

Office of Science

# 2017 Genomic Sciences Program Annual PI Meeting

Abstract Book

## February 6-8, 2017

Sponsored by the U.S. Department of Energy Office of Biological and Environmental Research Biological Systems Science Division **Department of Energy Washington, DC 20585** February 6, 2017

Dear Colleague:

On behalf of the Biological Systems Science Division (BSSD) welcome to the 2017 Genomic Science Annual PI Meeting! This year starts with some uncertainty due to the transition in administration. While numerous opinions abound in the media regarding DOE's programs, we remain optimistic about the future. Within the division, our basic science efforts in fundamental biology underpin advances in bioenergy and also form the basis for biotechnology development in general. Just as our earlier Human Genome Project sequencing technology development efforts resulted in large economic returns on initial investments, our current efforts to broaden the understanding of plant and microbial genomes for a range of beneficial purposes will produce similar returns. As a division, we will continue to stress how your accomplishments in DOE mission-relevant genomic science are providing the discoveries needed to underpin whole new biotechnology-based industries with enormous potential economic impact for the United States and the world.

This annual meeting is an opportunity for all principal investigators (PIs) funded within the Genomic Science Program to review their latest research results in the context of the entire program and develop new ideas with colleagues. I would especially urge new PIs to take full advantage of this opportunity to meet your colleagues and look for opportunities to further enhance the innovation and success of your research efforts. This is also an opportunity to meet with your DOE program staff, representatives from elsewhere within the Department of Energy, and colleagues from other Federal Agencies.

As a division we are continually seeking to craft a portfolio that pairs the creative talents of researchers with enabling capabilities and user facilities to maximize the potential for scientific discovery and achievement. At this meeting there will be plenary and poster presentations of a variety of enabling capabilities supported either directly by the division and/or BER, including the DOE Joint Genome Institute (JGI), the Environmental Molecular Science Laboratory (EMSL), the DOE Synchrotron Light and Neutron Sources, the DOE Systems Biology Knowledgebase (KBase), and new bioimaging technology development efforts. We are looking at creative ways to facilitate user access at these facilities to address the needs of multidisciplinary science. I would urge you to visit with representatives of these capabilities at this meeting to learn about some of these new efforts.

This past year has been another exciting and productive year for the Genomic Science program. We have planned a full agenda packed with compelling presentations throughout, highlighting the exceptional research results produced by the program. We hope that these presentations will spark fruitful discussions that can be carried into the poster sessions, which have become the heart of the meeting. We will hear the latest results from the Bioenergy Research Centers (BRCs) and talks from each element of the entire portfolio at this meeting. As most of you are aware, the BRCs are in their final year of funding. This year's BRC talks will highlight scientific contributions from early career researchers discussing their interdisciplinary research findings in the context of these large integrated research centers.

We are extremely pleased to welcome Dr. Robin Buell, MSU Foundation Professor of Plant Biology at Michigan State University, as our keynote speaker. Our plant biology efforts within the division are at the very core of our bioenergy and bioproducts research efforts. Dr. Buell's scientific contributions have been instrumental in helping to identify and develop beneficial traits in a broad range of agricultural and bioenergy-relevant plant species. Dr. Buell's approach to plant biology is an excellent example of fundamental research we hope to foster within our programs with broad applicability across numerous disciplines. We are confident Dr. Buell's presentation will be an excellent catalyst for scientific discussion throughout the meeting.

Continuing a long-standing tradition in alternate years, researchers supported by the joint USDA-DOE Plant Feedstock Genomics for Bioenergy program will also be in attendance. This program focuses on genomicsbased research leading to the improved use of biomass and plant feedstocks for producing biofuels and bioproducts, with recent targets on disease resistance in bioenergy feedstocks and oilseed crops. Convening with other Genomic Science program PIs will provide an ideal opportunity for the stimulation and exchange of ideas, sharing of expertise, and formation of new partnerships. Both a plenary session and a breakout session will be held to discuss some of the very important developments in feedstock genomics research that have taken place since these groups last met, and to highlight the most recent awards. The new USDA program manager, Dr. Jeffrey Steiner, National Program Leader at The National Institute of Food and Agriculture (NIFA), will be introduced.

This year's breakout sessions will cover a broad range of topics in the areas of plant and microbial biology for bioenergy purposes, including microbiome research. Additional sessions will highlight new bioimaging projects, KBase hands-on sessions to demonstrate and experience the latest capabilities and new aspects of the software development kit (SDK), and DOE user facilities capabilities and resources. We will also be co-hosting a breakout session in computational biology with our colleagues from the Scientific Discovery through Advanced Computing (SciDAC) program of DOE's Office of Advanced Scientific Computing Research (ASCR). We hope these breakout sessions are informative and help to foster new collaborations across DOE programs.

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

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

Sincerely,

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

Table of Contents (Click on Entries to advance to the abstract)	
Daniel Amador-Noguez: Limited thermodynamic driving force in glycolysis of cellulolytic PRESENTER: Tyler Jacobson	c clostridia Page 25
<b>Paul Adams</b> : High-quality protein production for structural studies: From plant cell wall microbial production of bioproducts.	synthesis to
PRESENTER: Giovani Pinton Tomaleri	Page 26
Harry Beller: Discovery of a Novel Bacterial Enzyme Enabling First-Time Biochemical Pro Toluene	duction of
PRESENTER: Harry Beller	Page 27
Nathan Hillson: j5 Software Through the Years: Insights from Aggregate Public Usage Me PRESENTER: Nathan Hillson	etrics Page 29
Taek Soon Lee: Isopentenyl diphosphate (IPP)-bypass mevalonate pathways for C5 alcoh PRESENTER: Taek Soon Lee	nol production Page 30
Jon Magnuson: Rhodosporidium toruloides for conversion of depolymerized cellulose, hemicellulos and lignin into bioproducts	
PRESENTER: Junko Yaegashi	Page 32
Jenny Mortimer: Characterization of Plant Golgi-Localized Nucleotide Sugar Transporter PRESENTER: Jenny Mortimer	rs Page 34
<b>Christopher Petzold</b> : High-throughput quantitative proteomic profiling of Escherichia co metabolism	li central carbon
PRESENTER: Christopher Petzold	Page 36
Pamela Ronald: Functional analysis of cell wall related genes in Sorghum bicolor PRESENTER: Tong Wei	Page 37
<b>Corinne Scown</b> : Life Cycle Assessment of Ionic Liquid-based Biofuel Production PRESENTER: Binod Neupane	Page 38
Steven Singer: A Bacterial Pioneer Leaves a Complex Legacy PRESENTER: Steven Singer	Page 40
Anup Singh: Droplet-Based Analog and Digital Microfluidic Platforms for High-Throughpe	ut Screening and
Synthetic Biology Applications PRESENTER: Kosuke Iwai	Page 42
Seema Singh: Ionic Liquid Pretreatment Technology: Challenges and Opportunities PRESENTER: Seema Singh	Page 44

Thomas Eng: Genome Wide Identification of Bacterial Membrane Capacity Determinant PRESENTER: Thomas Eng	s Page 45	
Hector Garcia Martin: Quantitative metabolic modeling at the Joint BioEnergy Institute PRESENTER: Hector Garcia Martin	(JBEI) Page 47	
<b>Dominique Loque</b> : Development of Novel Approaches to Optimize Energy Crops PRESENTER: Dominique Loque	Page 48	
Trent Northen: Linking Soil Biology and Chemistry Using Knowledge of Exometabolite Su	ubstrate	
PRESENTER: Tami Swenson	Page 49	
<b>Kenneth Sale</b> : Formation of a phenyl-choline ether structure in lignin reduces inhibition activity by lignin	of cellulase	
PRESENTER: Jijiao Zeng	Page 51	
Henrik Scheller: At the Interface: Glycoproteins, Glycolipids, and WAK-mediated Signalin for Plant-Microbial Symbiosis in Medicago truncatula	ng are Required	
PRESENTER: William Moore	Page 5 2	
<b>Paul Adams</b> : Exposure History Dependence of Microbial Mediated Substrate Transforma Groundwater	ation Rates in	
PRESENTER: Charles Paradis	Page 54	
<b>Paul Adams</b> : High-throughput testing of carbon source on microbial community assemb production	ly and antibiotic	
PRESENTER: Lauren Lui	Page 55	
<b>Paul Adams</b> : Identification of Novel Biosynthetic and Catabolic Pathways in Diverse Bacteria Using High-		
PRESENTER: Adam Deutschbauer	Page 56	
Paul Adams: Mechanism for Microbial Population Collapse in a Fluctuating Resource Empresenter: Serdar Turkarslan	vironment Page 58	
Paul Adams: Mechanisms of Uranium Reduction in Sulfate-Reducing Bacteria PRESENTER: Erica Majumder	Page 60	
Paul Adams: Microbes at the blurred boundary of natural and built environments PRESENTER: Fangqiong Ling	Page 61	
Paul Adams: Microbial Interactions with Natural Organic Matter Extracted from the Oak PRESENTER: Romy Chakraborty	Ridge FRC Page 62	
<b>Paul Adams</b> : Multiplex characterization of microbial traits using dual barcoded genome fragment		
PRESENTER: Vivek Mutalik	Page 63	

<b>Paul Adams</b> : Stress mediates relative importance of deterministic and stochastic assem groundwater microbial communities	bly in
PRESENTER: Daliang Ning	Page 65
Paul Adams: Systems Biology Guided by Global Isotope Metabolomics PRESENTER: Erica Forsberg	Page 67
Paul Adams: Temporal Variability and Microbial Activity in Groundwater Ecosystems PRESENTER: Heidi Smith	Page 69
<b>Paul Adams</b> : The role of adaptive evolution in shaping the structure and function of mo consortia	del microbial
PRESENTER: David Stahl	Page 71
<b>Paul Adams</b> : Understanding the thermodynamic Foundations of microbial Growth Effici and Field	encies in the Lab
PRESENTER: Frederick von Netzer	Page 73
<b>Robert Clubb</b> : Quantitative Measurements of Cellulase Display in the Model Gram+ Microbe Bacillus subtilis Define Determinants Required for Enzyme Display and Activity	
PRESENTER: Grace Huang	Page 75
Jennifer Reed: Approaches for Evaluating the Production Potential of High Volume Products in Microb Systems	
PRESENTER: Matthew Long	Page 76
<b>Sabeeha Merchant</b> : Functional analysis of copper and silver storage sites and their role homeostasis in Chlamydomonas	in metal
PRESENTER: Kristen Holbrook	Page 78
<b>Todd Yeates</b> : Design, Characterization, and Emerging Applications of Highly Symmetric Nanostructures	Protein
PRESENTER: Kevin Cannon	Page 79
<b>Daniel Amador-Noguez</b> : Development of metabolomic methods for investigating metabolic regulation	
PRESENTER: Julia Martien	Page 81
Federica Brandizzi: Optimizing MLG Production for Improved Biofuel Crop PRESENTER: Starla Zemelis-Durfee	Page 83
Joshua Coon: GLBRC Targeted Metabolomics PRESENTER: Alan Higbee	Page 86
<b>Cameron Currie</b> : Extracellular Products Mediate Bacterial Synergism in Cellulose Degrad PRESENTER: Camila Carlos	lation Page 88

<b>hi-You Ding</b> : Quantification of monolignol ferulate conjugate in Zip-lignin poplar by stimulated Rama cattering microscopy	
PRESENTER: Wei Shen	Page 91
<b>Timothy Donohue</b> : Using Genetic and Bioreactor Engineering to Produce Oleaginous Bac PRESENTER: Kimberly Lemmer	cteria Page 92
<b>Brian Fox</b> : A Cellulase Family Reunion: Observing and Predicting the Structural Changes the Evolution of GH5 Enzyme Specificity	Accompanying
PRESENTER: Evan Glasgow	Page 94
<b>Eric Hegg</b> : Using N and O Isotopes to Determine the Source of Microbial N2O Production PRESENTER: Joshua Haslun	n Page 95
Shawn Kaeppler: Transcriptional Analysis of Flowering Time in Switchgrass PRESENTER: Carl-Erik Tornqvist	Page 97
Robert Landick: Optimization of Isobutanol Production by Zymomonas mobilis PRESENTER: Indro Ghosh	Page 98
Christos T Maravelias: Process Systems Engineering for Biofuels and Bio-based Chemica PRESENTER: Rex Tong Lip Ng	ls Page 100
John Ralph: Suppression of CINNAMOYL-CoA REDUCTASE increases the level of monolig	nol-ferulates in
PRESENTER: Rebecca Smith	Page 102
Srivatsan Raman: Designing highly specific protein-based small molecule biosensorsDesigning highly specific protein-based small molecule biosensors	
PRESENTER: Vatsan Raman	Page 104
Curtis Wilkerson: Exploring RNA-Seq Expression Data With GxSeq PRESENTER: Nicholas Thrower	Page 105
Yaoping Zhang: Comparative Studies of Diverse Feedstocks and Identification of High Levels of	
PRESENTER: Yaoping Zhang	Page 107
Trent Northen: SELF-ASSEMBLED GOLD NANOPARTICLE FILM FOR NANOSTRUCTURE-INITIATOR MASS	
PRESENTER: Todd Duncombe	Page 109
Phil Robertson: Methods for high-throughput massively parallel soil ecofunctional gene PRESENTER: James Cole	analysis Page 111
G. Philip Robertson: Associative Nitrogen Fixation across a Nitrogen Input Gradient PRESENTER: G. Philip Robertson	Page 113

John Mullet: Utility of a Sorghum bicolor RNA atlas for improving energy sorghum stem PRESENTER: Brian McKinley	composition Page 115	
<b>G Philip Robertson</b> : Partitioning Nitrous Oxide (N2O) Emissions from Ammonia Oxidizing Bacteria (AOB) and Ammonia Oxidizing Archaea (AOA) in Corn and Switchgrass Ecosystems		
PRESENTER: Di Liang	Page 116	
<b>Paul Gilna</b> : Characterization of Populus Transgenic Plants Overexpressing PtDUF266A (C and Biofuel Production	XPtDUF266A)	
PRESENTER: Chang Geun Yoo	Page 118	
<b>Paul Gilna</b> : Defined Tetra-Allelic Gene Disruption of the 4-Coumarate:Coenzyme A ligase CRISPR/Cas9 in Switchgrass Results in Lignin Reduction and S/G Ratio Alteration	e 1 Gene by	
PRESENTER: Jongjin Park	Page 120	
Paul Gilna: Evaluation of Multiple Levers for Overcoming the Recalcitrance of Cellulosic PRESENTER: Evert Holwerda	Biomass Page 122	
Paul Gilna: Explaining and Predicting Biomass Recalcitrance with Rigidity Percolation The PRESENTER: Erica Gjersing	eory Page 124	
Paul Gilna: Field Experiments of Nine Switchgrass TOP Lines PRESENTER: Holly Baxter	Page 126	
Paul Gilna: From Gene to Network: Switchgrass TOP Line RNA-seq Data Analysis Pipeline PRESENTER: Yuhong Tang	e at BESC Page 128	
<b>Paul Gilna</b> : Gene Expression Differences between Clostridium thermocellum Biofilm and	l Planktonic Cells	
PRESENTER: Alexandru Dumitrache	Page 130	
Paul Gilna: Investigating the Role of Polysaccharide Methylation in the Plant Cell Wall PRESENTER: Breeanna Urbanowicz	Page 132	
Paul Gilna: Lignin Valorization through Biologial Funneling PRESENTER: Gregg Beckham	Page 134	
Paul Gilna: Metabolic Network Modeling of Clostridium thermocellum for Systems Biolo	ogy and	
PRESENTER: Cong Trinh	Page 136	
<b>Paul Gilna</b> : Modification of Pectin Biosynthesis Leads to Higher Biomass Yield and Sacch Bioenergy Feedstock	arification in	
PRESENTER: Ajaya Biswal	Page 138	
Paul Gilna: Modifying Carbon, Nitrogen, and Electron Metabolism in Clostridium thermocellum to Enhance Cellulosic Biofuel Yield and Titer		
PRESENTER: Adam Guss	Page 140	

