Operations are funded by National Science Foundation (DBI-1832184), U.S. Department of Energy (DE-SC0019749), and National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM133198.

## Integration of Enzyme Function Initiative Tools in the KBase Platform

Nils Oberg<sup>2</sup>, Douglas Mitchell<sup>2</sup>, John Gerlt<sup>2</sup>, Christopher S. Henry<sup>1\*</sup> (chenry@anl.gov)

<sup>1</sup>Argonne National Laboratory; <sup>2</sup>University of Illinois Urbana–Champaign

Goals: Protein families of unknown function are a significant challenge facing the DOE BER research community because they prevent comprehensive metabolic reconstructions of both individual microorganisms and microbiome systems. While many tools in KBase and elsewhere today permit the discovery of completely new protein families, very few tools exist, particularly in KBase, to study the function of these families. Fortunately, the Enzyme Function Initiative (EFI; http://enzymefunction.org) offers a suite of tools specifically designed to address this important problem. In this project researchers are working to fully integrate the EFI toolset into KBase, with complete ties to DOE BER sequencing sources including all sequence data in KBase as well as the JGI IMG Database. Advances in computational methods and DNA sequencing now allow for single projects to generate tens to hundreds of metagenome sequences and potentially tens of thousands of isolate or metagenomically assembled genomes (MAGs) from diverse ecosystems. In theory, computational inference of the protein products encoded by these genomes and the associated biochemical functions should allow for the accurate prediction and modeling of key microbial traits, organismal interactions, and ecosystem processes that drive biogeochemical cycles. Unfortunately, the rate and generation of metagenomes, isolate genomes, and MAGs, along with related multiomic datasets, currently far outpaces the ability to translate these genome-enabled findings into ecosystem-informed predictive knowledge.

One of the most significant challenges currently inhibiting the understanding of complex biological systems from genomic and multiomic data is the staggering number of proteins that have completely unknown functions. About 50% of the proteins encoded by the genes in complete microbial genomes, and an even higher proportion of those encoded by microbial genes from environmental samples, cannot be reliably assigned a function. These unknown functions translate into large gaps in the metabolic reconstructions, prevent researchers from explaining more than 25% of most metabolomes, and obfuscate the functional interdependencies that guide the structure of all microbiome systems. Despite these challenges, virtually all of the functional annotation tools currently available in KBase and other platforms focus largely on assigning functions to proteins that are very similar to other proteins of known function (e.g., via propagation of function based on close sequence homology). Because the sequence boundaries between functions cannot be specified in the absence of orthogonal information, homology-based annotations often are incorrect. Tools are needed that are designed to integrate multiple sources of evidence to decode the functions of uncharacterized protein families and understand the limits of annotation propagation.

The Enzyme Function Initiative (EFI) toolkit is designed to fill this exact niche in protein function discovery. The EFI tool pipeline is comprised of three analysis steps: (1) generation of sequence similarity networks (SSNs) enabling the semi-automated reconstruction of high-quality protein families built around any protein sequence of interest (EFI-EST; https://efi.igb.illinois.edu/efi-est/); (2) parallel exploration of the genome neighborhood context of a protein family across a diverse set of input genomes to discover potential functionally linked gene products/enzymes that can be used to infer novel enzymatic functions and metabolic pathways (EFI-GNT; https://efi.igb.illinois.edu/efi-gnt/); and (3) determination of metagenome abundance of clusters in the SSNs for protein families using chemically guided functional profiling (CGFP) to discover the physiological/ environmental context in which the proteins are expressed (EFI-CGFP; https://efi.igb.illinois.edu/efi-cgfp/). The EFI tools also provide links to structure data in Protein Data Bank (PDB) to gain further clues about protein function.

Researchers are deploying the EFI tools into the KBase platform, with re-engineering on the backend to permit a seamless integration with the KBase database of isolate, reference genome, metagenome, and MAG sequences. This will greatly enhance the value of the EFI tools to the DOE BER research community; greatly expand the ability of these tools to access more diverse sequence data and annotation sources (e.g., IMG); significantly ease the long term maintenance of the EFI tools by linking to the KBase data update cycle; and greatly enhance the capacity for users of the KBase platform to study protein families of unknown function.

**Funding Information**: Argonne National Laboratory is managed by UChicago Argonne, LLC for the U.S. Department of Energy under contract No. DE-AC02-06CH11357. This program is supported by the U.S. Department of Energy, Office of Science, through the Genomic Science program of the Biological and Environmental Research Program, under FWP PRJ39217.

### Design and Omics Exploration of Synthetic Microbial Communities in KBase

**Priya Ranjan**<sup>1\*</sup> (ranjanp@ornl.gov), Andrew Freiburger<sup>2</sup>, Alexis L. Marsh<sup>3</sup>, Dale A. Pelletier<sup>1</sup>, Myra B. Cohen<sup>3</sup>, David J. Weston<sup>1</sup>, Christopher S. Henry<sup>2</sup>, Paramvir S. Dehal<sup>4</sup>, Adam P. Arkin<sup>4</sup>, Robert Cottingham<sup>1</sup>, Mitchel J. Doktycz<sup>1</sup>

<sup>1</sup>Oak Ridge National Laboratory; <sup>2</sup>Argonne National Laboratory; <sup>3</sup>Iowa State University; <sup>4</sup>Lawrence Berkeley National Laboratory

https://www.kbase.us/research/doktycz-sfa/

Goals: Simple constructed communities with desired biological functions can be used to study bacterial processes involved in community establishment and mimic the behavior of natural communities to increase plant growth and disease resistance. In this collaboration with the Plant-Microbe Interfaces (PMI) science focus area (SFA) at Oak Ridge National Laboratory, team members are adding datasets and apps to KBase to simplify the selection of isolates for constructed community experiments. For designing the community, researchers can use KBase apps that annotate genomes with plant growth-promoting traits and secondary metabolism classes as well as an app that calculates metabolic dependencies between microbes of interest. The design process will be tested using experimental systems established in the PMI SFA for studying constructed communities. Further, the results from these experiments will

be integrated back into KBase to improve the mechanistic understanding of interactions and iteratively improve the design process.

Simple constructed communities offer a key experimental platform for examining how environmental perturbations affect the structure of the microbiome and host physiology and productivity. Computational simulations are a necessary tool to guide the design of simplified constructed communities. KBase-a DOE BER-funded, public, and freely accessible software and data science platform with a rich user interface—is ideal for developing such computational tools since it already offers a large and increasing number of diverse tools (e.g., functional annotation, metabolic modeling, auxotrophy prediction, substrate utilization and production of byproducts, taxonomic information, and predicting microbial traits). With these tools, it is possible to get some insights about a genome's biochemistry and general characteristics only based on its sequence. The following user interface applications (apps) are under development to expand the set of KBase tools in support of constructed community studies.

The first app, *Annotate Genomes with Plant Growth Promoting Traits,* applies annotation to genes in a genome based on the Plant Growth Promoting Trait ontology (PGPT). PGPT is a literature- and omics-curated, comprehensive, and hierarchical collection containing 6,900 PGPTs associated with 6,965,955 protein sequences. The ontology has several categories such as phytohormone production, plant signal production, biofertilization (potassium solubilization, iron acquisition), bioremediation (fluoride, heavy metal detoxification), colonizing plant system (chemotaxis, surface attachment, root colonization), plant immune response stimulation, stress control, competitive exclusion (quorum sensing, bacterial fitness, cell envelope remodeling). This app can help a user prioritize and select genomes for constructed community experiments.

The app Annotate Genomes with Secondary Metabolism Classes uses antiSMASH (1), a popular genomics tool, to annotate genes in a genome with secondary metabolism classes and identify biosynthetic gene clusters within the microbial genomes of interest. The biosynthetic gene cluster profile can be used to generate hypotheses regarding certain organisms having a higher potential for antagonistic and antimicrobial activity. This information can then be used to form subsequent hypotheses regarding the membership of stable communities. The team will also be exploring outputs from the JGI Secondary Metabolite Collaboratory as they work with them in the integration of the analysis workflow into KBase.