Paul Gilna: New Insights on the Lignin Pathway in Grasses	
PRESENTER: Juan Carlos Serrani-Yarce	Page 142
Paul Glina: Pleiotropy Decomposition of 609 Populus Trichocarpa Genotypes	
PRESENTER: Damei Jacobson	Page 144
Paul Gilna: Pyruvate Decarboxylase: Rationally Evolving Thermostable Enzymes for Met	abolic
Engineering	
PRESENTER: Deanne Sammond	Page 146
Paul Gilna: Rational Methyltransferase Expression in Escherichia coli for Transformation	of New
Organisms	
PRESENTER: Lauren Riley	Page 148
Paul Gilpa: Signaling Potween Switchgrass and Europal	
PRESENTER: Prasun Ray	
Thesenten. Thasan hay	1 age 150
Paul Gilna: Switchgrass Fermentation by Thermophilic Microbiomes	
PRESENTER: Xiaoyu Liang	Page 152
Paul Gilna: The Effect of Lignin and Hemicellulose Removal on Switchgrass Deconstructi	on by
Clostridium thermocellum	
PRESENTER: Ninad Kothari	Page 152
Paul Gilna: The Unique Mechanism of the Dominant Multi-Component Cellulase from	
Caldicellulosiruptor bescii	
PRESENTER: Yannick Bomble	Page 156
	C
Thomas Brutnell: High throughput phenotyping and quantitative genetics to understand productivity	
and drought traits in the model C4 grass Setaria	
PRESENTER: Ivan Baxter	Page 157
Themas Brutnelly Utilizing Cotorio visidie of a Madel for Malagular Characterization of la	cmanata
Mediated Growth and Defense Responses	Smonale-
PRESENTER: Christine Shvu	Page 158
Thomas Brutnell: Quantitative Trait Loci for Leaf Carbon Isotopic Signature and Transpiration Efficiency	
in the C4 grass Setaria	
PRESENTER: Asaph Cousins	Page 159
Rachel Brem: Functional Genomics of Lipid Accumulation in Rhodosporidium toruloides	Dago 161
FRESENTER, Samuel Colduelli	Lage TOT
Lawrence Smart: Genomics-Assisted Breeding for Leaf Rust (Melamosora) Resistance in Shrub Willow	
(Salix) Bioenergy Crops	
PRESENTER: Lawrence Smart	Page 162

John Dyer: Genomics and Phenomics to Identify Yield and Drought Tolerance Alleles for Camelina as a Biofuel Crop	Improvement of
PRESENTER: John Dyer	Page 164
John Dunbar: Ecological Factors Affecting Carbon Flow from Surface Litter PRESENTER: John Dunbar	Page 166
John Dunbar: Microbial Community Traits Linked to Carbon Flux Patterns in Soil PRESENTER: Renee Johansen	Page 168
<b>Cheryl Kuske</b> : Comparative Analyses of the Genomes and Secretomes of Ascomycota Fu Diverse Functions in Plant Biomass Decomposition	ngi Reveal
PRESENTER: Jean Challacombe	Page 169
<b>Cheryl Kuske</b> : Meta-omic, Enzymatic and Soil Chemical Measures Identify Surface Soil Decompositio Processes as Influenced by N Amendment in a Temperate Pine Forest	
PRESENTER: Cheryl Kuske	Page 171
Jay Chen: Plant-Microbe Interfaces: Populus genomics, genetics and molecular biology PRESENTER: Jay Chen	Page 173
<b>Caroline Harwood</b> : Plant-Microbe Interfaces: Probing the molecular mechanisms of plant-bacterial interactions	
PRESENTER: Caroline Harwood	Page 175
<b>Daniel Jacobson</b> : Plant-Microbe Interfaces: Plant-based genome-wide association virior analysis (GWAVA)	me/microbiome
PRESENTER: Gerald Tuskan	Page 177
Jennifer Morrell-Falvey: Plant-Microbe Interfaces: Comparative genomics and function characterization of Populus-associated endophytes	al
PRESENTER: Jennifer Morrell-Falvey	Page 179
<b>Dale Pelletier</b> : Plant-Microbe Interfaces: Understanding the Populus microbiome struct to host stress	ure in response
PRESENTER: Gregory Hurst	Page 181
<b>Christopher Schadt</b> : Plant-Microbe Interfaces: Dissecting the microbiome of Populus tree species from the soil to the canopy using amplicon sequencing and shotgun metagenomic analyses	
PRESENTER: Melissa Cregger	Page 183
<b>Christopher Schadt</b> : Plant-Microbe Interfaces: Linking host genotype fitness and soil conditions to microbiome community assembly in the Populus root-soil interface	
PRESENTER: Christopher Schadt	Page 184

<b>Timothy Tschaplinski</b> : Plant-Microbe Interfaces: Metabolic consequences of the introduction of a Populus trichocarpa lectin receptor-like kinase into Arabidopsis thaliana, a non-ectomycorrhizal host species.	
PRESENTER: Timothy Tschaplinski	Page 185
<b>Rytas Vilgalys</b> : Plant-Microbe Interfaces: Characterizing the diversity and function of the	
PRESENTER: Rytas Vilgalys	Page 187
<b>David Weston</b> : Plant-Microbe Interfaces: Model communities of Populus bacterial isolar mechanisms of microbiome function	tes to study
PRESENTER: David Weston	Page 188
<b>Xiaohan Yang</b> : Plant-Microbe Interfaces: Small-RNA (sRNA) and open reading from (sORF) endo- and eco-mycorrhizal symbioses in Populous	
PRESENTER: Xiaohan Yang	Page 190
James Moran: Spatially Resolved Rhizosphere Function for Elucidating Key Controls on Below-ground Nutrient Interactions	
PRESENTER: James Moran	Page 192
Kirsten Hofmockel: Molecular interactions of the plant-soil-microbe continuum of bioenergy ecosystems	
PRESENTER: Kirsten Hofmockel	Page 194
Jay Keasling: Engineering Polyketide Synthases to Generate Lightly Branched Biofuels PRESENTER: Constance Bailey	Page 196
Christof Meile: Potential Mechanisms of Anaerobic Methane Oxidation and Challenges to Slow-Grow Microbial Communities	
PRESENTER: Christof Meile	Page 197
Victoria Orphan: Anabolic activity in Geobacter biofilms as a function of distance to inso acceptor	luble electron
PRESENTER: Grayson Chadwick	Page 199
William Cannon: Multi-scale Modeling of Circadian Rhythms: From Metabolism to Regul PRESENTER: William Cannon	lation and Back Page 200
Maria Ghirardi: The Algal Ferredixin Interactome PRESENTER: Venkataramanan Subramanian	Page 202
<b>George Church</b> : Design, Synthesis, and Testing Toward a 57-Codon Genome PRESENTER: George Church	Page 204
George Church: Exploring Species Specificity of Lambda Red Recombination PRESENTER: Gabriel Filsinger	Page 205

<b>George Church</b> : Non-Standard Amino Acid Incorporation in the New Era of Recoded Ger PRESENTER: Aditya Kunjapur	nomes Page 206
George Church: Retron Library Recombineering: Construction and Functional Interrogation of Trackable	
PRESENTER: Max Schubert	Page 208
<b>Costas Maranas</b> : 13C-assisted flux elucidation using genome-scale carbon mapping mod PRESENTER: Saratram Gopalakrishnan	lels Page 209
ostas Maranas: Application of the MetRxn database to highlight multi-tissue/organisms and expansion include algorithms for predicting novel reactions and pathways	
	Page 211
James Cole: RDP: Data and Tools for Microbial Community Analysis PRESENTER: Benli Chai	Page 213
<b>Sarah Evans</b> : Is there a direct link between free-living nitrogen fixation rates and nitroger rates?	en mineralization
PRESENTER: Lisa Tiemann	Page 215
Sarah Evans: The Effects of Growth Promoting Rhizobacteria and Endophytes on Switchgrass Growth and Root Architecture	
PRESENTER: Bana Abolibdeh	Page 217
Jack Brown: Developing Non-food Grade Brassica Biofuel Feedstock Cultivars with High Content, and Oil Quality that are Suitable for Low Input Production Dryland Systems. PRESENTER: Jack Brown	Yield, Oil Page 218
Jennifer Martiny: Analysis of an Abundant Bacterial Genus in a Leaf Litter Community PRESENTER: Alexander Chase	Page 219
Jizhong Zhou: Long-term priming-induced changes in permafrost soil organic matter dec PRESENTER: Edward Schuur	composition Page 221
F <b>rank Löffler</b> : Multiomic Insights into the Activity and Dynamics of Soil Nitrifier Communities in Midwestern Agricultural Soils	
PRESENTER: Luis Orellana	Page 223
<b>David Stahl</b> : Factors Affecting Nitrous Oxide Production from Ammonia Oxidizers and Po Mitigation Options	ossible
PRESENTER: David Stahl	Page 225
Luca Comai: A population of copy number variants for poplar functional genomics PRESENTER: Isabelle Henry	Page 227
John Dueber: Repurposing the Yeast Peroxisome for Compartmentalizing Multi-enzyme PRESENTER: Jennifer Samson	Pathways Page 229

ennifer Reed: Application of machine learning and active learning to enhance chemical yields in nicrobes	
PRESENTER: Prashant Kumar	Page 231
James McKinlay: Factors Governing Mutualism Dynamics in a Hydrogen-Producing Cocu PRESENTER: James McKinlay	lture Page 233
Garret Suen: Evaluating the Cellulolytic Properties of Novel Fibrobacteres isolates from I Herbivores	Mammalian
PRESENTER: Garret Suen	Page 235
<b>Lawrence Smart</b> : Differential Expression, Regulatory Divergence, and Sex Dimorphism Personal Shrub Willow (Salix spp.) Transcriptome	ervade the
PRESENTER: Craig Carlson	Page 236
Clint Chapple: Model-Guided Metabolic Engineering of Increased 2-Phenylethanol Produ PRESENTER: Shaunak Ray	uction in Plants Page 238
<b>Clint Chapple</b> : Measurement and modeling of phenylpropanoid metabolic flux in Arabid PRESENTER: Peng Wang	opsis Page 240
John Morgan: Kinetic modeling of the phenylpropanoid pathway in Arabidopsis PRESENTER: Longyun Guo	Page 242
Martin Polz: Identifying the most efficient algal degrading communities: diversity, composition and	
PRESENTER: Xiaoqian Yu	Page 244
Martin Polz: The Ecology of Macroalgal Polysaccharide Utilization: Verrucomicrobia Isolates Initiate Fucoidan and Carrageenan Degradation Cascades	
PRESENTER: Christopher Corzett	Page 246
<b>Christopher Rao</b> : Characterization of the Alginate Lyases and Laminarinases from Vibrio PRESENTER: Christopher Rao	sp. Page 247
Huimin Zhao: Optimization of Alginate Utilization in Engineered Bacteria for Biofuels Pro PRESENTER: Jonathan Ning	oduction Page 249
Scott Baker: Nitrogen limitation and lipid production in Yarrowia lipolytica PRESENTER: Scott Baker	Page 250
Jens Nielsen: Leucine Biosynthesis is Involved in Regulating High Lipid Accumulation in Y lipolytica	arrowia
PRESENTER: Eduard Kerkhoven	Page 251
Jens Nielsen: Lipid Accumulation and its Impact on Amino Acid Metabolism in Saccharor PRESENTER: Michael Gossing	nyces cerevisiae Page 252

<b>Gregory Stephanopoulos</b> : Optimal feed control for maximization of lipid production fro acid by an engineered oleaginous yeast Yarrowia lipolytica	m dilute acetic
PRESENTER: Jingyang Xu	Page 253
<b>Gregory Stephanopoulos</b> : rewiring metabolism for maximum lipid production in oleagir Yarrowia lipolytica	ious yeast
PRESENTER: Kangjian Qiao	Page 254
<b>Thomas Brutnell</b> : Seteria viridis as a Model System to Accelerate Gene Discovery in Pan PRESENTER: Christine Shyu	icoid Grasses Page 255
Thomas Brutnell: Grasses Suppress Shoot-Borne Roots to Conserve Water During Droug PRESENTER: Jose Dinneny	ght Page 257
<b>Thomas Brutnell</b> : Development and application of novel phenotyping techniques to und genetic control of productivity and drought traits in the model C4 grass Setaria	lerstand the
PRESENTER: Andrew Leakey	Page 259
<b>Eduardo Blumwald</b> : Biocontainment in Perennial Grasses: Transgenic Pollen Ablation in sylvaticum and Panicum virgatum.	Brachypodium
PRESENTER: Roger Thilmony	Page 260
Eduardo Blumwald: Genome Editing of CENH3 in Switchgrass and Brachypodium: A Histone Varia	
PRESENTER: Christian Tobias	Page 261
John Cushman: Leveraging Agave and Kalanchoë Genomics Resources to Transfer Crass Metabolism (CAM) Modules into C3 Species Using Synthetic Biology Approaches	ulacean Acid
PRESENTER: Xiaohan Yang	Page 263
John Cushman: Engineering Crassulacean Acid Metabolism (CAM) to Improve Water-use Bioenergy Feedstocks	e Efficiency of
PRESENTER: John Cushman	Page 265
Brian Pfleger: RNase III as a Tool to Manipulate Transcript Stability: Identifying in vivo Ta Escherichia coli	argets in
PRESENTER: Brian Pfleger	Page 269
Michelle O'Malley: Engineering Anaerobic Gut Fungi for Lignocellulose Breakdown PRESENTER: Michelle O'Malley	Page 271
<b>Heather Coleman</b> : Expression of a Hyperthermophilic Endoglucanase in Poplar to Impro	ve Ethanol
PRESENTER: Yao Xiao	Page 273
Heather Coleman: Extreme Inducible Expression of Cellulases in Poplar PRESENTER: Charleson Poovaiah	Page 274

Heath Poplar	eather Coleman: Modified Cell Wall Composition through Expression of an Expansin-Like Protein in	
. op.a.	PRESENTER: Heather Coleman	Page 276
Ronan	<b>Fleming</b> : Automatic generation of genome-scale metabolism and expression mod PRESENTER: Ronan Fleming	dels for bacteria Page 277
<b>Ronan</b> netwo	<b>Fleming</b> : CHRR: Coordinate hit-and-run with rounding for uniform sampling of m rks	etabolic
	PRESENTER: German Predicat	Page 278
Ronan	<b>Fleming</b> : Comparative evaluation of atom mapping algorithms for metabolic read PRESENTER: German Predicat	ctions Page 279
Ronan	<b>Fleming</b> : DistributedFBA.jl: High-level, high-performance flux balance analysis in PRESENTER: Sylvain Arreckx	Julia Page 280
Ronan	<b>Fleming</b> : Software development of The Constraint-Based Reconstruction and Ana PRESENTER: Sylvain Arreckx	alysis Toolbox Page 281
Ronan	Fleming: Thermodynamically constraining a genome-scale metabolic model with	von Bertalanffy
PRESENTER: Ronan Fleming	Page 282	
Ronan	<b>Fleming</b> : Cardinality optimisation in constraint-based modelling of metabolism PRESENTER: Ronan Fleming	Page 283
<b>Daniel</b> Comm	<b>Buckley</b> : High Resolution DNA Stable Isotope Probing of Soil Indicates Changes in unity Metabolism Associated in Disturbance Due to Tillage PRESENTER: Daniel Buckley	Microbial Page 284
<b>Daniel</b> probin	<b>Buckley</b> : Mapping microbial food web dynamics in soil with high resolution stable	e isotope
	PRESENTER: Sam Barnett	Page 286
Mary I Structi	<b>Firestone</b> : Bacterial Traits Linking Avena Exudate Chemistry and Rhizosphere Bacterian During Root Development	erial Community
	PRESENTER: Heejung Cho	Page 288
Virgini	a Rich: Linking Arctic Lake Sediment Microbial Ecology to Carbon Biogeochemistry PRESENTER: Joanne Emerson	y Page 290
Virgini	a Rich: The IsoGenieDB: An Integrated Solution to Cross-Disciplinary Data Manage PRESENTER: Benjamin Bolduc	ement Page 291
Matth	<b>ew Sullivan</b> : Soil viral ecology along a permafrost thaw gradient PRESENTER: Matthew Sullivan	Page 293

<b>Ray Ming</b> : Pyramiding genes and alleles for improving energy cane biomass yield PRESENTER: Ray Ming	Page 294
<b>Jizhong Zhou</b> : Time-series metagenomics of experimentally warmed Alaskan tundra and temperate soils enables fine-resolution assessment of belowground C cycling feedbacks change	Oklahoma to climate
PRESENTER: Konstantinos Konstantinidis	Page 296
<b>Jizhong Zhou</b> : Successional Dynamics of Grassland Microbial Communities in Response t Precipitation Alternation, and Clipping	o Warming,
PRESENTER: Liyou Wu	Page 299
<b>Costas Maranas</b> : Ensemble cell-wide kinetic modeling of anaerobic organisms to suppor chemicals production	t fuels and
PRESENTER: Satyakam Dash	Page 302
<b>Costas Maranas</b> : k-ecoli457: A genome-scale Escherichia coli kinetic metabolic model satisfying flux data for multiple mutant strains	
PRESENTER: Ali Khodayari	Page 304
James Liao: Robust non-Oxidative Glycolysis in Escherichia coli PRESENTER: Alec Jaeger	Page 306
<b>Chris Rao</b> : Protein Acylomes in Fuel-Producing E. coli – Changes in Posttranslational Modifications with Different Carbon Sources	
PRESENTER: Birgit Schilling	Page 307
James Liao: Predicting multiple gene targets for optimal oil production in acetic acid me Yarrowia lipolytica by ensemble modeling	tabolism of
PRESENTER: Po-Wei Chen	Page 309
<b>Christopher Rao</b> : Amino Acids Are Preferred Over Glucose and Other Sugars in Escherich	nia Coli by a
PRESENTER: James Orr	Page 310
Joshua Rabinowitz: Glycolysis Balances Enzyme Efficiency and Metabolic Adaptivity PRESENTER: Junyoung Park	Page 312
Joshua Rabinowitz: Systems-Level Analysis of Mechanisms Regulating Yeast Metabolic F PRESENTER: Joshua Rabinowitz	lux Page 313
Janelle Thompson: Systems biology towards a continuous platform for biofuels producti an environmentally-isolated Bacillus strain for biofuel production and recovery under su PRESENTER: Janelle Thompson	on: Engineering percritical CO2. Page 315

<b>Inelle Thompson</b> : Systems biology towards a continuous platform for biofuels production: eterologous gene expression and isobutanol synthesis in B. megaterium SR7 and biofuel extraction order supercritical CO2	
PRESENTER: Janelle Thompson	Page 317
Mark Hildebrand: A Systems Biology and Pond Culture-Based Understanding and Improv Metabolic Processes Related to Productivity in Diverse Microalgal Classes for Viable Biof PRESENTER: Juergen Polle	vement of uel Production. Page 319
Mark Hildebrand: The Effect of Carbon Flux Topology and Synchronized Culture Growth Productivity	on Microalgal
PRESENTER: Mark Hildebrand	Page 321
<b>Amy Brunner</b> : Spatiotemporal Transcriptomics of Populus Growth in Response to Dayler Nutrient Availability	ngth and
PRESENTER: Hua Bai	Page 323
<b>Daniel Segre</b> : Synthetic microbial ecology for biofuel production from lignocellulose PRESENTER: Meghan Thommes	Page 324
John Mullet: Genomics of Energy Sorghum's Water Use Efficiency/Drought Resilience PRESENTER: John Mullet	Page 326
Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Depend PRESENTER: Cristal Zuniga	dent Interactions Page 327
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Depend PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus stra bioconversion</li> </ul>	dent Interactions Page 327 ins for lignin
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Depend PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus stra bioconversion PRESENTER: Tayte Campbell</li> </ul>	dent Interactions Page 327 ins for lignin Page 329
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Dependence PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus strabioconversion         PRESENTER: Tayte Campbell     </li> <li>Jonathan Schilling: Spatial Connectomics to Identify Agents Relevant to Lignocellulose D         Europi     </li> </ul>	dent Interactions Page 327 ins for lignin Page 329 econstruction in
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Dependence PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus strabioconversion PRESENTER: Tayte Campbell</li> <li>Jonathan Schilling: Spatial Connectomics to Identify Agents Relevant to Lignocellulose D Fungi</li> <li>PRESENTER: Jonathan Schilling</li> </ul>	dent Interactions Page 327 ins for lignin Page 329 econstruction in Page 330
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Dependent PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus stration bioconversion         <ul> <li>PRESENTER: Tayte Campbell</li> </ul> </li> <li>Jonathan Schilling: Spatial Connectomics to Identify Agents Relevant to Lignocellulose D Fungi         <ul> <li>PRESENTER: Jonathan Schilling</li> </ul> </li> <li>Jonathan Schilling: Distinct growth patterns and time-dependent secretome alterations taxonomically divergent brown rot fungi</li> </ul>	dent Interactions Page 327 ins for lignin Page 329 econstruction in Page 330 by two
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Dependent PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus strabioconversion PRESENTER: Tayte Campbell</li> <li>Jonathan Schilling: Spatial Connectomics to Identify Agents Relevant to Lignocellulose D Fungi PRESENTER: Jonathan Schilling</li> <li>Jonathan Schilling: Distinct growth patterns and time-dependent secretome alterations taxonomically divergent brown rot fungi PRESENTER: Gerald Presley</li> </ul>	dent Interactions Page 327 ins for lignin Page 329 econstruction in Page 330 by two Page 332
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Depend PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus stra bioconversion PRESENTER: Tayte Campbell</li> <li>Jonathan Schilling: Spatial Connectomics to Identify Agents Relevant to Lignocellulose D Fungi PRESENTER: Jonathan Schilling</li> <li>Jonathan Schilling: Distinct growth patterns and time-dependent secretome alterations taxonomically divergent brown rot fungi PRESENTER: Gerald Presley</li> <li>Daniel Schachtman: Systems Analysis of the Physiological and Molecular Mechanisms of Nitrogen Use Efficiency. Water Use Efficiency and Interactions with the Soil Microbiome</li> </ul>	dent Interactions Page 327 ins for lignin Page 329 econstruction in Page 330 by two Page 332 <sup>5</sup> Sorghum
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Depend PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus stra bioconversion PRESENTER: Tayte Campbell</li> <li>Jonathan Schilling: Spatial Connectomics to Identify Agents Relevant to Lignocellulose D Fungi PRESENTER: Jonathan Schilling</li> <li>Jonathan Schilling: Distinct growth patterns and time-dependent secretome alterations taxonomically divergent brown rot fungi PRESENTER: Gerald Presley</li> <li>Daniel Schachtman: Systems Analysis of the Physiological and Molecular Mechanisms of Nitrogen Use Efficiency, Water Use Efficiency and Interactions with the Soil Microbiome PRESENTER: Daniel Schachtman</li> </ul>	dent Interactions Page 327 ins for lignin Page 329 econstruction in Page 330 by two Page 332 Sorghum Page 334
<ul> <li>Michael Betenbaugh: Constraint-Based Community Modeling Reveals Condition-Depend PRESENTER: Cristal Zuniga</li> <li>Gautam Dantas: Aromatic tolerance and utilization in adapted Rhodococcus opacus strabioconversion PRESENTER: Tayte Campbell</li> <li>Jonathan Schilling: Spatial Connectomics to Identify Agents Relevant to Lignocellulose D Fungi PRESENTER: Jonathan Schilling</li> <li>Jonathan Schilling: Distinct growth patterns and time-dependent secretome alterations taxonomically divergent brown rot fungi PRESENTER: Gerald Presley</li> <li>Daniel Schachtman: Systems Analysis of the Physiological and Molecular Mechanisms of Nitrogen Use Efficiency, Water Use Efficiency and Interactions with the Soil Microbiome PRESENTER: Daniel Schachtman: Harvesting the Root Microbiome of Grasses toward Sustainable Incor Production</li> </ul>	dent Interactions Page 327 ins for lignin Page 329 econstruction in Page 330 by two Page 332 Sorghum Page 334 reases in Crop

<b>Christopher Rao</b> : Increasing growth yield and decreasing acetylation in Escherichia coli b optimizing the carbon-to-magnesium ratio in pentide-based media	stopher Rao: Increasing growth yield and decreasing acetylation in Escherichia coli by	
PRESENTER: David Christensen	Page 338	
<b>Himadri Pakrasi</b> : Systems Level Study of a Novel Fast-Growing Cyanobacterial Strain for Biofuel Production	Next Generation	
PRESENTER: Justin Ungerer	Page 340	
<b>Basil Nikolau</b> : Integrated and Dynamic Multispectroscopic In Situ Imaging of Plant Metal Level of Subcellular Compartments	bolism at the	
PRESENTER: Geng Ding	Page 342	
<b>Mary Firestone</b> : Monitoring fluxes of atmospherically-reactive gases (CO2, CH4 and N2C conversion of grasslands into a biofuel crop (Panicum virgatum)	)) during the	
PRESENTER: Jizhong Zhou	Page 344	
<b>Mary Firestone</b> : Plant-microbe and microbe-microbe interactions mediate switchgrass sustainability: following rhizosphere microbial communities during switchgrass establishment		
PRESENTER: Kelly Craven	Page 346	
<b>Mary Firestone</b> : Dissecting the Chemistry of Switchgrass-Microbe Interactions Using Cull Exometabolomics and Mass Spectrometry Imaging	tivation,	
PRESENTER: Kateryna Zhalnina	Page 348	
Mary Firestone: Faunal Population Dynamics Throughout Switchgrass Developmental St PRESENTER: Javier Ceja-Navarro	ages Page 350	
Peggy Lemaux: EPICON: Epigenetic Control of Drought Response in Sorghum PRESENTER: Peggy Lemaux	Page 353	
<b>Sarah Evans</b> : Switchgrass Growth and Transcriptomic Responses to Nitrogen Availability	and the	
PRESENTER: Alan Bowsher	Page 355	
<b>Sarah Evans</b> : The Contribution of Alternative Nitrogenases to Nitrogen Fixation in Switch	ngrass	
PRESENTER: Jeffrey Norman	Page 356	
Peter Balint-Kurti: Physiological and Molecular-Genetic Characterization of Basal Resista PRESENTER: Peter Balint-Kurti	ance in Sorghum Page 358	
<b>Thomas Juenger</b> : Climate adaption and sustainability in switchgrass: exploring plant-mic interactions across continental scale environmental gradients	robe-soil	
PRESENTER: Thomas Juenger	Page 360	
Hugo Cuevas: Genomic dissection of anthracnose resistant response in sorghum [Sorghu PRESENTER: Hugo Cuevas	um bicolor (L.)] Page 362	

Jeffrey Gardner: Systems analysis in Cellvibrio japonicus resolves predicted redundancy glucosidases and determines essential physiological functions	expected and the second secon	
PRESENTER: Cassandra Nelson	Page 364	
<b>Gloria Coruzzi</b> : EvoNet: A Phylogenomic and Systems Biology approach to identify genes	underlying	
PRESENTER: Kranthi Varala	Page 365	
<b>Kabir Peay</b> : Does mycorrhizal symbiosis determine the climate niche for Populus as a bio feedstock?	benergy	
PRESENTER: Kabir Peay	Page 367	
Serge Edme: Genetics and Genomics of Pathogen Resistance in Switchgrass PRESENTER: Serge Edme	Page 369	
<b>Bruce Hungate</b> : Quantitative stable isotope probing with 15N in soil microbial communi PRESENTER: Ember Morrissey	ties Page 371	
Bruce Hungate: Studying Microbial Stress Responses in Soil Ecosystems PRESENTER: Paul Dijkstra	Page 373	
Mary Lidstrom: Communal Metabolism of Methane and the Rare Earth Element Switch PRESENTER: Ludmila Chistoserdova	Page 375	
<b>Mary Firestone</b> : Directing traffic in the rhizosphere: how phage and fauna shape the flow root carbon through microbial pathways	w and fate of	
PRESENTER: Katerina Estera	Page 377	
Mary Firestone: Mapping the pathways of root carbon flow into and through soil microl PRESENTER: Evan Starr	pial food webs Page 379	
Mary Firestone: Roots stimulate expression of decomposition transcripts in the soil mice PRESENTER: Erin Nuccio	robiome Page 381	
Thomas D Bruns: A Simple Pyrocosm for Controlled, Replicated Studies of Post-Fire Soil Microbial		
PRESENTER: Akiko Carver	Page 383	
<b>Thomas Bruns</b> : Determination of the Roles of Pyrophilous Microbes in the Breakdown a Sequestration of Pyrolyzed Forms of SOM	nd	
PRESENTER: Thea Whitman	Page 385	
<b>Chaofu Lu</b> : Systems Biology to Improve Camelina Seed and Oil Quality Traits PRESENTER: Chaofu Lu	Page 387	
<b>Steven Allison</b> : Trait-based approaches for linking metagenomic data with microbial carbon cycling under drought conditions		
PRESENTER: Steven Allison	Page 389	

oria Orphan: Comparative genomics and functional characterization of assimilatory sulfate	
PRESENTER: Hang Yu	Page 391
Ana Paula Alonso: Development of Resources and Tools to Improve Oil Content and Qua	ality in
PRESENTER: Ana Paula Alonso	Page 392
<b>Kristen DeAngelis</b> : Get Your Model Out of the Clouds! Ground-Truthing Assumptions Ab Earth's Tiniest Engines Drive the Carbon Cycle	out how the
PRESENTER: Grace Pold	Page 394
Serita Frey: Fungal Responses To Elevated Temperature And Soil Nitrogen Availability PRESENTER: Kristen DeAngelis	Page 395
<b>Serita Frey</b> : Microbial Growth and Metabolism in Soil – Refining the Interpretation of Ca Efficiency	rbon Use
PRESENTER: Kristen DeAngelis	Page 396
Jamey Young: Mapping photoautotrophic metabolism of engineered cyanobacteria to identify reactions	
PRESENTER: Yi Ern Cheah	Page 397
Adam Arkin: Pooled Assembly, Genotyping and Scoring of Complex Synthetic Genomic L PRESENTER: Robert Egbert	ibraries Page 399
<b>Ryan Gill</b> : A novel design strategy for industrially relevant, unnatural modular megasynt PRESENTER: William Grau	hases Page 401
<b>Ryan T. Gill</b> : Combinatorial engineering of 3-hydroxypropionate production from hemice hydrolysate	ellulose
PRESENTER: Rongming Liu	Page 403
<b>Ryan Gill</b> : Developing a predictive method for tunable control over gene expression base interference technology	ed on CRISPR
PRESENTER: Katia Tarasava	Page 405
Ryan Gill: Enhanced CRISPR-based trackable protein engineering using modeling PRESENTER: Eun Joong Oh	Page 407
Ryan Gill: Extending functional genome annotations using high-throughput CRISPRi PRESENTER: Harneet Rishi	Page 409
<b>Ryan Gill</b> : Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli PRESENTER: Marcelo Bassalo	Page 411

<b>Ryan G</b> coli	II: Genome-scale design and engineering approach towards optimizing ethylene production in E.	
con	PRESENTER: Aparna Nagarajan	Page 413
Samuel	Payne: Enabling Metaproteomics Research	
	PRESENTER: Samuel Payne	Page 415
Brian D	avison: Combining Deuterium-Labeling and Neutron Scattering to GAin Molecula	r-Level Insights
neievai	PRESENTER: Hugh O'Neill	Page 416
Brian D Charact	<b>avison</b> : Conformation of Low-Molecular Weight Lignin in Water: Insights on the G ter of Lignin	Chemical
	PRESENTER: Loukas Petridis	Page 418
Brian D	vavison: Deuteration Effects on Switchgrass Structure and Mechanism: Lignin Dep Malls of Deuterated Switchgrass	osition Changes
	PRESENTER: Samarthya Bhagia	Page 420
<b>Brian D</b> Dilute A	avison: Real-Time Elucidation of Structure and Morphology of Native and Mutan Acid and Alkali Pretreatments Using Neutron Scattering	t Poplar During
	PRESENTER: Sai Venkatesh Pingali	Page 422
Welling symbio	<b>ston Muchero</b> : Species-specific evolution of membrane–bound receptors mediati nt specificity in the genus Salix	ng host-
·	PRESENTER: Wellington Muchero	Page 424
<b>Lee An</b> Produce	<b>n McCue</b> : Chemical Analysis of Carbon and Nitrogen Cycling Through the Extracel ed During the Formation of a Multi-Species Community	lular Matrix
	PRESENTER: Matt Marshall	Page 425
<b>Lee Ani</b> Partitio	n McCue: High-Resolution Spatial Analysis Reveals How Nitrogen Source Governs ning Between Members in a Phototrophic Consortium	Carbon
	PRESENTER: Christopher Anderton	Page 427
Lee Ani	n McCue: Metabolite Characterization in Complex Microbial Communities PRESENTER: Ryan Renslow	Page 429
Lee Anı	<b>n McCue</b> : Modeling Approaches for Understanding Metabolic Coupling in Microb PRESENTER: Hyun Song	ial Communities Page 431
<b>Lee An</b> Metabo	n McCue: Newly Identified Regulatory Roles for Vitamin B12 Suggest Coordination blism	n of Community
	PRESENTER: Lee Ann McCue	Page 433
Philipp	e Noirot: Multi-Omics Analysis of a Mycorrhizal System	Dage //35
		- 466 - 33