The app *Calculate Metabolic Interaction Score* builds pairwise metabolic models in a given environment and calculates several quantitative metrics that describe competitive and cooperative potential between the paired microbes. These metrics include: (1) the metabolic interaction potential (MIP), which approximates the potential for syntrophic cooperation between the organisms; (2) the metabolic resource overlap (MRO), which approximates competition for media substrates between the organisms; and (3) predicted growth rates of the isolates and community, which leverages other KBase Apps to approximate growth dynamics of the given community.

The aforementioned apps will be used by PMI researchers to prioritize genomes for constructed community experiments and generate testable hypotheses (e.g., the apps can identify microbial functions and possibly redundant community members, which can be tested by replacing one member with another microorganism whose functional and metabolic profile are predicted to be similar or different). Team members have also uploaded 550 PMI isolate genomes to KBase and are exploring design of improved user interface for searching, filtering, and selecting genomes for constructed communities. These datasets and apps will expedite the design-build-test-learn cycles of PMI SFA projects and enable new scientific explorations.

Blin, K., et al. 2021. "antiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities," *Nucleic Acids Research* **49**(1), W29– W35. DOI:10.1093/nar/gkab335.

**Funding Information:** Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract No. DE-AC05-00OR22725. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program of the Biological and Environmental Research Program, under FWP ERKPA39.
## NMDC: DOE NATIONAL MICROBIOME DATA COLLABORATIVE

#### Community Engagement Strategies of the National Microbiome Data Collaborative (NMDC)

Julia Kelliher<sup>1\*</sup> (jkelliher@lanl.gov), Alicia Clum<sup>3</sup>, Anastasiya Prymolenna<sup>2</sup>, Antonio Camargo<sup>3</sup>, Bin Hu<sup>1</sup>, Cameron Giberson<sup>2</sup>, Chien-Chi Lo<sup>1</sup>, Chris Mungall<sup>3</sup>, Donny Winston<sup>5</sup>, Elais Player Jackson<sup>1</sup>, Francie Rodriguez<sup>1</sup>, Grant Fujimoto<sup>2</sup>, Ingrid Ockert<sup>3</sup>, Jeff Baumes<sup>4</sup>, Kjiersten Fagnan<sup>3</sup>, Lee Ann McCue<sup>2</sup>, Marcel Huntemann<sup>3</sup>, Mark Flynn<sup>1</sup>, Mark Miller<sup>3</sup>, Michal Babinski<sup>1</sup>, Migun Shakya<sup>1</sup>, Mike Nagler<sup>4</sup>, Montana Smith<sup>2</sup>, Patrick Chain<sup>1</sup>, Patrick Kalita<sup>3</sup>, Paul Piehowski<sup>2</sup>, Po-E (Paul) Li<sup>1</sup>, Samuel O. Purvine<sup>2</sup>, Set Sarrafan<sup>3</sup>, Shane Canon<sup>3</sup>, Shreyas Cholia<sup>3</sup>, Simon Roux<sup>3</sup>, Sujay Patil<sup>3</sup>, Yan Xu<sup>1</sup>, Yuri E. Corilo<sup>2</sup>, **Emiley A. Eloe-Fadrosh<sup>3</sup>** 

<sup>1</sup>Los Alamos National Laboratory; <sup>2</sup>Pacific Northwest National Laboratory; <sup>3</sup>Lawrence Berkeley National Laboratory; <sup>4</sup>Kitware, Inc.; <sup>5</sup>Polyneme LLC

https://microbiomedata.org/

https://microbiomedata.org/ambassadors/ https://microbiomedata.org/community/championsprogram/

**Goals:** The vision of the National Microbiome Data Collaborative (NMDC) is to connect data, people, and ideas to advance microbiome innovation and discovery. With this vision in mind, the NMDC seeks to support a findable, accessible, interoperable, and reusable (FAIR) microbiome data sharing network through infrastructure, data standards, and community building that addresses pressing challenges in environmental sciences. The NMDC engagement strategy focuses on promoting a collaborative ecosystem for diverse microbiome researchers and implementing community feedback in all of the NMDC efforts and products.

The NMDC is a multinational laboratory initiative focused on advancing innovation and discovery in the field of microbiome science through the project's development of products and tools for the environmental microbiome research community (Wood-Charlson et al. 2020). The NMDC provides the community with three products: (1) The Submission Portal, (2) The Data Portal (Eloe-Fadrosh et al. 2022), and (3) NMDC EDGE, each aimed at making multiomics microbiome data FAIR. Each of these products was designed to specifically address the larger research community's specific needs and wants. The NMDC team implements recommendations and insights gleaned from usability testing and community feedback to continuously improve products. The team routinely engages with microbiome researchers to discuss how they want the NMDC products to look and operate, as well as understand what new functionality would benefit future research. In addition, the NMDC communicates and engages with many types of stakeholders, including funding agencies, publishers, institutions, programs, projects, and individual scientists. As part of these collaborative efforts, the NMDC hosts and co-hosts workshops (Vangay et al. 2021), webinars, presentations, panel discussions, and other events aimed at spreading awareness of and lowering barriers to adoption of FAIR principles in microbiome research and data generation. The NMDC Ambassador program allows early career researchers to host some of these events, thus expanding the overall reach of the content and training materials, while providing the Ambassadors with valuable experiences and career opportunities. The NMDC Champions program brings together microbiome researchers from diverse backgrounds to contribute to the NMDC (e.g., by beta-testing the NMDC products, co-authoring publications with the NMDC team, providing feedback). The NMDC will continue to prioritize community engagement as the products and network grow.

Eloe-Fadrosh, E. A., et al. 2022. "The National Microbiome Data Collaborative Data Portal: An Integrated Multi-Omics Microbiome Data Resource," *Nucleic Acids Research* **50**(D1), D828–36. DOI:10.1093/nar/gkab990.

Vangay, P., et al. 2021. "Microbiome Metadata Standards: Report of the National Microbiome Data Collaborative's Workshop and Follow-on Activities," *mSystems* **6**, e01194-20. DOI:10.1128/ mSystems.01194-20.

Wood-Charlson, E. M., et al. 2020. "The National Microbiome Data Collaborative: Enabling Microbiome Science," *Nature Reviews Microbiology* **18**, 313–14. DOI:10.1038/s41579-020-0377-0.

**Funding Information**: This work is supported by the U.S. Department of Energy Office of Science, through the Genomic Science program in the Biological and Environmental Research (BER) Program. The national laboratory partners are operated under contract Nos. DE-AC02-05CH11231 (LBNL), 89233218CNA000001 (LANL), and DE-AC05-76RL01830 (PNNL).