Philippe Noirot: Functional assignment of ligand specificities for Pseudomonas transpor PRESENTER: Peter Korajczyk	t proteins Page 437	
<b>Philippe Noirot</b> : Strain-Specific Transportomic Capacity of Pseudomonas fluorescens Lin Growth Promotion in Aspen Seedlings under Nutrient Stress	ked to Plant	
PRESENTER: Shalaka Shinde-Desai	Page 439	
Philippe Noirot: Genetic Systems to Enable Biosystems Design in Rhizosphere Pseudomo	onas	
PRESENTER: Philippe Noirot	Page 441	
<b>Eoin Brodie</b> : A Trait Based Dynamic Energy Budget Approach to Explore Emergent Micro Dynamics and Productivity	algal-Bacterial	
PRESENTER: Yiwei Cheng	Page 442	
Xavier Mayali: Isotope tracing and phylogenetic composition of simplified bacterial communities conferring growth and biomass enhancements to biofuel-producing microalgae		
PRESENTER: Jeffrey Kimbrel	Page 444	
Ali Navid: System-level analysis of metabolic trade-offs and changes during diurnal cycle Chlamydomonas reinhardtii	of	
PRESENTER: Ali Navid	Page 446	
Jennifer Pett-Ridge: Hyper-accumulated Mn, co-localized in Chlamydomonas reinhardtii acidocalcisomes with Ca and P, can be mobilized in Mn-deficient situations and protects	against	
PRESENTER: Peter Weber	Page 448	
<b>Rhona Stuart</b> : Exploring the functional aspects of microalgal interactions with associated predators and pathogens	d microbes,	
PRESENTER: Rhona Stuart	Page 450	
Peter Weber: Quantifying the contribution of viruses and phage to nutrient cycling with	NanoSIMS	
PRESENTER: Peter Weber	Page 452	
Jennifer Pett-Ridge: How Redox Fluctuation Shapes Microbial Community Structure and Organic Matter Relationships in a Humid Tropical Forest Soil	Mineral-	
PRESENTER: Ashley Campbell	Page 454	
<b>Karsten Zengler</b> : Omics-Driven Analysis of Clostridium ljundhalii Disentangles the Compl Conservation in Autotrophic and Heterotrophic Growth Conditions	exity of Energy	
PRESENTER: Mahmoud Al-Bassam	Page 456	
Ashley Shade: A New Synthetic Community System for Studying Microbial Interactions E	Driven by	
PRESENTER: John Chodkowski	Page 458	

<b>Karsten Zengler</b> : Using a Metabolic and Gene-Expression Model to Predict and Analyze t Response of Acetogen Clostridium liungdahlii	he Phenotypic	
PRESENTER: Joanne Liu	Page 460	
<b>Deanna Funnell-Harris</b> : Resistance to Stalk Pathogens for Bioenergy Sorghum PRESENTER: Deanna Funnell-Harris	Page 462	
<b>Yongqin Jiao</b> : Identification of a U/Zn/Cu responsive global regulatory two-component s	ystem in	
PRESENTER: Yongqin Jiao	Page 464	
<b>C. Robin Buell</b> : Exploiting Natural Diversity to Identify Alleles and Mechanisms of Cold Ad Switchgrass	daptation in	
PRESENTER: C Robin Buell	Page 465	
<b>Maria Harrison</b> : Genetic dissection of AM symbiosis to improve the sustainability of feed production	dstock	
PRESENTER: Maria Harrison	Page 466	
Patrick Brown: A Sorghum NAC Gene Affects Vascular Development and Biomass Proper PRESENTER: Patrick Brown	rties Page 467	
<b>Daniel Schachtman</b> : Metabolomics as a Molecular Tool to Understand the Mechanism of Nitrogen Use Efficiency and Interactions with the Soil Microbiome in Energy Sorghum.		
PRESENTER: Jessica Prenni	Page 468	
Gary Siuzdak: Systems Biology Guided by Global Isotope Metabolomics PRESENTER: Erica Forsberg	Page 469	
<b>M. David Marks</b> : Advancing Field Pennycress as a New Oilseed Biofuels Feedstock that does not Require New Land Commitments		
PRESENTER: John Sedbrook	Page 471	
Jillian Banfield: Microbial community proteogenomic analyses indicate extensive depth- oxidation and C1 metabolism in soil and increased capacity for N2O reduction with incre PRESENTER: Spencer Diamond	dependent CO eased rainfall Page 473	
Sabeeha Merchant: An O-Glycosylated Archaeal Flagellin PRESENTER: Rachel Ogorzalek Loo	Page 474	
John McKay: Germplasm Development for Sustainable Production of Camelina sativa Oi PRESENTER: John McKay	l Page 475	
<b>Dominique Loque</b> : Expression of S-adenosylmethionine Hydrolase in Tissues Synthesizing Secondary Cell Walls Alters Specific Methylated Cell Wall Fractions and Improves Biomass Digestibility		
PRESENTER: Aymerick Eudes	Page 477	

<b>Dominique Loque</b> : Transcript switches for fine-tuning of transgene expression PRESENTER: Yan Liang	Page 479
George Newcombe: Characterizing the Defense Hierarchy of Populus trichocarpa PRESENTER: George Newcombe	Page 480

### Limited thermodynamic driving force in glycolysis of cellulolytic clostridia

Tyler Jacobson,<sup>1,2</sup> David M. Stevenson,<sup>1,2</sup> Daniel Olson,<sup>2,3</sup> Lee R. Lynd,<sup>2,3</sup> and **Daniel Amador-Noguez**<sup>1,2</sup>\* (amadornoguez@wisc.edu)</sup>

<sup>1</sup>University of Wisconsin-Madison, Madison; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>Dartmouth College, Hanover, New Hampshire

**Project Goals:** The BioEnergy Science Center (BESC) is focused on the fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) designing plant cell walls for rapid deconstruction and (2) developing multi-talented microbes or converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. BESC research in biomass deconstruction and conversion targets CBP by studying thermophilic anaerobes to understand novel strategies and enzyme complexes for biomass deconstruction and manipulating these microorganisms for improved conversion, yields, and biofuel titer. BESC researchers provide enabling technologies in biomass characterization, 'omics, modeling and data management in order to (1) understand chemical and structural changes within biomass and (2) to provide insights into biomass formation and conversion mechanisms.

*C. thermocellum* and *C. cellulolyticum* are obligate anaerobes capable of converting cellulose into ethanol. The glycolytic pathways of these two microorganisms display unique cofactor utilization, resulting in an energy-efficient sugar catabolism that is thought to generate more usable energy in the form of high-energy phosphate bonds than canonical glycolytic pathways but at the cost of forward thermodynamic driving force.

Here, we have used a combination of <sup>13</sup>C-labeling and <sup>2</sup>H-labeling to measure absolute metabolite concentrations and fluxes in *C. thermocellum* and *C. cellulolyticum* and experimentally determine changes in free energy ( $\Delta G$ ) at each step in their glycolytic and ethanol fermentation pathways. Our experimental method relies on the fundamental principle that for any reaction,  $\Delta G$  is log proportional both to a concentration ratio (reaction quotient to equilibrium constant) and to a flux ratio (backward to forward flux), which can be reliably estimated from steady-state labeling data.

We found that the glycolytic and fermentation pathways in these two cellulolytic clostridia are surprisingly fully reversible under normal growth conditions. The overall thermodynamic driving force in the glycolytic pathways of *C. thermocellum* and *C. cellulolyticum* is significantly limited compared to canonical glycolytic pathways in model organisms such as *E. coli* or *S. cerevisiae* or non-cellulolytic thermophilic bacteria such as *T. saccharolyticum*. We also found that forward driving force is dependent on environmental inputs; for example, as ethanol accumulates during fermentation the thermodynamic driving force in glycolysis of *C. thermocellum* becomes even more limited. The limited forward driving force that we observe in the glycolytic and fermentation pathways of cellulolytic clostridia may constitute an evolutionary adaptation to growth in cellulose and the need to produce as much usable energy as possible per glucose.

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

### High-quality protein production for structural studies: From plant cell wall synthesis to microbial production of bioproducts.

Giovani P. Tomaleri<sup>1\*</sup> (gptomaleri@lbl.gov), Ditte Welner<sup>1</sup>, Andy DeGiovanni<sup>1</sup>, Ryan McAndrew<sup>1</sup>, Robert W. Haushalter<sup>3</sup>, Henrique Pereira<sup>1</sup>, Henrik V. Scheller<sup>2</sup>, Jay Keasling<sup>3</sup> and **Paul D. Adams**<sup>1</sup>.

<sup>1</sup>Technology, <sup>2</sup>Feedstocks and <sup>3</sup>Fuels Synthesis Divisions, DOE Joint BioEnergy Institute, Emeryville, CA;

Project Goals: JBEI's mission is to provide the scientific basis for converting lignocellulosic biomass to renewable, drop-in, liquid transportation fuels and other important bioproducts. In order to achieve this, JBEI researchers use high high-resolution X-ray crystallography and cryo-electron microscopy methods to understand the atomic structures of proteins involved in producing advanced biofuels and bioproducts by providing the foundational understanding of enzyme mechanism and structure/function relationships as well as the ability to engineer function and stability. An important step to obtaining high-resolution structural information is sample preparation, which requires the expression and purification of high-quality protein that typically includes a series of screening steps followed by up scaling and optimization. The workflow will be described for enzymes involved in the synthesis of the plant cell wall, and enzymes used for the microbial production of bioproducts.

Glycosyltransferases (GTs) are enzymes involved in the biosynthesis of plant cell-wall polysaccharides. The cell walls are complex structures that play a key role in plant fitness. Furthermore, these cell walls have been proposed to be a source of renewable energy in the form of biofuels. GTs catalyze the connection of simple monosaccharide sugars into complex polysaccharide sugars, collectively known as hemicelluloses. Due to their essential function in plant cell wall biosynthesis we are interested in their structural characterization and mechanisms of catalysis.

Type I Fatty acid synthases (FAS1) are megadalton-sized enzymes, which contain multiple catalytic domains and function like molecular assembly lines. Corynebacteria utilize the multifunctional type I FAS for the *de novo* biosynthesis of fatty acids to generate, for example, palmitic acid (C16:0), oleic acid (C18:1) and stearic acid (C16:0 or C18:0), which are incorporated into the cell-membrane phospholipids or used as precursors for mycolic acid biosynthesis. Fatty acids are interesting candidates for future biofuels, because they are highly reduced and have high energy densities. Due to the complex interactions of multiple catalytic domains, the engineering of FAS1 for biofuel production will benefit from detailed structural analysis.

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.

### Discovery of a Novel Bacterial Enzyme Enabling First-Time Biochemical Production of Toluene

**Harry R. Beller**<sup>1</sup>\* (HRBeller@lbl.gov), Andria Rodrigues,<sup>1</sup> Kamrun Zargar,<sup>1</sup> Avneesh Saini,<sup>1</sup> Susannah G. Tringe,<sup>2</sup> Jay D. Keasling,<sup>1</sup> and Christopher J. Petzold<sup>1</sup>

<sup>1</sup>Joint BioEnergy Institute (JBEI), Emeryville, CA; <sup>2</sup>Joint Genome Institute (JGI), Walnut Creek, CA

# Project Goals: The Joint BioEnergy Institute (JBEI) aims to produce a chemically diverse suite of biofuels from lignocellulosic biomass. Our objective in this project was enzyme discovery to enable first-time biochemical production of toluene, an important octane booster and petrochemical with a global annual market of 29 million tons, from cellulosic sugars.

Although anaerobic bacterial biosynthesis of toluene from phenylacetic acid was reported more than two decades ago, the biochemistry underlying this novel metabolism has never been elucidated. Here we report the discovery of a toluene synthase (phenylacetate decarboxylase) from an anaerobic, sewage-derived enrichment culture that quantitatively produces toluene from phenylacetate. The discovery process (Zargar et al. 2016) included metagenome sequencing of the culture (which included more than 340,000 protein-coding genes), anaerobic FPLC (Fast Protein Liquid Chromatography) of cell-free extracts of the culture, and differential metaproteomic analyses to identify proteins present in active (toluene-producing) FPLC fractions but absent in adjacent inactive FPLC fractions (i.e., toluene synthase candidates). Toluene synthase candidates included a novel glycyl radical enzyme (GRE) and its cognate activating enzyme [AE; a radical SAM (S-adenosyl-L-methionine) enzyme]. Recombinant, N-terminally tagged, codon-optimized versions of the GRE and AE genes were expressed in E. coli and purified under anaerobic conditions. After in vitro reconstitution of the AE to restore its [4Fe-4S] cluster, its activity was confirmed in vitro by measuring conversion of SAM to methionine. In vitro assays with the purified GRE, AE, and SAM were shown to successfully convert <sup>13</sup>Clabeled phenylacetate to <sup>13</sup>C-labeled toluene, whereas no toluene was produced in control assays lacking SAM. Thus, using an omics-enabled approach, we have discovered a novel glycyl radical enzyme (only 6 are currently known) that catalyzes decarboxylation of phenylacetate to form toluene. Heterologous expression of the GRE and AE in a phenylacetate-overproducing microbial host should enable toluene production from lignocellulosic biomass.

### **Publications**

1. Zargar, K., R. Saville, R. Phelan, S. G. Tringe, C. J. Petzold, J. D. Keasling, and H. R. Beller. 2016. *In vitro* characterization of phenylacetate decarboxylase, a novel enzyme catalyzing toluene biosynthesis in an anaerobic microbial community. *Scientific Reports (Nature)* **6**, 31362 doi: 10.1038/srep31362.

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

### j5 Software Through the Years: Insights from Aggregate Public Usage Metrics

Nathan J. Hillson<sup>1,2,3,4,5</sup>\* (njhillson@lbl.gov)

<sup>1</sup>Fuels Synthesis and <sup>2</sup>Technology Divisions, DOE Joint BioEnergy Institute, Emeryville, CA; <sup>3</sup>DNA Synthesis Science Program, DOE Joint Genome Institute, Walnut Creek, CA; <sup>4</sup>DOE Agile BioFoundry, Emeryville, CA; <sup>5</sup>Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA.

### https://j5.jbei.org

### **Project Goals:**

The publicly-accessible j5 (DNA assembly design automation) webserver has been active since August 2010 (four and a half years ago). Since then, nearly two thousand academic, non-profit, and government researchers across almost five hundred institutions world-wide have become registered users of j5. While the analysis presented here only covers public j5 webserver usage (excludes, for example, j5 use through the JBEI/JGI/Agile BioFoundry DIVA webservers or TeselaGen's commercial bioCAD/CAM platform), it provides key insights into DNA assembly design trends, and speaks to the significant contributions that DOE-supported software can broadly make to academic and government research efforts.

#### References

- Hillson, N.J.\*, Rosengarten, R.D., and Keasling, J.D. (2012) j5 DNA Assembly Design Automation Software. ACS Synthetic Biology 1 (1), 14-21.
- Hillson, N.J. (2014) j5 DNA Assembly Design Automation. DNA Cloning and Assembly Methods, Valla S., and Lale, R. (Editors), Humana Press, 245-269.

This work was part of the DOE Joint BioEnergy Institute (http:// www.jbei.org) and part of the DOE Joint Genome Institute (http://jgi.doe.gov) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, and was part of the Agile BioFoundry (http://agilebiofoundry.org) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-ACO2-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

### Title: Isopentenyl diphosphate (IPP)-bypass mevalonate pathways for C5 alcohol production

Aram Kang<sup>1,2</sup>, Corey W. Meadows<sup>1,2</sup>, Nicolas Canu<sup>1</sup>, Jay D. Keasling<sup>1,2,3,4</sup>, **Taek Soon Lee**<sup>1,2\*</sup> (<u>tslee@lbl.gov</u>)

<sup>1</sup>Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA 94608, USA; <sup>2</sup>Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; <sup>3</sup>Department of Bioengineering, University of California, Berkeley, CA 94720, USA; <sup>4</sup>Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, USA

https://www.jbei.org/research/divisions/fuels-synthesis/metabolic-engineering/

https://www.jbei.org/jbei-invention-leads-to-more-efficient-biofuel-production-for-industrial-application/

Project Goals: The Joint BioEnergy Institute (JBEI) aims to produce a chemically diverse suite of biofuels from lignocellulosic biomass, and isoprenoid-based biofuels have been of great interest due to their superb fuel properties. Mevalonate (MVA) pathway is one of the major biosynthetic pathways of isoprenoid fuel production. Various engineering strategies and tools for this pathway have been explored to identify the bottlenecks of the pathway and to achieve higher production of these biofuels. The toxicity of key intermediates and the intrinsic energy demands of this pathway, which result a significant operational cost for aeration, have been two important problems when a production in large scale is exploited. In this work, we present modified version of the MVA pathway that will address these issues for isoprenoid biofuel production, and our recent efforts to improve isopentenol production using this modified pathway.

Isopentenol (3-methyl-3-butenol) is a promising biofuel with favorable combustion properties and a precursor molecule for the production of isoprene (1). A mevalonate (MVA)-based isoprenoid biosynthetic pathway for C<sub>5</sub> alcohols was constructed in *E. coli* using genes from several organisms, and the pathway was optimized to achieve over 50% theoretical yield (2, 3). The MVA and MEP pathways intersect at isopentenyl diphosphate (IPP), the direct precursor to isoprenoid-derived C<sub>5</sub> alcohols and initial precursor to longer chain terpenes, which makes independent regulation of the pathways difficult.

In pursuit of the "decoupling" of the MVA pathway from native cellular regulation, we designed novel IPP-bypass MVA pathways for C<sub>5</sub> alcohol production by utilizing promiscuous activities of two enzymes, phosphomevalonate decarboxylase (PMD) and an *E. coli*-endogenous phosphatase (4).

The IPP-bypass pathways have reduced energetic requirements, are further decoupled from intrinsic regulation, and are free from IPP-related toxicity. In addition to these benefits, we demonstrate that reduced aeration rate has less impact on this bypass pathway than the original MVA pathway. Finally, we showed that performance of the bypass pathway was primarily determined by the activity of PMD. We designed a growth-linked screening platform to select PMD mutants with improved activity and demonstrated titer increases in the mutant strains. This modified pathway would be a good platform for industrial production of isopentenol and isoprene.



#### References

- 1. Mack, J.H., Rapp, V.H., Broeckelmann, M., Lee, T.S., Dibble, R.W. (2014). Fuel, 117, 939-943
- George, K.W., Chen, A., Jain, A., Batth, T.S., Baidoo, E., Wang, G., Adams, P.D., Petzold, C.J., Keasling, J.D., Lee, T.S. (2014), *Biotech. Bioeng*. 111, 1648-1658
- George, K.W., Thompson, M.G., Kang, A., Baidoo, E., Wang, G., Chan, L.J.G., Petzold, C.J., Adams, P.D., Keasling, J.D., Lee, T.S. (2015), *Sci. Rep.* 5, 11128
- 4. Kang, A., George, K.W., Wang, W., Baidoo, E., Keasing, J.D., Lee, T.S. (2016), Metabol. Eng. 34, 25-35

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

### *Rhodosporidium toruloides* for conversion of depolymerized cellulose, hemicellulose, and lignin into bioproducts

Junko Yaegashi<sup>1,2\*</sup> (JYaegashi@lbl.gov), James Kirby<sup>1,3</sup>, Masakazu Ito<sup>4</sup>, Jian Sun<sup>1,5</sup>, Tanmoy Dutta<sup>1,5</sup>, Mona Mirsiaghi<sup>6</sup>, Eric R. Sundstrom<sup>6</sup>, Edward Baidoo<sup>1,6</sup>, Deepti Tanjore<sup>6</sup>, Todd Pray<sup>6</sup>, Kenneth Sale<sup>1,5</sup>, Seema Singh<sup>1,5</sup>, Jay D. Keasling<sup>1,3,6,7,8</sup>, Blake A. Simmons<sup>1,6</sup>, Steven W. Singer<sup>1,6</sup>, Adam P. Arkin<sup>4,6,8</sup>, Jeffrey M. Skerker<sup>4,6</sup>, **Jon K. Magnuson<sup>1,2</sup>**, **John M. Gladden<sup>1,5</sup>** 

<sup>1</sup> Joint BioEnergy Institute, Emeryville, CA, USA; <sup>2</sup> Pacific Northwest National Laboratory, WA, USA; <sup>3</sup> California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, CA, USA; <sup>4</sup> Energy Bioscience Institute, Berkeley, CA, USA; <sup>5</sup> Sandia National Laboratory, Livermore, CA, USA; <sup>6</sup> Lawrence Berkeley National Laboratory, Berkeley, CA, USA; <sup>7</sup> Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA, USA; <sup>8</sup> Department of Bioengineering, University of California, Berkeley, CA, USA.

### https://www.jbei.org/research/divisions/deconstruction/fungal-biotechnology/

Project Goals: Conversion of all major components of lignocellulose into advanced bioproducts has been a long elusive goal that is essential for enabling a robust bioeconomy. This project, performed in the Fungal Biotechnology group in the Deconstruction Division at JBEI, seeks to establish and utilize *Rhodosporidium toruloides* as a platform organism that can convert depolymerization products of cellulose, hemicellulose and lignin into biofuels and bioproducts.

Economical conversion of lignocellulosic biomass to biofuels and bioproducts is central to the establishment of a robust bioeconomy. Efficient conversion of lignocellulose requires the utilization of all its primary components (cellulose, hemicellulose, and lignin), yet no microbe in commercial use today can achieve this feat. To that end, we explored the utility of *Rhodosporidium toruloides* as a new platform organism for the production of biofuels and bioproducts from deconstructed plant biomass. In this study, Rhodosporidium toruloides was engineered to produce two non-native terpenes with biofuel (bisabolene) and pharmaceutical (amorphadiene) applications from a mixture of depolymerized cellulose (glucose), hemicellulose (xylose), and lignin (p-coumaric acid). The titers of these bioproducts in synthetic defined media with 2% glucose were 294mg/L and 36g/L, respectively. It was also cultivated on corn stover hydrolysates prepared by two different pretreatment methods, including one using the novel biocompatible ionic liquid choline  $\alpha$ -ketoglutarate. This organism was able to tolerate the ionic liquid that was used for pretreatment and was also able to utilize all of the three major carbon sources in the hydrolysate to support growth and bisabolene production. Finally, we demonstrate that this organism is amenable to high-gravity fed-batch fermentation, reaching a titer of 680mg/L in alkaline corn stover hydrolysate. This study establishes *R. toruloides* as a new platform for the simultaneous conversion of depolymerized cellulose, hemicellulose, and lignin into biofuels and bioproducts.

This work was performed as part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

### Characterization of Plant Golgi-Localized Nucleotide Sugar Transporters

Julien Sechet<sup>1,2</sup>, Beibei Jing<sup>1,2</sup>, Soe Htwe<sup>1,2</sup>, Fekadu Anderbehan<sup>1,2</sup>, Noriko Inada<sup>3</sup>, **Jenny C Mortimer**<sup>1,2</sup>\* (jcmortimer@lbl.gov)

<sup>1</sup>Joint BioEnergy Institute, Emeryville, CA; <sup>2</sup>Biosciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>3</sup>The University of Tokyo, Japan

### http://jbei.org/

Project Goals: The most abundant organic material on earth is lignocellulosic biomass or non-food plant material. JBEI's mission is to convert biomass to biofuels. The goal is to provide the nation with clean, renewable transportation fuels identical to gasoline, diesel and jet fuel. Building a successful lignocellulosic biofuels industry depends, in part, on developing specialized biofuel crops or feedstocks that are optimized for deconstruction into sugars and fermentation into biofuels and bioproducts.

Abstract: The majority of the nucleotide sugar substrates required for plant polysaccharide biosynthesis are made in the cytosol. Non-cellulosic polysaccharides, including hemicelluloses and pectins, are synthesized inside the Golgi lumen by glycosyltransferases. Nucleotide sugar transporters (NSTs) are therefore required to move the substrates to the correct compartment. Whilst Arabidopsis is estimated to have ~50 NSTs, only 4 are predicted to transport GDP-sugars. Arabidopsis synthesizes four different GDP-sugars: GDP-D-mannose, GDP-D-glucose, GDP-L-galactose and GDP-L-fucose which glycosylate pectins, hemicelluloses, proteins and sphingolipids.

Recent work has shown that the Arabidopsis GOLGI NUCLEOTIDE SUGAR TRANSPORTER1 (GONST1) specifically transports GDP-D-mannose for sphingolipid glycosylation<sup>1</sup> and GOLGI FUCOSE TRANSPORTER (GFT) transports GDP-L-fucose for multiple purposes, including xyloglucan fucosylation<sup>2</sup>. Here, we have characterized the remaining two predicted GDP-sugar transporters: GONST2 and GONST3, and identified specific functions for both. These data will be discussed in the context of engineering plant cell walls for improved production of biofuels and biochemicals.

### References

- 1. Mortimer et al. (2013) Plant Cell 25: 1881.
- 2. Rautengarten et al. (2016) Nature Communications 7:12119.

### Funding statement.

This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research,

through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy

### High-throughput quantitative proteomic profiling of *Escherichia coli* central carbon metabolism

Yan Chen,<sup>1,2</sup> Jonathan Vu,<sup>1,2</sup> Marcin P. Joachimiak,<sup>2,3,4</sup> David Ando,<sup>1,2</sup> Leanne Jade G. Chan,<sup>1,2</sup> Paul D. Adams,<sup>1,2,5</sup> Héctor García Martín,<sup>1,2,4</sup> **Christopher J. Petzold**<sup>1,2,4\*</sup> (cjpetzold@lbl.gov)

<sup>1</sup>DOE Joint BioEnergy Institute, Emeryville, CA; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>3</sup>DOE Systems Biology Knowledgebase, Emeryville, CA; <sup>4</sup>DOE Agile BioFoundry, Emeryville, CA <sup>5</sup>University of California, Berkeley, CA.

https://www.jbei.org

#### **Project Goals:**

Achieving a bioeconomy that produces "drop-in" and novel fuels drives many aspects of metabolic engineering and synthetic biology research. Yet, despite large investments in time and money many challenges remain to realizing economic conversion of sugar to biofuels. Thus, it is crucial to modify the host metabolism to produce large amounts of target molecules. These efforts benefit from large -omics datasets that inform predictions, however, many samples must be analyzed to understand the host metabolism and develop actionable information. This requires significant resources both to efficiently develop methods for new hosts and analyze them in a high-throughput manner. In this work, we describe a workflow using data acquired from discovery proteomic experiments and demonstrate its utility by quantifying 73 proteins from *E. coli* central metabolism in over 500 KEIO collection knockout strains grown in two media conditions with biological replicates. This work enables systems biology research which drives scientific discoveries, adds statistical power to biological hypotheses, and provides an extensive resource for metabolic engineering in *E. coli*.

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 and was part of the Agile BioFoundry (http://agilebiofoundry.org) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office.
#### Functional analysis of cell wall related genes in Sorghum bicolor

Tong Wei<sup>1,2\*</sup> (tongwei@lbl.gov), Guotian Li<sup>1,2</sup>, Zhanguo Xin<sup>3</sup>, Rashmi Jain<sup>2</sup>, Zhangyin Hao<sup>1</sup>, Devon Birdseye<sup>1</sup>, Dominique Loque<sup>1</sup>, Henrik Scheller<sup>1</sup>, **Pamela Ronald**<sup>1,2</sup>

<sup>1</sup> Feedstocks Division, Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>2</sup> Department of Plant Pathology and the Genome Center, University of California Davis, Davis, CA; <sup>3</sup> Plant Stress and Germplasm Development Unit, USDA-ARS, 3810 4th Street, Lubbock, TX

https://www.jbei.org/

# Project Goals: We are performing a functional genomics study of the cell wall related genes to improve biofuel-related traits in *Sorghum bicolor*.

It is of great importance to study and modify bioenergy crops to reduce the dependence on fossil fuels and release of climate-changing greenhouse gasses. Sorghum is tolerant to drought and grows on marginal lands, which makes it a potential bioenergy crop. To identify the genes involved in cell wall biosynthesis, we searched the sorghum genome and found 570 glycosyltransferase (GT) genes and 377 glycoside hydrolase (GH) genes. More than 60% of these genes have at least one mutant in our sorghum TILLING mutant collection. Here, we exemplified the sorghum functional genomics study by characterizing a sorghum cellulose synthase-like F6 (CslF6) in mixed-linkage  $\beta$ -glucan biosynthesis. Among nine sorghum mutants carrying mutations in CslF6, two lines developed obvious morphological phenotypes, the reduced plant height resembling the rice knockout mutant of *cslf6*. The gene expression assay showed that CslF6 is ubiquitously expressed across different tissue types, with higher expression in stems and young leaves. Further assays demonstrate that the mutants contain less mixed linkage glucan content compared with the wild type BTx623 plants. Our study demonstrates that sorghum CslF6 participates in mixed-linkage  $\beta$ -glucan deposition. The cell wall-related genes we identified here lay the foundation for the future biofuel research in sorghum.

The work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the US Department of Energy under Contract No. DE-AC02-05CH11231.

#### Life Cycle Assessment of Ionic Liquid-based Biofuel Production

Binod Neupane (bneupane@lbl.gov)\*, N.V.S.N. Murthy Konda, Blake A. Simmons, Seema Singh, and **Corinne D. Scown** 

Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA 94608

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

The main goals of this project are to investigate the environmental performance of various biofuel production pathways and develop a model for conducting life cycle environmental impact assessments of biofuels and chemicals developed at the Joint BioEnergy Institute. As biofuel production is a complex process requiring expertise in multiple scientific fields, this project requires close collaboration with other division members at the Lab including chemical engineers to incorporate process modeling and experimental results into life-cycle assessment models. This project will identify key opportunities for improving the economic and environmental performance of chemicals and fuels. The other goals include facilitating collaborations with academic, government, and industrial entities, and participate in educational and outreach.

#### Abstract text. Please limit to 2 pages.

There is a pressing need for efficient processes that can deconstruct biomass to sugars and subsequently convert sugars to biofuels, while achieving targeted environmental and socioeconomic benefits. Pretreatment of lignocellulosic biomass is a key step to overcome biomass recalcitrance and make the sugars and lignin available for further deconstruction and conversion. Although ionic liquids (ILs) are considered to be potential alternatives to conventional biomass pretreatment solvents, their greenhouse gas (GHG) emissions-intensity and water footprint are not yet well understood. In this study, we developed a life cycle assessment model to evaluate GHG emissions of IL production, specifically Cholinium Lysinate ([Ch][Lys]), and of its subsequent use as a pretreatment solvent in a corn stover-to-ethanol production. The results are then compared with the dilute acid (DA) pretreatment method (in the context of a cellulosic biorefinery) and conventional gasoline. The results suggest that, depending on the location of lysine production, GHG emissions for [Ch][Lys] production range from 6 to 8 kg CO<sub>2</sub>e/kg. Based on the biorefinery design considered, GHG emissions of IL-based biofuel range from 21 to 26 gCO<sub>2</sub>e/MJ (78 % reduction compared to gasoline). Additionally, we found that high IL recovery (>99 wt. %) is important from GHG emissions reduction perspective. Major contributors to the total GHG emissions includes fertilizers and GHG-intensive chemicals such as hydrochloric acid (HCl). This emphasizes the need for development of processes with high yields (thereby minimizing the GHG emissions associated with feedstock production) while minimizing the need for resource-intensive chemicals. Furthermore, the credits due to electricity export (i.e., by dispacing fossil derived electricity from grid) can be very high, which highlights

the need for energy-efficient biorefinery processes. Given the uncertainty with the several process parameters, detailed sensitivity analysis is conducted to better understand potential impact on emission intensity.

#### Funding statement.

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

### A Bacterial Pioneer Leaves a Complex Legacy

Sebastian Kolinko,<sup>1,2</sup> Yu-Wei Wu<sup>1,2</sup>, Firehiwot Tachea<sup>2,3</sup>, Leanne Jade Chan<sup>1,2</sup>, Paul D. Adams<sup>1,4</sup>, Todd Pray<sup>2,3</sup>, Deepti Tanjore<sup>2,3</sup>, Christopher J. Petzold<sup>1,2</sup>, John M. Gladden<sup>1,5</sup>, Blake A. Simmons<sup>1,2</sup> and **Steven Singer<sup>1,2</sup>**\*(SWSinger@lbl.gov)<sup>2</sup>

<sup>1</sup> Joint BioEnergy Institute, Emeryville, CA, USA; <sup>2</sup> Biological and Systems Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; <sup>3</sup> Advanced Biofuels Process Demonstration Unit, Emeryville, CA, USA; <sup>4</sup>Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; <sup>5</sup>Biological and Materials Science Center, Sandia National Laboratories, Livermore, California, USA

http://jbei.org

### Project Goals: The goal of this project, performed in the Microbial Communities group in the Deconstruction Division at JBEI, is to identify enzymes critical for polysaccharide hydrolysis in microbial consortia adapted to grow on biomass substrates and use these new enzymes to improve the conversion of biomass to biofuels

Cellulases have traditionally been identified and characterized from fungal and bacterial isolates. In natural environments, microbial consortia are responsible for lignocellulose deconstruction and may produce enzymes previously unobserved in isolates. To obtain new enzymes, a cellulolytic consortium was established by adapting aerobic compost microbial communities to grow with crystalline cellulose as the sole carbon source at elevated temperature. Soluble cellulases from this consortium provided the base for an enzymatic mixture, called Jtherm, that deconstructed biomass at temperatures up to 80°C. The consortium was scaled to 300 L to enable large-scale production of Jtherm. Time series metagenomic analysis of the consortium grown at 300 L indicated that an uncultivated Paenibacillaceae population was abundant early in the cultivation but became <1% abundant when the culture was harvested. Recovery of the genome of this Paenibacillaceae population, named *Candidatus* "Reconcillibacillus cellulovorans', from the metagenomic data demonstrated that it contained a putative operon with genes for a unique suite of multi-domain glycoside hydrolases (GH9, GH48, GH6/5 and GH10) and a polysaccharide monooxygenase (AA10). All of these proteins were multi-domain cellulases containing from one to three CBM3 binding modules in addition to the catalytic subunits, a structure similar to glycoside hydrolases from Caldicellulosiruptor. Biochemical purification of the Jtherm components demonstrated that the active components were the C. 'R. cellulovorans' GH9, GH6/5 and GH48, which were arranged in protein complexes that fundamentally differed from cellulosomes isolated from anaerobic Firmicutes. Recombinant expression of the three components of the complexes demonstrated that the GH9 and GH6/5 proteins had substantial activity on crystalline cellulose. This work demonstrates that the detailed study of microbial consortia can provide a deep understanding of community dynamics and provide access to novel protein structures for biomass deconstruction.