# Portals of Discovery: The NMDC Infrastructure and Products for Microbiome Research

Montana Smith<sup>1\*</sup> (montana.smith@pnnl.gov), Alicia Clum<sup>3</sup>, Anastasiya Prymolenna<sup>1</sup>, Antonio Camargo<sup>3</sup>, Bin Hu<sup>2</sup>, Cameron Giberson<sup>1</sup>, Chien-Chi Lo<sup>2</sup>, Chris Mungall<sup>3</sup>, Donny Winston<sup>5</sup>, Elais Player Jackson<sup>2</sup>, Francie Rodriguez<sup>2</sup>, Grant Fujimoto<sup>1</sup>, Ingrid Ockert<sup>3</sup>, Jeff Baumes<sup>4</sup>, Julia Kelliher<sup>2</sup>, Kjiersten Fagnan<sup>3</sup>, Lee Ann McCue<sup>1</sup>, Marcel Huntemann<sup>3</sup>, Mark Flynn<sup>2</sup>, Mark Miller<sup>3</sup>, Michal Babinski<sup>2</sup>, Migun Shakya<sup>2</sup>, Mike Nagler<sup>4</sup>, Patrick Chain<sup>2</sup>, Patrick Kalita<sup>3</sup>, Paul Piehowski<sup>1</sup>, Po-E (Paul) Li<sup>2</sup>, Samuel O. Purvine<sup>1</sup>, Set Sarrafan<sup>3</sup>, Shane Canon<sup>3</sup>, Shreyas Cholia<sup>3</sup>, Simon Roux<sup>3</sup>, Sujay Patil<sup>3</sup>, Yan Xu<sup>2</sup>, Yuri E. Corilo<sup>1</sup>, **Emiley A. Eloe-Fadrosh<sup>3</sup>** 

<sup>1</sup>Pacific Northwest National Laboratory; <sup>2</sup>Los Alamos National Laboratory; <sup>3</sup>Lawrence Berkeley National Laboratory; <sup>4</sup>Kitware Inc, New York; <sup>5</sup>Polyneme LLC, New York

#### https://microbiomedata.org

**Goals**: The vision of the National Microbiome Data Collaborative (NMDC) is to connect data, people, and ideas to advance microbiome innovation and discovery. The team is committed to creating the needed infrastructure to answer tomorrow's research questions. With this vision in mind, the NMDC seeks to support a Findable, Accessible, Interoperable, and Reusable (FAIR) microbiome data sharing network—through infrastructure, data standards, and community building—that addresses pressing challenges in environmental sciences. The infrastructure and portals that the NMDC has developed provide the research community with a platform to share their microbiome research and data in accordance with the FAIR principles, thereby promoting data reuse and accelerating scientific discoveries.

The NMDC is committed to FAIR multiomics microbiome data. The NMDC infrastructure supports a collaborative,

integrative science ecosystem that empowers the research community to contribute, explore and investigate microbiome data. This is accomplished through the three key NMDC products that are openly available to the research community: (1) the Submission Portal, (2) the Data Portal, and (3) NMDC EDGE. The NMDC Submission Portal provides users a place to contribute sample metadata in a standardized manner with in-sheet validation to ensure machine readability and findability. The Data Portal consumes this metadata and presents search tools to find and access information about the research studies and data generated from the samples. The Data Portal also provides links to the associated omics data processed through the NMDC's standardized bioinformatics workflows. NMDC EDGE is an user-friendly interface for the NMDC standardized bioinformatics workflows. These three NMDC products are built specifically with the BER research community in mind and refined through a process of continual collaboration that is based on user-focused feedback. Leveraging each of the NMDC products enables environmental microbiome researchers to adhere to FAIR data principles, thus expanding potential research questions, comparisons, and scientific discovery.

Eloe-Fadrosh, E. A. et al. 2022. "The National Microbiome Data Collaborative Data Portal: An Integrated Multi-omics Microbiome Data Resource," *Nucleic Acids Research*. 7(60), (D1):D828–D836. DOI:10.1093/nar/gkab990.

Wood-Charlson, E. M. et al. 2020. "The National Microbiome Data Collaborative: Enabling Microbiome Science," *Nature Reviews Microbiology* **18**, 313–314. DOI:10.1038/s41579-020-0377-0.

**Funding Information:** This work is supported by the U.S. Department of Energy Office of Science, through the Genomic Science Program in the Biological and Environmental Research (BER) Program. The national laboratory partners are operated under contract Nos. DE-AC02-05CH11231 (LBNL), 89233218CNA000001 (LANL), and DE-AC05-76RL01830 (PNNL).

# Index

## A

Abadoul, Aghiad 131 Abraham, Joshua P. 49 Abraham, Paul E. 110, 111, 148 Acharya, Shwetha M. 142 Acheson, Justin F. 14, 16 Adams, Catharine A. 142 Adams, Michael W. 114, 139, 140 Adams, Paul D. 138, 139, 140 Addison, Bennett 13 Adler, Benjamin A. 143 Ahamed, Firnaaz 145 Ahkami, Amir H. 54, 55 Aladesanmi, Omolara T. 142 Alexander, Bill 10, 71 Alexander, William 108 Alharthi, Dalal 132 Allemann, Marco N. 89 Allen, Benjamin 162, 163, 164 Allen, Doug K. 49, 74, 75 Allen, Jonathan 105 Allison, Steven D. 117, 118 Allman, Andrew 85 Altenburg, Sara 139 Altpeter, Fredy 6 Amador-Noguez, Daniel 40, 42 Amezola, Melissa 103 AminiTabrizi, Roya 130 Andeer, Peter F. 99, 142, 153 Anderson, Jeremiah 66 Anderton, Christopher 22 Andrews, Holly 125 Ané, Jean-Michel 52, 53 Antoniewicz, Maciek 85 Appidi, Manasa R. 148 Arana, Sebastian E. Toro 34 Aristilde, Ludmilla 40 Arkin, Adam P. 138, 139, 140, 162, 163, 164, 169 Arnold, Jonathan 22 Arnolds, Kathleen L. 103, 104 Assato, Carol Kiam 36 Aufrecht, Jayde 54, 55 Avalos, José L. 7, 41 Avala-Ortiz, Christian 132 Azubuike, Christopher C. 89

#### B

Baas-Thomas, Maximilien 78 Babinski, Michal 171, 172 Babusci, Erica 131 Baidoo, Edward 19, 29 Baig, Nameera F. 142 Bailey-Bale, Jack 8 Baker, Scott 92 Balasubramanian, Vimal 54 Baldino, Henri 107, 108 Bale, Jacob 51 Baliga, Nitin S. 139, 140 Balmant, Kelly 30, 31 Banaei-Kashani, Farnoush 155 Banan, Darshi 54, 55 Banerjee, Shivali 6 Banfield, Jillian 142, 143, 150 Banta, Amy B. 17 Baral, Nawa 20 Barnes, Elle M. 58, 153 Barrangou, Rodolphe 143 Barrett, Jennifer 69 Bartley, Laura E. 57 Bates, Philip D. 74, 75 Battacharyya, Amrita 152 Baugh, Alyssa 90 Baumes, Jeff 171, 172 Baumohl, Jason 162, 163, 164 Bawal, Sarthak 25 Baxter, Ivan 69, 70 Baym, Michael 78 Bechtold, Emily 134 Beckham, Gregg T. 40, 43 Beckley, John 143 Beecroft, Chris 160 Behsaz, Bahar 156 Beilsmith, Kathleen 162, 163, 164 Bell, Elizabeth 43 Bell, Sheryl L. 119, 153 Bennetzen, Jeff 22 Benucci, Gian 17 Berg, Maureen 160 Bernacchi, Carl J. 2 Bezrutczyk, Margot 22 Bharadwaj, Vivek S. 13, 38 Bhat, Aditi 66, 67 Bibao, Aivett 115 Biggs, Bradley W. 138 Bingman, Craig A. 14, 16

Bing, Ryan G. 114 Bin, Hu 171 Blaby-Haas, Crysten E. 66, 67, 88 Blaby, Ian K. 78, 144 Blair, David Dakota 162, 163, 164 Blair, Elaina 92 Blakely, Bethany J. 2 Blazewicz, Steven J. 64, 119, 150, 151, 152 Blenner, Mark 71, 94 Bock, Beatrice 22 Bohrer, Gil 133, 134 Bolduc, Benjamin 130 Bomble, Yannick J. 11, 13, 38, 82 Bonito, Gregory 17 Bonnette, Jason 57 Borton, Mikayla A. 43, 133, 134, 155 Bouffier-Landrum, Amanda 22 Boyle, Nanette R. 72, 85 Bozdag, Nimet Beyza 25 Brailey-Jones, Phil 22 Bramer, Lisa M. 115 Brandizzi, Federica 15 Brandner, Jacob R. 77 Braud, Maxwell 25, 27 Brisson, Vanessa 63, 64, 128 Brizendine, Richard 43 Brock, Jordan 27 Brodie, Eoin 124, 125, 151, 160 Brooks, Matthew 4 Brophy, Jennifer A. N. 70 Brown, Christine 83 Brown Jr., David M. L. 145 Brzostek, Edward R. 2 Buchko, Garry W. 144 Budin, Itay 73 Buell, C. Robin 74 Buie, Cullen 63 Bu, Lintao 13 Burdick, Leah 24, 89, 148 Burley, Stephen K. 167 Burnum-Johnson, Kristin E. 115, 144 Busov, Victor 23 Byers, Neil 160

## С

Cahill, Jesse 112 Cahoon, Edgar B. 74, 75 Callaghan, Melanie M. 40, 42 Camacho, Dimitrios J. 85, 88 Camargo, Antonio 171, 172 Cameron, Jeffrey C. 10 Canagaratna, Manjula 124, 125 Cannon, William R. 43 Canon, Shane 162, 163, 164, 171, 172 Canto Encalada, Gabriela 100 Cao, Liu 156 Carlson, Hans K. 138, 142 Carothers, James M. 76, 77 Carpenter, Timothy S. 97 Carper, Dana L. 24, 111, 148, 149 Carr, Alex 139, 140 Carrell, Alyssa A. 24, 111, 148, 149 Carruthers, David 85 Casas Mollano, J. Armando 93 Casey, John 64 Castro, Claudia 107 Ceja-Navarro, Javier A. 127 Chain, Patrick 136, 137, 171, 172 Chakraborty, Romy 124, 125, 138, 139, 142 Chandonia, John-Marc 138, 162, 163, 164 Chanton, Jeffrey P. 123, 133 Chapla, Digantkumar Gopaldas 38 Chaudhury, Debayan 16 Chen, Fang 10 Chen, Jav 110 Chen, Mingfei 138 Cheong, Jia Hwei 22 Chhetri, Hari 8 Chiappino-Pepe, Anush 77, 78 Chigumba, Desnor N. 156 Chirania, Payal 11 Chivian, Dylan 162, 163, 164 Chlebek, Jennifer L. 105, 106 Cholia, Shreyas 171, 172 Choppararu, Sai Akhil 75 Chopra, Ratan 34 Choudhary, Hemant 19, 20 Chrisler, William 5 Chuang, Derrick 87 Chuckran, Peter 150 Chung, Anny 22 Chung, Sun Woo 54, 55 Church, George M. 77, 78, 79, 80 Chu, Rosalie K. 115

Ciesielski, Peter 13 Claflin, Megan 124, 125 Clarke, Taylor L. 70 Clark, Lauren 112 Clark, Meara 125 Clark, Miranda 24, 111 Clemente, Thomas 30 Clendinen, Chaevien S. 36, 115, 125 Clum, Alicia 171, 172 Coclet, Clement 160 Cogan, Joanna 48 Cohen, Myra B. 169 Cole, Benjamin 22 Coleman-Derr, Devin 107, 108 Coleman, Gary 23 Comai, Luca 28 Connelly, Louis 69 Coon, Joshua J. 42 Corilo, Yuri E. 132, 171, 172 Correr, Fernando Henrique 57 Cors, Jonathan 93 Cort, John R. 144 Corulli, Charles Joseph 38 Cottingham, Robert 162, 163, 164, 169 Couvillion, Sneha 146 Craig, Rory J. 85, 88 Craven, Kelly 64 Crawford, Jason 81 Cregger, Melissa A. 24 Cress, Brady F. 143 Crockett, Zachary 162, 163, 164 Crosby, James 114 Cross, Eben 124 Crowley, Meagan 13 Crowley, Mike 13 Crump, Byron C. 155 Cullen, Daniel 101 Currie, Cameron R. 115 Czech, Lucas 34

## D

da Costa, P. B. 17 Dai, Yi 85 Daletos, Georgios 47 Daly, Rebecca 43 Dang, Chansotheary 120 D'Angelo, Grace 121 Dao, Thai 22 Dardick, Christopher 74 Datta, Sujit 85 Davila, Xiomy Pincha 22 Davis, James J. 155 Davison, Brian H. 60, 61 Davison, Michelle R. 144 Davoudi, Saeedeh 155 Dean, Bruce 7 Deenik, Jonathan 126 Deep, Kamal M. 43 Dehal, Paramvir S. 162, 163, 164, 165, 169 De, Kuntal 110 Demieville, Jeffrey 25, 125 Deng, Kai 18 de Raad, Markus 138 Dervinis, Christopher 30, 31, 32 Deshavath, Narendra Naik 6 De Silva, Kithmee 4 Deutschbauer, Adam M. 54, 108, 138, 139, 142, 143 Devos, Katrien M. 10, 22 Dexter, Gara 108 D'haeseleer, Patrik 166 Diamond, Spencer 142 Diao, Jinjin 90 DiBiase, Beth 40 DiDonato, Nicole 152 Dien, Bruce S. 6, 47 Dilkes, Brian 27 Ding, Lin 103 Ding, Xinxin 36 Dinneny, José R. 25, 70 Doktycz, Mitchel J. 148, 149, 169 Domigues, Tomas 29 Dominguez-Huerta, Guillermo 130 Donohue, Timothy J. 15, 17, 42 Doolev, David 47 Doty, Sharon L. 54, 55 Doudna, Jennifer A. 143 Dow, Ellen G. 162, 163, 164 Downs, Nick 142 Drake, Meghan 162, 163, 164 Duchesneau, Katherine 123 Duenas, Marco A. 88 Duncan, Keith E. 70 Durdic, Karla Ilic 43 Durrett, Timothy P. 74, 75 Dutta, Shuchismita 167 Dwivedi, Dipankar 140

## E

Echauri, Sergio A. Garcia 41 Eckert, Carrie A. 10, 71 Edger, Patrick 27 Edirisinghe, Janaka N. 145, 155, 162, 163, 164, 167 Edwards, Jackson 47 Egbert, Robert 107, 108 Eisenstein, Edward 23 Elias, Damien A. 139 Elkins, James G. 60 Ellenbogen, Jared 130, 133, 134 Ellisman, Mark 73 Ellis, Perry 43 Elmore, Joshua 107, 108 Eloe-Fadrosh, Emiley A. 160, 171, 172 EMERGE Biology Integration Institute Team 130 EMERGE Coordinators 130 Emerson, Joanne B. 116, 117, 118 Enalls, Brandon C. 138 Ence, Daniel 32 Erickson, Erika 43 Estera-Molina, Katerina 64, 126, 127, 150, 152 Eudes, Aymerick 19, 21, 29 Evans, Barbara 61 Eveland, Andrea L. 25, 27 Exposito-Alonso, Moises 34

#### F

Fackler, Nicholas 83 Fagnan, Kjiersten 160, 171, 172 Fan, Yupeng 140 Faria, José P. 162, 163, 164 Farinas, Olivia R. 131 Farris, Yuliya 108, 146 Faulkner, Ian D. 76 Feist, Adam M. 19 Fenster, Jacob 10 Fernandes-Júnior, Paulo Ivan 52 Fernandez, Josue 22 Ferreira, Djennyfer 134 Fields, Matthew W. 139, 140 Fillman, Jason 162, 163, 164 Firestone, Mary K. 64, 127, 128, 150, 152 Fisher, Michael 75 Flaherty, David 7 Flores, Kevin 8 Flynn, Mark 171, 172 Flynn, Regan 78 Flynn, Rory 133, 155 Foley, Megan M. 152 Fonseca, Citlali 107 Fossum, Christina 127 Fox, Brian G. 14, 16 Fox, Jerome M. 10, 74