# This work was performed as part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and

*Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.* 

#### Droplet-Based Analog and Digital Microfluidic Platforms for High-Throughput Screening and Synthetic Biology Applications

Kosuke Iwai<sup>1,2</sup>\* (KosukeIwai@lbl.gov), Philip C. Gach<sup>1,2</sup>, Peter W. Kim<sup>1,2</sup>, Manasi Raje<sup>1,2</sup>, Joshua V. Heinemann<sup>1,3</sup>, Todd A. Duncombe<sup>1,2</sup>, Kai Deng<sup>1,2</sup>, Trent R. Northen<sup>1,3</sup>, Nathan J. Hillson<sup>1,3,5</sup>, Paul D. Adams<sup>1,4,6</sup>, and **Anup K. Singh**<sup>1,2</sup>

<sup>1</sup>Joint BioEnergy Institute, Emeryville, CA; <sup>2</sup>Biological and Engineering Sciences, Sandia National Laboratories, Livermore, CA; <sup>3</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>4</sup>Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>5</sup>DOE Joint Genome Institute, Walnut Creek, CA; and <sup>6</sup>University of California, Berkeley, CA

https://www.jbei.org/research/divisions/technology/microfluidic-assays/

#### **Project Goals:**

The JBEI mission is to conduct basic and applied research to enable economically-viable conversion of lignocellulosic biomass into biofuels to provide the nation with clean, renewable transportation fuels identical to gasoline, diesel and jet fuel. The goal of this project, performed in the Microfluidic Assays group in the Technology Division at JBEI, is to deliver the robust and easy-to-use microfluidic platforms to automate the synthetic biology process for advancing biofuels development.

In recent years, synthetic biology has dramatically grown and became significantly important for both of scientific researches and industrial applications such as biofuel and pharmaceutical applications. However, multiple genetic engineering steps required for synthetic biology are often time-consuming and labor-intensive with repetitive pipetting and plating. Therefore, automated and efficient processes to perform molecular biology assays have been long desired. Microfluidic assays and devices with aqueous droplets (microliter to picoliter in volume) suspended in oil phase as compartmentalized reaction chambers have attracted a significant attention for performing biochemical reactions and analysis as they provide drastic improvements over their macroscale counterparts with various benefits such as faster reaction time, less volume of reagent consumption required, better control of experimental environment, and higher throughput with multiplexed processes.

We are involved in developing innovative microfluidic assays and integrated devices for many biofuel research applications including enzyme screening, enzyme evolution and synthetic biology. Our hybrid microfluidic platforms utilize continuous-flow (analog) microfluidics that manipulate the droplets by controlling the hydrodynamic force, and digital microfluidics (DMF) that utilize surface tension from electrowetting on dielectric with arrayed electrodes. The systems can handle large numbers of droplets at once as well as actively manipulate target droplets in a programmable manner, and are capable of multiple steps of droplet manipulation

including formation of aqueous droplets and encapsulation of reagents and cells, hydrodynamic capture and array of the droplets, electric-field driven merge and split of the droplets to add specific amount and concentrations of various reagents, and incubation process with localized temperature control. In addition, we integrate optical fibers in the microchannels to add on-chip capability for fluorescence-based detection of encapsulated cells and enzymatic activities in the discrete droplets, and for triggering sorting of droplets. We are also integrating this platform with mass spectrometry to allow sensitive, label-free detection of chemicals and biofuels produced by the engineered cells.

One example application of our system is automation of synthetic biology experiments. Optimization of pathways can involve very large number of experiments as multiple variants are available for each gene. Our platform can integrate and automate the processes of DNA assembly, transformation, and cell culture in one device. We show that the platform is capable of accurate DNA assembly, efficient transformation, and cell culture and is compatible with many cloning methods (e.g., Golden Gate and Gibson) and chassis organisms (e.g., bacteria, yeast and fungus). We additionally demonstrate capability of the system for on-chip gene editing of *Saccharomyces cerevisiae* utilizing the CRISPR-Cas9 based cloning-free toolkit.

Unlike conventional microtiter plate based reactions, our analog-digital microfluidic platforms with on-chip fluorescence detection allow completely automated genetic engineering steps using 10-100-fold lower amounts of reagents and can be useful for application requiring high throughput screening and reactions, and integration with mass spectrometry enables higher sensitivity detection.

#### References

 S.C.C. Sigh, G. Goyal, P.W. Kim, N. Koutsoubelis, J.D. Keasling, P.D. Adams, N.J. Hillson and A.K. Singh, "A Versatile Microfluidic Device for Automating Synthetic Biology," *ACS Synth. Biol.*, 2015, 4, 1151-1164.
S.C.C. Shih, P.C. Gach, J. Sustarich, B.A. Simmons, P.D. Adams, S. Singh and A.K. Singh, "A Droplet-to-Digital (D2D) Microfluidic Device for Single Cell Assays," *Lab Chip*, 2015, 15, 225-236.
P.C. Gach, S.C.C. Shih, I. Sustarich, I.D. Keasling, N.L. Hillson, P.D. Adams and A.K. Singh, "A Droplet

3. P.C. Gach, S.C.C. Shih, J. Sustarich, J.D. Keasling, N.J. Hillson, P.D. Adams and A.K. Singh, "A Droplet Microfluidic Platform for Automating Genetic Engineering," *ACS Synth. Biol.*, 2016, **5**, 426-433.

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. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy's Nuclear Security Administration under contract DE-AC04-94AL85000.

#### Ionic Liquid Pretreatment Technology: Challenges and Opportunities

Seema Singh\* (ssingh@lbl.gov), Blake Simmons, Jian Sun, Tanmoy Dutta, Kwang-Ho Kim and Dajiang Liu Deconstruction Division, Joint Bioenergy Institute, Emeryville, CA

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

Ionic Liquid (IL) pretreatment technology is overcoming some of the barriers that have historically prevented industrial implementation of biochemical conversion technologies for lignocellulosic biomass utilization for biofuels and bioproducts. IL pretreatment technology is evolving rapidly due to the development of new ILs, process integration, and task specificity of certain ILs. Despite the enormous potential, there are certain challenges that have to be addressed in order to make IL pretreatment an economically viable technology in the biorefinery context. The critical next step for the biorefinery implementation of IL conversion technology is fundamental study and design of new ILs that are cheap, biocompatible and renewable, and development of cost effective approaches for IL recycling and product recovery. In this presentation, we will discuss what makes IL pretreatment unique, why we must continue to examine IL technology and what are the technical challenges and opportunities for IL based technology for lignocellulosic biomass processing

#### Funding Acknowledgement:

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.

#### Genome Wide Identification of Bacterial Membrane Capacity Determinants

Thomas Eng,<sup>1,2</sup> (tteng@lbl.gov), Heather M. Jensen,<sup>1,2</sup> Robin Herbert,<sup>1,2</sup> and Aindrila Mukhopadhyay<sup>1,2</sup>

<sup>1</sup>Joint Bioenergy Institute, Emeryville, CA; <sup>2</sup>Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Project Goals: The production of fuels and bioproducts in microbial hosts require significant development and optimization of not only the biosynthetic pathway but also the host chassis. However, the initial choice of microbial host has a considerable yet often overlooked impact the success of the project. The goals of this project are focused on evaluating the determinants governing membrane capacity, as membrane-bound proteins are key to many aspects of the host chassis optimization. We assembled a high-throughout method to rapidly screen a saturated transposon library in *Escherichia coli* covering all non-essential genes for single gene mutants that increase the overall cellular membrane capacity.

The production of fuels and bioproducts in a microbial host requires significant development and optimization of any given heterologous gene pathway. However, the initial choice of microbial host for gene expression has a considerable yet often overlooked impact the success of the project, as the microbial host must withstand a range of environmental and physiological perturbations to yield high levels of the desired product. In this project we have specifically aimed to understand the determinants governing membrane capacity, as membrane-bound proteins play a central role in many functions relevant to optimal carbon utilization and bioproduction. Examples include cellular transporters that regulate the import of the carbon source, sensors that respond to the composition of the cytoplasm and most importantly, the export of toxic intermediates or end-products from the intracellular-milieu.

While eminently useful, the overexpression of such membrane proteins often causes growth inhibition placing an inherent upper limit on expression. Membrane-protein overexpression can result in nonfunctional or mislocalized protein complexes, morphological deformities, trigger cellular stress response and a metabolic shunt to acetate. We hypothesized that it would be possible to increase the total cellular capacity for membrane protein expression by inactivating single genes that are dispensable for growth under laboratory conditions.

In this study, we assembled a high-throughout method to rapidly screen a saturated transposon library in *Escherichia coli* covering all non-essential genes for single gene mutants which increased the overall cellular membrane capacity. We assessed membrane protein expression

using the fluorescence of a membrane protein fused to GFP as a proxy. We report the discovery of single gene knockouts that significantly improve the expression of candidate membrane proteins in these mutant strains. These candidate mutant alleles were then validated using independently derived isogenic deletion mutants. We provide evidence that the overexpression of these membrane-proteins, achieved using expression in gene deletion backgrounds, leads to the desired improvement in the related phenotypic fitness. Our observations provide a microbial-host chassis that may be immediately applicable to any membrane protein expression system without further optimization of transcription or translation.

#### Funding:

This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

#### Quantitative metabolic modeling at the Joint BioEnergy Institute (JBEI)

**Héctor García Martín**,<sup>1,2,3</sup>\* (hgmartin@lbl.gov), David Ando,<sup>1,2</sup> Zak Costello,<sup>1,2,3</sup> Tyler Backman,<sup>1,2,3</sup> Jennifer Gin,<sup>1,2</sup> Jason Brice,<sup>1,2</sup> Christopher Beitel,<sup>1,2</sup> William Morrell,<sup>1,4</sup> Garrett Birkel,<sup>1,2</sup> Teresa Lopez,<sup>1,2</sup> Mark Forrer,<sup>1,4</sup> Nathan Hillson,<sup>1,2,3</sup> Christopher Petzold,<sup>1,2,3</sup>

<sup>1</sup>DOE Joint BioEnergy Institute, Emeryville, CA; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>3</sup>DOE Agile BioFoundry, Emeryville, CA; <sup>4</sup>Biotechnology and Bioengineering and Biomass science and Conversion Department, Sandia National Laboratory, Livermore, CA

https://www.jbei.org

#### **Project Goals:**

The Quantitative Metabolic Modeling (QMM) directorate at the Fuels Synthesis Division is devoted to developing models of metabolism which are both quantitative and predictive, in order to improve biofuel production in a rationally directed fashion. Here, we present our work in three different thrusts: machine learning modeling, mechanistic modeling, and software development. Machine learning modeling is used to systematically improve biofuel production without a detailed understanding of pathway dynamics and, separately, is combined with kinetic modeling in order to predict time-dependent metabolomic data from measured protein levels. Mechanistic models involving <sup>13</sup>C Metabolic Flux Analysis for genome-scale models are used to produce actionable items for metabolic engineering that have improved fatty acid production by 70%. All these methods are supported by a suite of software tools that improve productivity: the Experiment Data Depot (EDD) stores data and metadata for a variety metabolic engineering experiments. EDD allows easy visualization of these data and its access in a standardized manner for use with modelling approaches. Arrowland (https://public-arrowland.jbei.org/) provides an intuitive and interactive tool to explore -omics data and make predictions on possible metabolic engineering approaches. The JBEI QMM library is a python open source library that provides the tools for all the aforementioned capabilities.

This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, and was part of the Agile BioFoundry (http://agilebiofoundry.org) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

### **Development of Novel Approaches to Optimize Energy Crops**

#### Dominique Loqué

Joint BioEnergy Institute, Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory; Plant and Microbial Biology department, University of California, Berkeley dloque@lbl.gov

Development of highly performant energy-crops is needed to provide sustainable and large resources of biomass to feed our biorefineries and to support cost-effective conversion of the biomass into biofuels and bio-products. Plant biomass is mainly composed of polymers made of fermentable sugars (cellulose and hemicellulose) that are embedded in a robust aromatic polymer called lignin. Recalcitrant to degradation, lignin inhibits efficient extraction and hydrolysis of cell wall polysaccharides and prevents low-cost lignocellulosic-biofuel production. Unfortunately, content and composition of these polymers cannot be drastically manipulated to the same extend in all tissues without causing deleterious consequences on plant productivity. Therefore, technologies allowing the precise manipulation of content and composition of various components of plant cell walls should be developed to facilitate the production of cheap and large quantities of fermentable sugars without compromising plant growth. Moreover, engineering plants with complex metabolic pathways or multiple traits is often inhibited by the number of genes that are required to reach the final product. It shows the need of synthetic biology tools to express multiple genes with controllable expression strengths and in specific tissues. Here we highlight our progress in synthetic biology to support plant engineering which includes in vitro and in vivo DNA assembly methods to stack multiple gene cassettes, promoter libraries developed for root specific expression, and novel devises to fine-tune gene expression in plants. We believe the development of these tools and approaches have the potential to support scientists and engineers who are looking at stacking and controlling multiple genes and interested in manipulating endogenous metabolic pathways.

This work was part of the DOE Early Career Award and the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research; and U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

#### Linking Soil Biology and Chemistry Using Knowledge of Exometabolite Substrate Preferences

Tami L. Swenson<sup>1</sup>\* (tlswenson@lbl.gov), Ulas Karaoz<sup>1</sup>, Joel M. Swenson<sup>1</sup>, Richard Baran<sup>2</sup>, Rebecca Lau<sup>1</sup>, Benjamin P. Bowen<sup>1</sup>, **Trent Northen**<sup>1,3</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>2</sup>Thermo Fisher Scientific, San Jose, CA; <sup>3</sup>DOE Joint Genome Institute, Walnut Creek, CA

http://www.northenlab.org/research/biological-soil-crusts-biocrusts/

Project goals: The Department of Energy has made major investments in soil sequencing efforts that have the potential to revolutionize predictive models of soil nutrient cycling. However, we lack vital data to link sequence data to metabolic transformations in soils. This program aims to help bridge this gap by pioneering new soil metabolomics approaches that link microbial community structure to soil organic matter dynamics.

The specific connections between soil metabolites and microbial community structure are not well-understood. The dramatic expansion of sequencing data provides a window into microbial community structure and metabolic potential; however, linking these data to exogenous metabolites that microbes process and produce (the exometabolome) remains challenging. Here we explore the ability to use exometabolite profiles of biological soil crust (biocrust) isolates to relate the relative abundance of microbes and metabolites *in situ* (intact biocrusts). Biocrusts are microbial communities that inhabit the surface of soils found in arid regions comprising more than 40% of Earth's terrestrial surface and play a critical role in soil stabilization and nutrient cycling.

Recently we characterized the substrate preferences of seven biocrust bacterial isolates using exometabolomics with the finding that there was very little overlap in metabolite utilization indicating strong niche partitioning (Baran et al, 2015). We next wanted to determine if these metabolite preferences link the relative abundance of these bacteria to soil metabolites in intact biocrusts. To facilitate correlation between soil microbes and metabolites, we focused on the cascade of microbial activity set in motion upon wetting dry soil along a successional gradient. At five timepoints and across four successional stages following wetup, soil water was collected and analyzed by normal-phase liquid chromatography/ mass spectrometry for metabolites and DNA was extracted and sequenced on the HiSeq sequencing platform. Ribosomal protein (L15) was used as a phylogenetic marker.

Biocrust wetting caused a dramatic shift in both microbial community structure and metabolite profiles. Four dominant organisms in the biocrust were close phylotypes of the exometabolite-profiled isolates: a cyanobacterium (*Microcoleus* spp.), two firmicutes (*Anoxybacillus* sp. and *Bacillus* sp.) and an actinobacterium (*Blastococcus* sp.). In general, soil metabolites displayed the expected directionality (positive or negative correlation) with isolate phylotype abundance for 76% of the metabolites examined. Specifically, 78% of the metabolites that were consumed by an isolate were negatively correlated with their phylotypes *in situ* and 73% of released metabolites were positively correlated. Our results demonstrate the potential of exometabolite profiling in providing a functional link between microbial community structure and chemical composition suggesting that large-scale exometabolite profiling using environmentally relevant metabolites may enhance interpretation of sequence-based microbial community analysis.

#### References

Baran, R., Brodie, E.L., Mayberry-Lewis, J., Hummel, E., Da Rocha, U.N., Chakraborty, R., Bowen, B.P., Karaoz, U., Cadillo-Quiroz, H., Garscial-Pichel, G. & Northen, T.R. Exometabolite niche partitioning among sympatric soil bacteria. *Nat Commun.*, **6**, 8289 (2015).