Fram, Benjamin 43 Francis, Ryan 108 Frank, Andrew 107, 108 Fraser, L. 17 Freiburger, Andrew 169 Freimuth, Paul 158 Fritschi, Felix 17 Fudyma, Jane D. 116 Fu, Jie 5 Fujimoto, Grant 171, 172

#### G

Gado, Japheth 43 Gaffrey, Matthew J. 5 Gallaher, Sean D. 85, 88 Gallo, Luiza 29 Gao, Mu 157 Gao, Yu 19, 21 Gao, Yuqian 115, 144 Garcia, Guadalupe C. 73 García, Lizzette Moreno 98 Garcia, Marci 146 Gattidi, Sai Deepak 25 Gautam, Liza 36 Gauthier, Nicholas P. 43 Geffre, Parker 124 Gehan, Malia 74 Geonczy, Sara E. 116 Gerlt, John 168 Ge, Xiaoxuan 139 Ge, Yufeng 30 Gezan, Salvador 32 Giannone, Richard J. 11, 47, 89 Giberson, Cameron 171, 172 Gille, Derek M. 15 Gil-Loaiza, Juliana 124, 125 Gimode, Winnie 10 Gittrich, Marissa R. 131 Gladden, John M. 19, 20 Gladman, Nicolas 66 Glassman, Sydney I. 117, 118 Glass, N. Louise 142 Goff, Jennifer 139, 140 Gomez-Cano, Fabio 27 Gomez-Rivas, Emma 54 Gonzalez, Emmanuel Miguel 25, 27 Goodchild-Michelman, Isabella 79 Goodstein, David 56 Goudeau, Danielle 22 Grabski, Steven 131 Greenlon, Alex 152 Griffiths, Marcus 35 Grigoriev, Igor 101

Grigory, Justin 125 Grimwood, Jane 56 Groover, Andrew 36 Grosjean, Nicolas 66 Grotewold, Erich 27 Guarnieri, Michael T. 98, 103, 104 Guatney, Lucia S. 155 Guestella, Melina 14 Guest, Jeremy 7 Guler, Mustafa 156 Gunn, Shayla L. 35 Guo, Jiarong 130 Gupta, Chiarg 66 Gupta, Suresh Kumar 27 Gurevich, Alexey 156 Guss, Adam 71, 108 Gu, Tianhao 162, 163, 164

#### Η

Haddad, Eli 97 Hagan, David 124 Ha, Khanh 47 Hall, Ashley N. 17 Hall, Harrison 94 Halvorsen, Tiffany M. 97 Hamilton, Matthew 78 Hammer, Rachel 64 Hanafy, Radwa 95 Handakumbura, Pubudu 36, 107 Hanson, Buck 136, 137 Harman-Ware, Anne E. 10, 13 Harmon-Smith, Miranda 78 Hartman, Melannie D. 2 Harwood, Caroline 108, 149 Hateley, Shannon 34 Hawkes, Christine 64 Hawkins, Charles 36 Hayer, Michaela 119, 120 Hazen, Terry C. 140 He, Fei 75 Hendrickson, Matt 54, 55 Hengge, Neal N. 98 Hennen, Samantha 38 Henriksen, James 79 Henry, Christopher S. 134, 145, 155, 162, 163, 164, 166, 167, 168, 169 Henry, Isabelle M. 28 Hernandez Hernandez, Diana 103 Herrera, Daisy 54 Hessler, Tomas 143 Hestrin, Rachel 64 Hettich, Robert L. 11, 148

Higgins, Riley C. 104 Hillman, Ethan 95 Hofmockel, Kirsten S. 119, 120, 144, 145, 146, 153 Holwerda, Evert K. 11 Homyak, Peter M. 117, 118 Honeker, Linnea 124, 125 Hooker, Casey 95 Howard, Joel 92 Howard-Varona, Cristina 130 Hoyt, David 125, 133, 134 Hren, Andrew 10 Hsu, Chloe 105 Huang, Ching-Ting 22 Hubbard, Allen 69 Huber, George 7 Hu, Bin 172 Hu, Yifeng 90 Hudson, Matt 3, 7, 75 Huenemann, Jay 108 Hungate, Bruce A. 119, 120, 151, 152 Huntemann, Marcel 171, 172 Hunt, Kristopher 139, 140 Hyatt, Doug 8 Hyman, Paul 131

## I

Im, Jinwoo 140 Ireland, A. J. 162, 163, 164 Isaacs, Farren J. 81, 112

## J

Jackson, Elais Player 171, 172 Jacobson, Daniel 8, 32, 36, 111 James, Brandon 5 Jansson, Janet K. 144 Jarboe, Laura 82 Jardine, Kolby J. 29, 124, 125 Jaroensuk, Juthamas 49 Javaid, Tasleem 38 Jawahir, Vannessica 35 Jawdy, Sara S. 24, 148, 149 Jenkins, Jerry 56 Jensen, Rasmus O. 83 Jeon, Jun Hyuk 54, 55 Jernigan, Robert L. 82 Jewett, Michael C. 83, 112 Jha, Ramesh K. 90 Jha, Shantenu 158 Jia, Kejue 82 Jiang, Hui 69, 70

Jian, Siyang 126, 127 Jiao, Yongqin 105, 106 Ji, Heng 7 Jilani, Bilal 45 Jin, Hongyu 30 Jin, Shiyan 18 Jin, Yong-Su 7 Joachimiak, Marcin P. 162, 163, 164 Johnson, Carl H. 49 Johnson, Elizabeth 48 Johnson, Nancy Collins 22 Joshi, Trupti 74 Joyner, Dominique C. 140 Ju, Rebecca 64 Juenger, Thomas E. 17, 56, 57 Juice, Stephanie M. 2 Jungbluth, Sean 162, 163, 164

#### K

Kainer, David 8, 32, 111 Kakouridis, Anne 127, 128 Kalita, Patrick 171, 172 Kamat, Neha P. 40 Kambhampati, Shrikaar 69 Kamimura, Roy 162, 163, 164 Kang-Yun, Christina 105, 106 Kan, Megan 64, 128 Kannan, Aditya 156 Karaoz, Ulas 151, 160 Karlen, Steven D. 14, 16 Katsaras, John 60 Keasling, Jay D. 18, 19, 20, 21 Keller, Keith 162, 163, 164 Kelliher, Julia 136, 137, 171, 172 Kelly, Megan 52 Kelly, Robert M. 114 Kennedy, Nolan 48 Kersten, Roland D. 156 Kessell, Aimee K. 145 Khade, Pranav M. 82 Khangura, Rajdeep S. 27 Khasanova, Albina 142 Kiattisewee, Cholpisit 76 Kijpornyongpan, Teeratas 46 Kimbrel, Jeffrey 63, 64, 128, 151, 152, 166 Kim, Hyu 63 Kim, Jiwoo 153 Kim, Peter 128 Kim, Roy 92 Kim, Soo-Hyung 54, 55 Kim, Young-Mo 54, 108 King, Ethan 43

Kinney, Kyle 17 Kirst, Matias 30, 31, 53 Klauer, Ross 71, 94 Kleese van Dam, Kerstin 158 Kleiner, Manuel 121 Klein, Marie 8 Klingeman, Dawn M. 24, 148 Klos, Dan 162, 163, 164 Koch, Benjamin J. 120, 151 Ko, Dae Kwan 15 Koley, Somnath 49 Kolpakovas, Tadas 83 Konstantinidis, Kostas 123 Kontur, Wayne S. 15 Koo, Abraham J. 75 Köpke, Michael 48, 83 Korenek, Abigail 121 Kosina, Suzanne M. 88 Kostka, Joel E. 123 Kouris, Angela 121 Krechmer, Jordan 124, 125 Krishnamoorthy, Sankarganesh 94 Krummenacher, Benjamin 156 Kuatsjah, Eugene 46 Kubis, Matthew R. 11 Kuchina, Anna 77 Kuehl, Jennifer V. 138 Kuhn, McKenzie 133 Kumar, Ambuj 82 Kumar, Indrajit 25, 27 Kumar, Manish 97, 100 Kumaran, Desigan 66, 67 Kumari, Sunita 66 Kyle, Jennifer E. 115 Kyrpides, Nikos 144

### L

Lachowiec, Jennifer 57 Laemthong, Tunyaboon 114 Lafler, Marissa 64, 128, 152 Lagergren, John 8, 36 Lahiri, Arushi 143 Lai, Edith 128 Lalli, Priscila M. 115 Land, Miriam 162, 163, 164 Landwehr, Grant 83 Lane, Matthew 32 Lankiewicz, Thomas S. 18 Lastiri-Pancardo, Gustavo 98 Lauersen, Kyle J. 88 Lawson, Christopher 92 Leakey, Andrew D. B. 3, 69 Leang, Ching 83

LeBoldus, Jared 111 Lee, Christopher 73 Lee, Daniel 14 Lee, Seungho 25 Lee, Taek Soon 20 Lee, Yi-Yuan 156 Leggieri, Patrick 92 Leinas, James 51 Leonard, Sean P. 105, 106 Leonelli, Laurie 4 Leopold, Jonathan E. 131 Lerma-Ortiz, Claudia 167 Lesea, Hira 138 Li, Gabriella 103, 104 Li, Song 25 Li, Xiaoping 69 Liang, Luna 61 Lieng, Chloe 97 Liew, Fungmin (Eric) 48 Lillington, Stephen P. 18 Lim, Samuel 43 Linger, Jeffrey G. 98, 103, 104 Lingwan, Maneesh 49 Lin, Hsiao-Han 142 Lin, Xiaoxia (Nina) 85 Lin, Yang-Tsung 88 Li, Po-E (Paul) 171, 172 Lipscomb, Gina L. 114 Lipton, Andrew 21 Lipton, Mary S. 85, 121, 144 Liquet, Jose 152 Listgarten, Jennifer 105 Liu, Alexander E. 35 Liu, Degao 69 Liu, Filipe 155, 162 Liu, Haodong 156 Liu, Qun 66, 158 Liu, Yang 110 Liu, Yanmeng 85 Liu, Yihua 121 Llases, Maria 43 Lo, Chien-Chi 171, 172 Lodewyk, Ella 14 Long, Ben 22 Long, Steve 4 Longley, Reid 136, 137 Lovell, John T. 17, 56 Lowry, David B. 17, 57 Lubbert, Collin 69 Lu, Chaofu 57, 58, 74 Lu, Li Chieh 153 Lu, Ting 7 Lu, Yiwei 75 Luebbert, Collin 69

Luecke, Noah 64 Lui, Filipe 163, 164 Lui, Lauren M. 138, 140 Lunny, Elizabeth 124, 125 Lynd, Lee R. 11, 40 Lyons, Eric 25

#### Μ

Maaz, Tai 126 MacIntyre, April 52 Mageeney, Catherine 112 Magni, Lorenzo 80 Mahapatra, Arijit 73 Maitra, Shraddha 6 Malani, Saurabh 41 Malmstrom, Rex 22 Mansfield, Shawn D. 14, 16 Maranas, Costas D. 7, 45 Markillie, Lye Meng 5 Marks, Debora 43 Marks, M. David 34 Marple, Maxwell 152 Marquis, James 140 Marschmann, Gianna 151 Marsh, Alexis L. 169 Marshall-Colon, Amy 5 Martí, Jose Manuel 105 Martinez-Baird, Jessica 108 Masiello, Caroline 153 Maslov, Sergei 7 Mason, Christopher 79 Mau, Rebecca L. 120 Mayali, Xavier 63 May, Jody C. 49 McBrien, Jamie 24 McCalder, Janine 121 McCarthy, Heather R. 57 McClure, Ryan 107, 108, 144, 146 McCue, Lee Ann 171, 172 McDermott, Jason E. 144, 145 McFarlane, Karis 128 McGivern, Bridget 133, 134 McGraw, Robert 158 McIlroy, Simon 133 McKinley, Brian 5 McLean, John A. 49 McLennan, David 24 McQuillan, Josie 85 McReynolds, Melanie 7 Mehta, Radhika 88 Meng, Xianzhi 61 Merchant, Sabeeha S. 85, 87, 88 Meredith, Laura K. 124, 125

Metcalf, Alex J. 72 Meyers, Kaya A. 16 Michael, Jonathan 140 Michener, Joshua K. 89 Milazzo, Marco 79 Miller, Cameron 156 Miller, Christopher S. 133, 155 Miller, Mark 171, 172 Mills, Carolyn 48 Min, Kyungjin (KJ) 128 Mishra, Somesh 6 Mitchell, Douglas 168 Mitchell, Hugh D. 5, 101 Mitta, Neel 156 Mockler, Todd C. 27 Mohimani, Hosein 156 Mondibrown, Sarah 22 Monroe, Matthew E. 115 Monsaint-Queeney, Victoria 119 Monteiro, Lummy Maria Oliveira 71.94 Moon, Tae Seok 90 Moore, Caitlin E. 2 Moore, Ronald J. 115 Morales, Demosthenes 136, 137 Morales, Elisa Y. 35 Moremen, Kelley W. 38 Morgan, Elizabeth 43 Morgan, Mallory 32 Morin, Timothy 133 Morrell-Falvey, Jennifer 60, 148, 149 Morris, Megan 63 Morrissey, Ember M. 120 Mortimer, Jenny C. 19, 21, 142 Moseley, Jeffrey L. 85, 88 Moyne, Oriane 97, 100 Muchero, Wellington 37, 110, 111 Mueller, Alexander P. 48, 83 Mukhopadhyay, Aindrila 20 Mullet, John 5 Mungall, Chris 171, 172 Muñoz Briones, Javier 95 Munoz, Nathalie 54, 115 Mural, Ravi V. 30 Murao, Ayako 103 Murphy, Natasha 43 Murray, Dylan 21 Mutalik, Vivek K. 131, 139, 160, 165 Mutan, Nida M. 35 Myers, Erik Alan 69

## Ν

Nagaraju, Shilpa 83 Nagler, Mike 171, 172 Nakajima An, Davi 157 Nakamura, Akira 81 Narayan, Keshav 156 Narrowe, Adrienne B. 43 Navid, Ali 63, 64, 166 Neal, Maxwell 100 Neely, Christopher 164, 165 Neidle, Ellen L. 90 Nelson, Amelia R. 117 Nelson, William C. 108, 144, 145 Nethery, Matthew 143 Newcomer, Michelle E. 140 Nguyen, Nhu H. 126, 127 Nickels, Jonathan D. 60 Nico, Peter S. 152 Nicolas, Alexa 150 Nicora, Carrie D. 85, 115 Nielsen, Torben 138 Nisonoff, Hunter 105 Niyogi, Krishna K. 87 Noack, Marcus 153 Noel, Cara M. 131 Noguera, Daniel R. 15 Nolan, Alexa 30 North, Justin A. 43 Northen, Trent R. 18, 88, 99, 128, 138, 139, 142, 143, 153 Norton-Baker, Brenna 43 Norton, Grant 97 Novak, Vlastimil 99 Nuccio, Erin 64, 127, 128 Nusinow, Dmitri 35 Nyerges, Akos 78