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

## Formation of a phenyl-choline ether structure in lignin reduces inhibition of cellulase activity by lignin

Jijiao Zeng<sup>1</sup>, Sun Jian<sup>1</sup>, John Gladden<sup>1</sup>, Seema Singh<sup>1</sup>, Blake A. Simmons<sup>3</sup>, Michael kent<sup>2</sup> and **Kenneth L. Sale<sup>1\*</sup>** (klsale@lbl.gov)

<sup>1</sup>Deconstruction Division, Joint BioEnergy Institute / Sandia National Laboratories, Emeryville, CA, <sup>2</sup>Sandia National Laboratories, Albuquerque, NM, <sup>3</sup>Lawrence Berkeley National Laboratories, Berkeley, CA

#### **Project Goals:**

Ionic liquids have been extensively studied as solvents for biomass pretreatment and have been shown to be excellent solvents for selective isolation of cell wall components. Work in the deconstruction division at the Joint BioEnergy institute (JBEI) has focused on developing lower cost, biocompatible and adaptable ionic liquid technologies for application in biorefineries. Our work has demonstrated that choline-based ionic liquids are highly efficient at solubilizing lignin during biomass pretreatment. However, the mechanism by which lignin from different biomass sources is solubilized by choline-based ionic liquids is still unclear. In this study, we show that lignin is chemically modified by choline during pretreatment with choline glutamate ([Ch][Glu]) and choline  $\alpha$ -ketoglutarate ([Ch][aKg]). Interestingly, these choline-based ionic liquids modify lignin by forming a new phenyl-choline ether (4-O-C) bond, which dramatically increases the solubility of modified lignin in aqueous solution. We also analyzed the thermal stability and inhibition kinetics of a selected endo-1.4-β-D-glucanase (Cel5A from *Thermotoga maritima*) in the presence of choline-modified and unmodified lignin. Our results show that both cholinemodified lignin and unmodified lignin enhance the melting temperature of Cel5A and bovine serum albumin (used as a control), indicating a stabilizing interaction. We show that native lignin inhibits Cel5A via an uncompetitive inhibition mechanism, whereas choline modified lignin is less inhibitory, having a 10X higher IC50. These results support the development of cholinebased bio-compatible one-pot pretreatment and saccharification technologies.

#### Funding statement.

This work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. This work was also supported by the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

#### At the Interface: Glycoproteins, Glycolipids, and WAK-mediated Signaling are Required for Plant-Microbial Symbiosis in Medicago truncatula

William Moore<sup>1,2\*</sup> (wmoore@lbl.gov), Candace Chan<sup>1</sup>, Oge Nnadi, <sup>1</sup> Paul Hussey<sup>1</sup>, Nathan Hillson<sup>1</sup>, and **Henrik Scheller**<sup>1,2</sup>

<sup>1</sup>Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, Emeryville, CA <sup>2</sup>Department of Plant & Microbial Biology, University of California, Berkeley

In nature plants interact with beneficial root associated bacteria and fungi that aid in nutrient uptake and promote plant growth and resilience. Here we use Medicago truncatula as a model plant to investigate the role of the plant cell surface in beneficial plant-microbe interactions with arbuscular mycorrhizal fungi (AMF) and nitrogen-fixing Sinorhizobium. The plant cell surface plays a key role in symbiosis by forming a specialized plant-microbial interface composed of plant-derived membranes, cell wall polysaccharides, and protein complexes, through which nutrients and information are bi-directionally exchanged<sup>1,2</sup>. It has previously been reported in the literature through the use of glycan-directed monoclonal antibodies that arabinogalactan proteins (AGPs) aggregate at the symbiotic interface in a variety of plant-microbe mutualisms, however, the identity of the genes encoding these glycoproteins has remained unknown<sup>3-6</sup>. Here we report the discovery of several AGP encoding genes in *M. truncatula* that are specifically expressed during symbiosis with either AMF or Sinorhizobium meliloti. Functional studies using RNAimediated knockdown of specific AGPs result in drastic symbiotic phenotypes. including poor mycorrhizal colonization and impaired nitrogen fixation in Sinorhizobum infected root nodules. In parallel, we have identified a wall-associated kinase (WAK)-like receptor and a glycosyltransferase enzyme involved in sphingolipid biosynthesis that are also required for symbioses. Our data indicate that glycopeptides, glycolipids, and WAK mediate signaling at the cell surface is necessary for establishing and maintaining symbiosis in *M. truncatula*.

This work was supported by the NSF graduate research fellowship program and was part of the DOE Joint BioEnergy Institute (http://<u>www.jbei.org</u>) 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

#### References

- 1. Balestrini, R. & Bonfante, P. The interface compartment in arbuscular mycorrhizae: A special type of plant cell wall? *Plant Biosystems An International Journal Dealing with all Aspects of Plant Biology* **139**, 8–15 (2005).
- 2. Brewin, N. J. Plant Cell Wall Remodelling in the Rhizobium–Legume Symbiosis. *Critical Reviews in Plant Sciences* **23**, 293–316 (2004).
- 3. Jackson, O., Taylor, O., Adams, D. G. & Knox, J. P. Arabinogalactan Proteins Occur in the Free-Living Cyanobacterium Genus Nostocand in Plant– NostocSymbioses. *Mol. Plant Microbe Interact.* **25**, 1338–1349 (2012).

- 4. Bonfante-Fasolo, P., Tamagnone, L. & Peretto, R. Immunocytochemical location of hydroxyproline rich glycoproteins at the interface between a mycorrhizal fungus and its host plants. *Protoplasma* (1991).
- 5. Tsyganova, A. V. *et al.* Distribution of legume arabinogalactan protein-extensin (AGPE) glycoproteins in symbiotically defective pea mutants with abnormal infection threads. *Cell Tiss. Biol.* **3**, 93–102 (2009).
- 6. Berry, A. M. *et al.* Arabinogalactan proteins are expressed at the symbiotic interface in root nodules of Alnus spp. *New Phytologist* **155**, 469–479 (2002).

### **Exposure History Dependence of Microbial Mediated Substrate Transformation Rates in Groundwater**

Charles J. Paradis<sup>1,\*</sup>(charlesjparadis@gmail.com), **Terry C. Hazen**<sup>1, 2</sup>, Adam P. Arkin<sup>3</sup>, and Paul D. Adams<sup>3</sup>

<sup>1</sup>University of Tennessee, Knoxville, TN, <sup>2</sup>Oak Ridge National Laboratory, Oak Ridge, TN, <sup>3</sup>Lawrence Berkeley National Laboratory, Berkeley, CA, <sup>\*</sup>Presenting author

http://enigma.lbl.gov/research/environmental-microbiology/

# Project Goals: The goals of this project are to: (1) demonstrate the exposure history dependence of microbial mediated substrate transformation rates in groundwater at the field scale and (2) elucidate the microbial mechanism(s) which control the exposure history dependence of microbial mediated substrate transformation rates

The rates at which natural microbial communities can transform a substrate in groundwater have been shown to increase after repeated exposures to the substrate; herein referred to as the "memory effect". The objectives of this study were to determine: (1) how long the memory effect can last and (2) how the memory effect can alter the structure and function of natural microbial communities. Ethanol substrate was injected into a single groundwater test well for six consecutive weeks in order to establish a memory effect. The groundwater control well, located up-gradient of the test well, was not injected with ethanol during this time. The rate of ethanol removal in the test well was negligible the first week whereas subsequent rates were significant. The test and control wells will be monitored for six additional weeks under ambient conditions. Ethanol substrate will then be injected into both test and control wells in order to determine: (1) if the test well retained its memory effect and (2) if the rate of ethanol removal in the control wells is negligible.

Here we present the hydrological, geochemical, and microbiological data and analyses in hand from the study site and the experimental well pair. This includes: (1) the direction and magnitude of groundwater velocity, (2) the effective porosity of the groundwater system, (3) diffusive mass transport in the experimental well pair, (4) the rates of ethanol removal in the test well, (5) the extent of nitrate, sulfate, and uranium removal in the test well, (6) the extent of limiting metal nutrient and/or co-factor removal in the test well, (7) microbial community structure (16S rRNA sequencing) at the study site, and (8) microbial community function (GeoChip) at the study site.

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

## Title: High-throughput testing of carbon source on microbial community assembly and antibiotic production

Authors: Lauren Lui<sup>1,2</sup>\* (lmlui@lbl.gov), Hans Carlson<sup>1</sup>, Andrew Sczesnak<sup>1</sup>, Adam M. Deutschbauer<sup>1</sup>, Adam P. Arkin<sup>1</sup>, and Paul D. Adams<sup>1</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley CA

#### http://enigma.lbl.gov

Project Goals: Utilize a high-throughput culture-based approach to more easily understand ecological constraints that affect microbial community structure. Specifically these experiments aim to identify how selective pressures of carbon resource type and diversity contributed to changes in community structures.

**Abstract**: We are working with 12 isolates of Pseudomonas spp., the most common isolates from ENIGMA studies at the Oakridge Field Research Center. This synthetic community provides a system to understand how environmental conditions affect community assembly and how closely related species interact with each other. Typically, closely related microbial species inhabit similar ecological niches and will inhibit the growth and survival of each other as they compete for the same resources. Pseudomonas are known to produce bacteriocins, which inhibit closely related strains and species, and other toxic small molecules, such as phenazines.

We are testing the effect of 96 carbon sources on the structure of our synthetic community in regards to (1) growth on selective carbon sources and (2) antibiotic production. Carbon sources, such as sugars, have been demonstrated to regulate antibiotic production in various bacterial species, including Pseudomonas species. We report our initial results on identifying selective carbon sources for growth and carbon sources that induce microbes to produce inhibitory compounds. We are also using a new microfluidics method developed in our lab to help determine genes important for species interactions in these conditions. In the future we can apply these same methods to study other ENIGMA isolates and the effects of other substrates and conditions such as metal ions, nutrient levels, and proportions of nutrients on community assembly.

#### Identification of Novel Biosynthetic and Catabolic Pathways in Diverse Bacteria Using High-throughput Genetics

Morgan N. Price<sup>1</sup>, Kelly M. Wetmore<sup>1</sup>, R. Jordan Waters<sup>2</sup>, Mark Callaghan<sup>1</sup>, Jayashree Ray<sup>1</sup>, Jennifer V. Kuehl<sup>1</sup>, Ryan A. Melnyk<sup>1</sup>, Jacob S. Lamson<sup>1</sup>, Yumi Suh<sup>1</sup>, Zuelma Esquivel<sup>1</sup>, Harini Sadeeshkumar<sup>1</sup>, Romy Chakraborty<sup>3</sup>, Benjamin E. Rubin<sup>4</sup>, Grant M. Zane<sup>5</sup>, Judy D. Wall<sup>5</sup>, James Bristow<sup>2</sup>, Matthew J. Blow<sup>2</sup>, Adam P. Arkin<sup>1,6</sup>, **Adam M. Deutschbauer<sup>1\*</sup>** (**AMDeutschbauer@lbl.gov**), Paul D. Adams<sup>6,7</sup>

<sup>1</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>2</sup>Joint Genome Institute, Walnut Creek, CA; <sup>3</sup>Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory; <sup>4</sup>Division of Biological Sciences, University of California, San Diego; <sup>5</sup>Biochemistry Division, University of Missouri, Columbia, Missouri; <sup>6</sup>Department of Bioengineering, University of California, Berkeley; <sup>7</sup>Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA

#### http://fit.genomics.lbl.gov/; http://enigma.lbl.gov/;

Project Goals: For most bacteria with sequenced genomes, we do not understand how they synthesize some amino acids or how they consume some carbon sources. This makes it challenging to reconstruct their metabolism or to predict their ecological role from their genome sequence. We used genome-wide mutant fitness data from 14 different genera of heterotrophic bacteria to fill gaps in amino acid biosynthesis pathways and to identify novel catabolic routes for several sugars.

The genomes of many free-living bacteria do not appear to encode some of the necessary enzymes for amino acid biosynthesis. This has led to speculation that these bacteria cannot make all 20 amino acids and that they are cross-feeding each other amino acids. However, when we tested 25 heterotrophic Proteobacteria from 14 genera, we found that 24 of them grew in minimal media without any added amino acids. In contrast, comparative genomics tools predict that all of these bacteria, except for *Escherichia coli*, are auxotrophic for multiple amino acids. We examined representatives of 10 genera in more detail and identified 11 gaps that we could not fill using current knowledge. Using genome-wide mutant fitness data, we identified novel enzymes that fill 9 of the 11 gaps and hence explain the biosynthesis of methionine, threonine, serine, or histidine by bacteria from six genera. We also found that the sulfate-reducing bacterium *Desulfovibrio vulgaris* synthesizes homocysteine (which is a precursor to methionine) by using DUF39, NIL/ferredoxin, and COG2122 proteins, and that homoserine is not an intermediate in this pathway. It appears that most free-living bacteria can make all 20 amino acids but we do not yet know how.

We also used the genome-wide fitness data to identify novel catabolic pathways for several sugars:

- Oxidation of 2-deoxy-D-ribose, a component of DNA, by *Pseudomonas simiae* (4 novel reactions)
- Oxidation of D-arabinose by *Sinorhizibium meliloti* (2 novel reactions)
- Additional dehydratase steps in the oxidation of L-fucose by Sinorhizobium meliloti

• The first identification of a gene for D-glucosaminate ammonia-lyase, which is required for glucosamine catabolism in many *Pseudomonas* 

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

#### Mechanism for Microbial Population Collapse in a Fluctuating Resource Environment

Serdar Turkarslan<sup>1</sup>\*(sturkarslan@systemsbiology.org), Arjun V Raman<sup>1</sup>, Anne W Thompson<sup>1</sup>, Mark A Gillespie<sup>1</sup>, Kristina L Hillesland<sup>4</sup>, Christina E Arens<sup>1</sup>, Sergey Stolyar<sup>1</sup>, Adrian López García de Lomana<sup>1</sup>, David J Reiss<sup>1</sup>, Frederick von Netzer<sup>2</sup>, Drew Gorman-Lewis<sup>5</sup>, Grant M Zane<sup>3</sup>, Jeffrey A Ranish<sup>1</sup>, Judy D Wall<sup>3</sup>, David A Stahl<sup>2</sup>, and **Nitin S Baliga<sup>1</sup>**, Adam P Arkin<sup>6</sup>, Paul D. Adams<sup>6</sup>

<sup>1</sup>Institute for Systems Biology, Seattle, WA, <sup>2</sup>Civil and Environmental Engineering, University of Washington, Seattle, WA, <sup>3</sup>Department of Biochemistry, University of Missouri, Columbia, MO, <sup>4</sup>Biological Sciences, University of Washington Bothell, Bothell, WA, <sup>5</sup>Earth and Space Sciences, University of Washington, Seattle, WA, <sup>6</sup>Lawrence Berkeley National Laboratory, Berkeley, CA.

Project URL: enigma.lbl.gov, http://networks.systemsbiology.net/syntrophy

#### **Project Goals:**

Microbial generalists are metabolically versatile and able to cope with periodically or randomly fluctuating natural environments by appropriately up- or down-regulating genes to optimize resource utilization. Not surprisingly, many experimental evolution studies have reported that mutations accumulate in regulatory genes during metabolic specialization of a generalist to a single resource environment. However, it is generally assumed and not demonstrated whether or how regulatory mutations foster adaptive evolution to new environments. The goal of this project is to provide insight into the complex intertwined relationship between regulatory architecture of a generalist and the spatiotemporal structure of environmental changes. We aim to bring together technologies and concepts in a systems biology framework to provide comprehensive characterization of microbial resilience from genome-wide and population scale to detailed molecular and single cell level.

#### Abstract:

Managing tradeoffs through gene regulation is believed to confer resilience to a microbial community in a fluctuating resource environment. To investigate this hypothesis we imposed a fluctuating environment that required the sulfate-reducer *Desulfovibrio vulgaris* to undergo repeated ecologically-relevant shifts between retaining metabolic independence (active capacity for sulfate respiration) and becoming metabolically specialized to a mutualistic association with the hydrogen consuming *Methanococcus maripaludis*. Strikingly, the microbial community became progressively less proficient at restoring the environmentally-relevant physiological state

after each perturbation and most cultures collapsed within 3-7 shifts. Counter-intuitively, the collapse phenomenon was prevented by a single regulatory mutation. We have characterized the mechanism for collapse by conducting RNA-seq analysis, proteomics, microcalorimetry, and single cell transcriptome analysis. We demonstrate that the collapse was caused by conditional gene regulation, which drove precipitous decline in intracellular abundance of essential transcripts and proteins, imposing greater energetic burden of regulation to restore function in a fluctuating environment.