## 0

Oberg, Nils 168 Ockert, Ingrid 171, 172 Oda, Yasuhiro 108, 149 Ogden, Aaron 108 Oh, Joonseok 81 Oliverio, Angela 133 Olson, Daniel G. 40, 45 O'Malley, Michelle 18, 92 O'Neill, Hugh M. 60, 61 Onyenemezu, Onyeka 82 O'Quinn, Hailey 114 Orlova, Margarita 49 Orr, Galya 5 Orton, Daniel 115 Ostien, Rebecca 48 Ostrov, Nili 78 Osuna-Kleist, Paul 28 Otwell, Anne 139 Owen, Siân V. 78 Owen, Trenton K. 142 Ozersky, Philip 27, 69

## P

Paape, Timothy 66, 67 Palmer, Claire 52 Palmero, Brett 48 Palmo, Joseph 124 Panahabadi, Rahele 57 Pandhal, Jagroop 85 Pang, Changxu 66 Parasurama, Sriram 54, 55 Pareek, Manish 95 Park, Dan 105, 106 Parker, Charles 160 Park, Gilchan 165 Parks, Jerry M. 157 Parrott, Wavne 74 Parton, William J. 2 Pasa-Tolic, Ljiljana 152 Pasquini, Miriam 66 Passi, Anurag 100 Patel, Jaymin 81 Patil, Sujay 171, 172 Pauli, Duke 25, 27, 125 Paulitz, Timothy 58 Pavicic, Mirko 8, 32 Pawlowski, Emily 27 Peabody V, George 108 Pearson, Erik 162, 163, 164 Pelech, Elena 4 Pelletier, Dale A. 148, 149, 169 Pena, Maria J. 38 Pendergast IV, Tom H. 10, 22 Peng, Hao 58 Pereira, Wendell 31 Perveen, Irum 47 Peter, Gary F. 32 Petersen, Bryan M. 2 Peters, Jason M. 17 Petridis, Loukas 61 Petro, Caitlin 123 Pettiford, Jordan Edward 25 Pettinga, Dean 107 Pett-Ridge, Jennifer 64, 120, 126, 127, 128, 150, 151, 152 Petzold, Christopher J. 18 Pfleger, Brian F. 7, 49

Piatkowski, Bryan T. 148, 149 Piehl, Dennis 167 Piehowski, Paul 115, 171, 172 Pingali, Sai Venkatesh 60, 61 Platts, Adrian 27 Poirier, Brenton 108 Poole, Farris 139, 140 Pope, Maddison 69 Prabhakar, Pradeep Kumar 38 Pratama, Akbar Adjie 130 Presley, Gerald N. 89 Price, Gavin 162, 163, 164 Prymolenna, Anastasiya 171, 172 Purcell, Alicia M. 120 Purvine, Samuel O. 85, 171, 172 Putt, Andrew 140 Pu, Yucen 14, 16 Pu, Yunqiao 61 Pyne, Saptarshi 52, 53

## Q

Qi, Yiping 23 Quatrini, Paola 79 Quinto, Laura 81, 112

## R

Rabinowitz, Joshua 7 Rabus, Jordan 132 Rackers, Rachel 22 Ragauskas, Arthur J. 61 Raimondo, Morgan 54 Rajakaruna, Sumudu 130, 132 Ralph, John 14, 16 Ramírez-Flores, María del Rosario 24 Ramon, Christina 64 Rand, Eleanor A. 78 Ranft, Hannah 83 Rangamani, Padmini 73 Ranjan, Priya 37, 162, 163, 164, 169 Rao, Christopher 7 Ray, Prasun 64 Regulski, Michael 66 Rellán-Álvarez, Rubén 34 Rempfert, Kaitlin R. 119 Reynoso, Vinicio 83 Rhee, Seung Y. 36 Ricci, Dante P. 105, 106 Rich, Virginia 133 Ridgeway, Joanna R. 2 Riehl, William 162, 163, 164 Riley, Lauren 108

Riley, William 134 Roberts, Avery 143 Robinson, Aaron 136, 137 Rochereau, Charlotte 105 Rodionov, Dmitry 114 Rodriguez, Alberto 20 Rodriguez, Francie 171, 172 Rodriguez-Zaccaro, Daniela 36 Rodríguez-Zapata, Fausto 34 Rogers, Alistair 158 Rollins, Nathan 43 Romero, Sandra Gomez 72 Roscioli, J. Rob 124, 125 Roth, Bryan L. 156 Roth, Melissa S. 87 Roth, Spencer 24, 123 Roux, Simon 160, 171, 172 Roy, Jetendra K. 82 Roy, Sushmita 30, 31, 52, 53 Rudolph, Alexandra 78 Rush, Tomas 110, 111 Rutter, Erica 8

## S

Ryon, Krista 79

Ryu, Seunghyun 47

Sachdeva, Rohan 150 Sachsenberg, Timo 121 Sacks, Erik J. 6 Sadkhin, Boris 162, 163, 164 Sakkaff, Zahmeeth 155 Saleska, Scott 133 Salvachúa, Davinia 46 Sànchez i Nogué, Violeta 98 Sanchez, Jesus 107 Sanciangco, Millicent 57 Sander, Chris 43 Sanderson, Courtney M. 131 Sangha, Amandeep 51 Santibañez, Rodrigo 100, 103 Santos-Medellín, Christian 116 Sarrafan, Set 171, 172 Sauro, Herbert M. 76 Savage, David F. 97 Sawyer, Cole 111 Schadt, Christopher W. 123 Schaefer, Amy 149 Scharwies, Johannes Daniel 70 Scheller, Henrik 19, 20, 21 Schilling, Jonathan S. 101 Schindel, Heidi 48, 83 Schlatter, Daniel 58 Schmitz, Bob 74

Schmutz, Jeremy 56 Schnable, James C. 30 Schoeniger, Joe 112 Schoonen, Martin A. 158 Schroeder, Wheaton L. 45 Schubert, Max G. 77, 79 Schwartz, Alexa 46 Schwartz, Egbert 120 Scott, Haden L. 60, 61 Scown, Corinne D. 20 Seaver, Samuel 66, 162, 163, 164, 167 Sebastian, Jose 70 Sebesta, Jacob 103 Sedbrook, John 34, 35, 36 Seelig, Georg 77 Seleman, Alan 162, 163, 164 Sethian, James 153 Shabek, Nitzan 36 Shaffer, Michael 155 Shah, Manesh 32 Shah, Riddhi 61 Shakya, Migun 171, 172 Shanklin, John 6, 7, 158 Shao, Zengyi 7 Sharma, Bishal D. 40 Shaw, Trey 75 She, Fukang 42 Shen, Yihui 7 Shen, Yue 78 Shepard, Hawkins S. 49 Sher, Andrew W. 54, 55 Shih, Patrick 20, 74 Shin, Janis 76 Shorter, Joanne 124, 125 Shrestha, Ritu 107, 108 Shu, Mengjun 110 Shu, Shengqiang 56 Sieradzki, Ella T. 150 Sieren, Hailey 14 Silberg, Jonathan J. 153 Simmons, Blake A. 18, 19, 20 Simmons, Tuesday 108 Singer, Steven 18, 20 Singh, Atinder 4 Singh, Jyoti 71, 94 Singh, Vijay 6, 7 Singleton, Derick 155 Sinha, Saurabh 7 Skolnick, Jeffrey 157 Slanzon, Giovana Simão 127 Slessarev, Eric 128, 152 Slocum, Samuel T. 156 Smanski, Michael J. 74, 93

Smith, Brayden Paul 38 Smith, Clyde A. 144 Smith, Heidi 139 Smith, Jeremy C. 60, 61 Smith, Micholas Dean 60, 61 Smith, Montana 144, 171, 172 Smith, Rebecca A. 14, 16 Sokol, Noah W. 152 Solomon, Kevin 71, 94, 95 Som, Suman 29 Song, Hyun-Seob 145 Sorensen, Patrick O. 160 Soto, Carlos X. 165 Span, Elise 51 Spangler, Margaret 10 Spann, Sedona 22 Spatafora, Davide 79 Spencer, Owen 78 Stahl, David A. 139, 140 Stanley, Robert 22, 125 Starker, Colby G. 69, 70 Stegen, James 155 Stein, Johannes 80 Stemple, Brooke 133 Stephanopoulos, Gregory 47 Stevenson, David M. 40 St. John, Peter 82 Stone, Bram W. G. 119, 120 Streich, Jared 8 Strous, Marc 121 Stuart, Rhona 63, 64, 166 Studt, Jacob E. 2 Sullivan, Matthew B. 130, 131 Sun, Zimou 14, 16 Suzuki, Yo 103 Swaminathan, Kankshita 5 Sweedler, Jonathan 3, 7

#### Т

Tadesse, Dimiru 66, 67 Tamborski, Janina 70 Tamburro, Jacob 85 Tandukar, Nirwan 34 Tang, Jinyun 133 Tan, Grace 69 Tang, Tzu-Chieh 79 Tan, Luoxi 60 Tannous, Joanna 110, 111 Tanwee, Tania N. N. 114 Tas, Huseyin 77 Tayengwa, Reuben 23 Taylor, Gail 8, 24 Tec-Campos, Diego 100 Tejeda-Lunn, Daniel 69 ter Horst, Anneliek M. 116 Terry, Gwyneth 162, 163, 164 Tfaily, Malak M. 123, 124, 125, 130, 132, 133 Thadani, Nicole 43 Thairu, Margaret W. 115 Thapa Magar, Roniya 160 Thelen, Jay J. 74, 75 Thiruppathy, Deepan 97, 98, 100 Thomas, David J. 57 Thomas, Jason 34, 36 Thorgersen, Michael 139 Thusoo, Eashant 40 Tian, Funing 130 Tian, Yang 19 Tian, Yuxin 90 Tibebu, Redeat 69 Tibocha-Bonilla, Juan D. 103 Tiedge, Kira 38 Tierney, Braden 79 Timilsena, Prakash Raj 25 Tolić, Nikola 121 Tonelli, Marco 16 Topp, Christopher N. 35, 70 Torkelson, Ella R. 16 Torn, Margaret 125 Torres Béjar, Marta 142 Torres, Isaac 22 Tournay, Robert 54, 55 Tran, Quoc 77 Tringe, Susannah G. 58, 153 Trinh, Cong T. 47 Trippe, Richard 22 Trotter, Valentine V. 108, 138 Trujillo, Edna 42 Tsai, Chung-Jui 37, 74 Tsai, Helen 28 Tullman-Ercek, Danielle 48 Tumen-Velasquez, Melissa 90 Turco, Gabriele 79 Turkarslan, Serdar 139 Tuskan, Gerald A. 8, 10, 11, 13 Tyson, Gene 133 Tzeng, Jung-Ying 34

#### U

Upadhyaya, Shivani 87 Upton, Racheal 57 Urbanowicz, Breeanna R. 38, 74 Urvoy, Marion 130 Ushizima, Daniela 153

#### V

Vailionis, Jason 114 Valentino, Hannah R. 89 Valenzuela, Jacob J. 139 Vallat, Brinda 167 Van Fossen, Elise 108 VanLoocke, Andy 2, 3 Van Wallendael, Acer 17 Varner, Ruth 133 Vasdekis, Andreas 47 Vazquez, Julio Rivera 42 Vela, Saddie 52 Velickovic, Dusan 22, 54, 115, 125 Velickovic, Marija 115 Venado, Rafael E. 53 Venkatraman, Kailash 73 Venturelli, Ophelia 52 Vermerris, Wilfred 30, 31, 52 Viana, Willian Goudinho 70 Vigliotti, Mark 85 Vilbert, Avery C. 15 Villa, Jorge 133, 134 Vinke, Svenja 78 Violette, Marlene 121 Vögeli, Bastian 83 Vogel, John P. 99 Voigt, Christopher 51 Voytas, Daniel F. 69

### W

Wainaina, James M. 130 Wakao, Setsuko 87 Waldbauer, Jacob 40 Walian, Peter J. 138, 139 Walker, Kathleen F. 140 Walkup, Jeth 120 Walley, Justin 36 Wang, Bo 49 Wang, Chao 120 Wang, Daifeng 66 Wang, Harris 105 Wang, Jia 148 Wang, Shen 156 Wang, Shenqi 81 Wang, Shi 160 Wang, Sumin 162, 163, 164 Wang, Yajiao 140 Wang, Yunxuan 61 Ware, Doreen 66 Webb, Amber 149 Webb-Robertson, Bobbie-Jo M. 115 Weber, Peter 63, 64

Weekley, Brock 75 Wei, Hairong 23 Weisenhorn, Pamela 162, 163, 164 Weitz, David 43 Welti, Ruth 74 Wergeles, Nicholas 75 Werner, Allison Z. 40, 43 Wessel, Austin 63 Weston, David J. 24, 148, 149, 169 Wilker, Jennifer 52 Wilkes, Rebecca A. 40 Wilkins, Michael J. 117, 118, 134 Williams, Jonathan 42 Williams, Kelly 112 Williams, Sarai 115 Willis, Morgan J. 38 Willmore, Cody 58 Wilson, Andrew 107, 108 Wilson, James 139 Wilson, Rachel M. 123, 133 Winkler, Tanya 54 Winston, Donny 171, 172 Wolf, Emily 52 Wollard, Jessica 64, 128 Wood-Charlson, Elisha 134, 162, 163, 164 Woodworth, Sean 46 Wright, Aaron 94 Wrighton, Kelly C. 43, 130, 133, 134, 155 Wu, Chao 82, 103 Wu, Chao-ting 80 Wühr, Martin 7 Wu, Ruonan 115, 144 Wu, Xing 34 Wyatt, Gabby 38 Wymore, Ann 24

## X

Xiao, Yuguo 25, 27 Xie, Meng 66, 67 Xiong, Wei 103 Xu, Chunhui 75 Xu, Dong 74, 75 Xu, Wentao 43 Xu, Yan 171, 172 Xu, Yao 49

## Y

Yaguchi, Allison 46 Yang, Bin 103 Yang, Jinliang 30 Yang, Wendy H. 2 Yang, Xiaohan 110, 149 Yang, Zhifeng 127 Yang, Ziming 162, 163, 164 Yan, Shirui 78 Yee, Estella 67 Yee, Mon O. 142 Yin, Chuntao 58 Yin, Peng 80 Yoo, Shinjae 162, 163, 164, 165 York, Larry 8 Yoshikuni, Yasuo 3, 7, 78, 144 Young, Jamey D. 49 Young, Robert 125 Young, Sarah, J. 43 Yuan, Mengting 126, 127 Yu, Jianping 103 Yung, Mimi Cho 97, 106

#### Ζ

Zahn, Michael 46 Zalatan, Jesse G. 76, 77 Zanghellini, Alexandre 51 Zarei, Ariyan 25 Zayed, Ahmed A. 130 Zemaitis, Kevin 115 Zengler, Karsten 97, 98, 99, 100, 103 Zerbe, Philipp 38 Zhai, Chengxiang 7 Zhai, Yi 99 Zhalnina, Kateryna 142 Zhang, Jiwei 101 Zhang, Ke 114 Zhang, Lifang 66 Zhang, Qizhi 162, 163, 164, 167 Zhang, Shufan 22

Zhang, Xinge 43 Zhang, Ying 114 Zhao, Huimin 3, 6, 7 Zhao, Jianxin 10 Zhao, Rui 115 Zhou, Jizhong 126, 127, 140 Zhou, Nanquing 40 Zhu, Ying 115 Ziegler, Samantha J. 38 Zimmerman, Amy E. 54, 119, 146 Zucker, Jeremy 76 Zuniga, Baltazar 49 Zwart, Peter 153