#### **Publications:**

Turkarslan S., Raman AV., Thompson AW., Gillespie MA., Hillesland KL., Arens CE., Stolyar S., López García de Lomana A., Reiss DJ., von Netzer F., Gorman-Lewis D., Zane GM., Ranish JA., Wall JD., Stahl DA., and Baliga NS. (2016) Mechanism for Microbial Population Collapse in a Fluctuating Resource Environment. *Mol Sys Biol* (in revision)

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

#### Title: Mechanisms of Uranium Reduction in Sulfate-Reducing Bacteria

Author(s): Erica L-W. Majumder<sup>1\*</sup>(<u>majumdere@missouri.edu</u>), Sharien Fitriasari<sup>1</sup>, Judy D. Wall<sup>1</sup>, Adam Arkin<sup>2</sup>, Paul D. Adams<sup>2</sup>(<u>PDAdams@lbl.gov</u>)

**Author Affiliation:** <sup>1</sup>University of Missouri, Columbia MO; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley CA

#### http://enigma.lbl.gov

**Project Goals:** The goals of this project are to understand how Sulfate-Reducing Bacteria handle metal contamination on a molecular level. The Oak Ridge Reservation Field Research Site has many contaminating toxic metals that along with other geochemical factors drive community formation. SRB are found throughout the site in significant numbers. In the work, we are determining how the SRB tolerate, resist and chemically transform one of the major contaminating metals, uranium.

Abstract: Sulfate-reducing bacteria (SRB) are anaerobic soil bacteria that interact with many metals, over twenty have been have been characterized to date. Some interactions are detrimental to human interests, such as bio-corrosion of pipes and ships, but others are potentially beneficial, such as the reduction of soluble uranium (VI) to insoluble U(IV). The process of uranium reduction has been investigated in the model SRB, Desulfovibrio vulgaris Hildenborough because of the presence of SRB at the Oak Ridge Reservation in the areas contaminated with uranium and other metals. Single deletion mutants were created for each of the genes proposed in the literature to be involved in the reduction process: cytoplasmic atypical thioredoxins on a putative metal reducing operon, periplasmic tetraheme cytochrome  $c_3$  and extracellular pili forming protein PilA. Mutant strains were tested for their ability to reduce uranium, tolerate uranium, and respire uranium. The thioredoxin mutants were able to reduce and tolerate almost as much uranium as wild-type suggesting that they are not the main contributor to uranium reduction. Both the cytochrome  $c_3$  mutant and the pili-less mutant were significantly impaired in their ability to reduce and tolerate uranium suggesting that extracellular and periplasmic mechanisms are more important for this interaction than cytoplasmic. Reduction rates were increased for wild type and mutant strains under pyruvate fermentation rather than sulfate respiration indicating that metal reduction might be more favorable when the organism employs fermentative pathways. Key electron donors, acceptors and proteins were identified for this SRB mechanism of uranium reduction.

#### Microbes at the blurred boundary of natural and built environments

Fangqiong Ling<sup>1\*</sup> (<u>qiong@mit.edu</u>), Johnathan Friedman<sup>1</sup>, Shijie Zhao<sup>1</sup>, Mark B. Smith<sup>1</sup>, Andrea M. Rocha<sup>2,3</sup>, Charles J. Paradis<sup>2,3</sup>, Jizhong Zhou<sup>4</sup>, Terry C. Hazen<sup>2,3</sup>, **Eric J. Alm<sup>1</sup>**, Adam P. Arkin<sup>5,6</sup> and Paul D. Adams<sup>5,6</sup>

<sup>1</sup>Massachusetts Institute of Technology, Cambridge, MA; <sup>2</sup>University of Tennessee, Knoxville, Knoxville, TN; <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge, TN; <sup>4</sup>University of Oklahoma, Norman, OK; <sup>5</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>6</sup>University of California, Berkeley, CA

http://enigma.lbl.gov

### Project Goals: This project aims to better understand the assembly of low-diversity communities in a groundwater ecosystem with genomic tools.

Humans today spend more time working, living, and recreating in the manmade environment than the natural environment. However, the boundary between built and natural environments isn't always clear. The Oak Ridge Field Research Center (FRC) presents a case where past anthropogenic contamination to the natural environment caused by uranium enrichment influences the current built environment through the channel of water supply. Microbes living in this environment have been shown to be useful as quantitative biosensors of contamination. In the present study, we found that microbial communities at the FRC sites exhibited an unusual range of diversity, which correlates to the performance of the microbial sensor. We show that the variation in diversity is poorly explained by chemical gradients or cell count, yet well explained by migration when viewed from a dispersal-limited community assembly perspective. Our on-going work applies shotgun sequencing to infer strain-level diversity that can provide further insights on the community assembly processes.

#### Funding statement.

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

Fangqiong Ling is supported by Alfred P. Sloan Foundation Microbiology of the Built Environment Postdoctoral Fellowship

#### Microbial Interactions with Natural Organic Matter Extracted from the Oak Ridge FRC

**Romy Chakraborty**<sup>1</sup>\* (rchakraborty@lbl.gov), Xiaoqin Wu<sup>1</sup>, Terry C. Hazen<sup>2</sup>, Yina Liu<sup>3</sup>, Nancy Hess<sup>3</sup>, Matthew W. Fields<sup>4</sup>, Ping Zhang<sup>5</sup>, Liyou Wu<sup>5</sup>, Jizhong Zhou<sup>5</sup>, Qinghao Li<sup>1</sup>, Wanli Yang<sup>1</sup>, Adam P. Arkin<sup>1</sup> and Paul D. Adams<sup>1</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley CA; <sup>2</sup>Oak Ridge National Laboratory, Oak Ridge, TN & University of Tennessee, Knoxville, TN; <sup>3</sup>Pacific Northwest National Laboratory, Richland, WA; <sup>4</sup>Montana State University, Bozeman, MT; <sup>5</sup>University of Oklahoma, Norman, OK

http://enigma.lbl.gov

Project Goals: Natural organic matter (NOM) is central to microbial food webs and microbially mediated NOM transformations determine much of the carbon (C) flux in subsurface environments. However, little is known about the molecular signature of this pool of C and the microbial activities that regulate NOM turnover are still poorly resolved. The goal of this project was to study the interactions between NOM (extracted from the field site) and native microbial communities present in groundwater at a background site (FW305) at Oak Ridge Field Research Center, TN.

Water-soluble NOM was extracted from sediment samples collected from the background uncontaminated site, and the extraction efficiencies were 3.2% for organic carbon and 1.6% for inorganic carbon. Extracted NOM was used as the sole source of carbon in controlled lab incubations, and groundwater from FW305 well served as the microbial inoculum. Subsamples were harvested at several time points during a 50 day incubation for both chemical and microbial analyses. Results indicated a rapid decrease of total organic carbon within the first 1.5 days, concomitant to a rapid burst in CO<sub>2</sub>, and increased in cell numbers. 16S rRNA gene amplicon sequencing suggested that Massilia spp dominated in the original inoculum, and gradually decreased to below 5% after 50 days, while Azospirillum spp and Cupriavidus spp gradually increased from < 5% to 20-30%. Advanced chemical techniques including FTICR-MS and sXAS were used to characterize the C pool that included NOM metabolites and microbial byproducts produced during the incubation period. The molecular mass of C pool generally ranged from 200-600. Relative abundance of compounds with mass around 400-600 increased with incubation time and were considered to be produced from microbial activities. Also, the C pool shifted during incubation, the proportion of lignin in cultures increased, while proteins decreased. Further, GeoChip was used to identify the changes of microbial communities and expression of functional genes during transformation of the NOM.

In conclusion, the data clearly showed that microbial community present at Oak Ridge FRC responded to NOM from the site, and that the community shifted to reveal dominant members in sequence over time in response to transformation of the different functional groups in NOM.

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

# Multiplex characterization of microbial traits using dual barcoded genome fragment expression library in diverse bacteria

**V. K. Mutalik**<sup>1</sup> (vkmutalik@lbl.gov), P.S. Novichkov<sup>1</sup> S. Carim<sup>2</sup>, M. Callaghan<sup>1</sup>, H. Liu<sup>1</sup>, M. Price<sup>1</sup>, A.M. Deutschbauer<sup>1</sup>, A.P. Arkin<sup>1,2</sup> and Paul D.Adams<sup>1</sup>

Lawrence Berkeley National Lab; <sup>2</sup>University of California at Berkeley;

#### http://enigma.lbl.gov

**Project Goals:** This ENIGMA project aims to develop a large-scale high-throughput genefunction characterization technology that is scalable, quantitative, less laborious, cheaper than currently available technologies and allow multiplex trackable quantification of microbial fitness under hundreds of conditions in different organisms. This project also aims to overcome technological limitations of RB-TnSeq<sup>1</sup> in characterizing essential genes and genes with no significant phenotypes in diverse conditions. One of the key goals of this project is to support ENIGMA's vision in connecting genes to function space for microbial communities and link ENIGMA environmental parameters to single-gene phenotypes in the laboratory.

#### ABSTRACT

Advances in DNA sequencing have had a tremendous impact on genomic research. Thousands of microbial genomes, metagenomes and single cell genomes have been sequenced, bioinformatically studied and predictions have been made about architecture of biodiversity and in vivo potential of microorganisms. Among this vast number of sequenced microbes, only a handful of them have been experimentally characterized for functions and fitness traits (partly because of uncultivability of most microorganisms). Current state-of-the-art technologies such as transposon mutagenesis and shotgun expression libraries coupled with next generation sequencing have been helping to fill some of the knowledge gap in sequence-to-function space. In this context, the ENIGMA SFA is developing tools and technologies for discovering and interconnecting gene functions within individual microbes and microbial communities in the field. ENIGMA is developing high-throughput tools such as randomly barcoded transposon sequencing (RB-TnSeq<sup>1</sup>) for cultivable microorganisms to rapidly assay loss-of-function mutant phenotypes, but it remains a challenge to infer functions of essential genes in addition to cultivating and mutagenizing individual microbial isolates from the field. One complimentary approach is to generate shotgun expression libraries from diverse microbes to perform gain-offunction assays in model organisms, but it is a very low throughput method and labor intensive. There is an urgent need for high-throughput characterization technology that is scalable, quantitative, less laborious, cheaper and allow multiplex trackable quantification of microbial fitness under hundreds of conditions in different organisms. Such a technology will provide access to the unexplored sequence-function space of hundreds to thousands of microbial

genomes including the uncultivable ones and offer novel insights into functions encoded by microbial communities.

Here we develop a next generation technology **Dual Barcoded Shotgun Expression Library** Sequencing (Dub-SEQ) by combining shotgun expression library method with RB-TnSeq and BarSeq techniques developed within the ENIGMA SFA. Dub-SEQ uses dual random barcoded shotgun expression libraries on broad host plasmids and increases the throughput of functional screens by using barcode sequencing (Barseq) to assess the microbial fitness. The Dub-SEQ library can have the source DNA from diverse microbial/viral genomes, metagenomes and single cell genomes and can be heterologously expressed in diverse microbes for gain-of-function screens. Each library of different size donor genomic fragments are associated with dual barcodes only once and the microbial fitness conferred by heterologous DNA is quantified by simple BarSeq protocol after assaying in different conditions expressed from different bacteria.

We have implemented the Dub-SEQ methodology in *E coli* to validate different experimental and computational methods, and to prove the interpretability and the power of the method before moving on to other library hosts and genetic material sources. Here we present pooled fitness assays across 52 stress conditions (~total 200 assays) using the *E coli* Dub-SEQ library to demonstrate the scalability of our approach. These massive selections of gene functions are of relevance to ENIGMA mission including metal tolerance/resistance, nitrate/nitrite reduction, and utilization of different carbon and nitrogen sources. As these Dub-SEQ pooled fitness assays are obtained in conditions equivalent to those carried out with RB-TnSeq<sup>1</sup> technology developed within ENIGMA previously, we can directly compare the gain-of-function (Dub-SEQ) dataset to the loss-of-function dataset (RB-TnSeq<sup>1</sup>). Here we demonstrate that the technology is scalable, quantitative, less laborious and cheaper than current gain-of-function assays.

As Dub-SEQ technology can be applied to metagenomic DNA this approach represents a highthroughput platform to link ENIGMA environmental parameters to single-gene phenotypes in the laboratory and gain insights into functions encoded by microbial communities. In addition, as the Dub-SEQ libraries can be expressed in multiple hosts, it provides a seamless platform to screen gene-functions in multiple genetic backgrounds. We believe this technology will not only support ENIGMA's vision in connecting genes to function space for microbial communities, but also broadly applicable in building functional compendium of genome fragments from diverse single amplified genomes, viruses and metagenomes from diverse environments.

#### References

1. Wetmore KM, et al. (2015), mBio, 6(3):e00306-15.

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

### Stress mediates relative importance of deterministic and stochastic assembly in groundwater microbial communities

Daliang Ning<sup>1</sup>\*(ningdaliang@ou.edu), **Jizhong Zhou**,<sup>1</sup> Zhili He,<sup>1</sup> Ping Zhang,<sup>1</sup> Joy D. Van Nostrand,<sup>1</sup> Liyou Wu,<sup>1</sup> Renmao Tian,<sup>1</sup> Eric J. Alm,<sup>2</sup> Terry C. Hazen,<sup>3,4</sup> Dwayne Elias,<sup>3</sup> Matthew W. Fields,<sup>5</sup> Michael W. W. Adams,<sup>6</sup> Romy Chakraborty,<sup>7</sup> David Stahl,<sup>8</sup> Judy Wall,<sup>9</sup> Adam P. Arkin,<sup>7,10</sup> and **Paul D. Adams**<sup>7,10</sup>

<sup>1</sup>University of Oklahoma, Norman; <sup>2</sup>Massachusetts Institute of Technology, Cambridge; <sup>3</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>4</sup>University of Tennessee, Knoxville; <sup>5</sup>Montana State University, Bozeman; <sup>6</sup>University of Georgia, Athens; <sup>7</sup>Lawrence Berkeley National Lab, Berkeley, California; <sup>8</sup>University of Washington, Seattle; <sup>9</sup>University of Missouri, Columbia; <sup>10</sup>University of California at Berkeley, Berkeley.

#### http://enigma.lbl.gov/

Project Goals: Understanding the mechanisms controlling community diversity, distribution and succession is a central, but poorly understood, issue in ecology, particularly in microbial ecology. Although both stochastic and deterministic processes are believed to play roles in shaping community diversity and distribution, their relative importance is hotly debated. The importance of ecological stochasticity in shaping microbial community composition and structure is far less appreciated. Moreover, despite recent intensive studies on ecological community assembly, the factors mediating the relative importance of deterministic vs stochastic processes in shaping community composition and structure remain elusive. Thus, the major goal of this study is to illustrate the relative roles of deterministic and stochastic processes in shaping community structure and the factors controlling their relative importance.

To determine whether and how environmental factors mediate community assembly processes, about 100 wells representative of no or low, medium, high and extremely high stress were sampled and more than 200 environmental variables were measured. Null model analysis based on phylogenetic diversity of 16S rRNA gene revealed that the groundwater microbial communities at control or low stress wells without contamination were largely stochastic (~67%). As environmental stresses increased, the communities became less and less stochastic, with 41% of stochasticity at the extremely stressed wells. Also, quantitative analysis showed that variable selection (24~49%) and dispersal limitation (25~57%) played dominant roles while homogeneous selection (7.6~10%), homogeneous dispersal (0~1.6%), and undominated (or drift) (3.9~14%) play minor roles. Environmental stresses had strong positive correlation with variable selection (r=0.96), and negative correlations (r=-0.93) with dispersal limitation. Interestingly, drift (e.g. undominated) were higher at both low (8.6%) and extremely (14%) stressed wells than medium (4.7%) and high (3.9%) stressed wells. The spatial patterns of various processes were consistent with spatial distributions of various contaminants. In addition,

further statistical analyses indicated that selection on the groundwater microbial communities was largely imposed by heavy metals, oxygen, pH, and some anions (e.g., phosphate), whereas dispersal limitation was significantly correlated with geographic distance. Collectively, our results revealed stress mediates relative importance of deterministic and stochastic assembly in groundwater microbial communities.

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

#### Systems Biology Guided by Global Isotope Metabolomics

Erica M. Forsberg<sup>1,\*</sup> (<u>forsberg@scripps.edu</u>), Tao Huan,<sup>1</sup> Erica L. Majumder,<sup>2</sup> Judy W. Wall,<sup>2</sup> Michael P. Thorgersen,<sup>3</sup> Matthew Fields,<sup>4</sup> Michael W. Adams,<sup>3</sup> **Gary Siuzdak**,<sup>1</sup> Adam P. Arkin,<sup>4</sup> and Paul D. Adams<sup>4</sup>

<sup>1</sup>The Scripps Research Institute, La Jolla, California; <sup>2</sup>University of Missouri, Columbia; <sup>3</sup>University of Georgia, Athens; Montana State University, Bozeman; <sup>4</sup>Lawrence Berkeley National Lab;

#### http://enigma.lbl.gov; https://xcmsonline.scripps.edu

Project Goals: The goal of our ENIGMA research is to employ global metabolomics on microbial systems to better understand how they function, what molecular species they consume/produce and identifying metabolic pathways that are effected by various stressors. Examples of stressors include metal contamination and nitrate stress. Here we apply numerous bioinformatic tools to process raw liquid chromatography mass spectrometry data obtained from cellular extracts to identify statistically significant dysregulated metabolites and the pathways they are involved in. We also utilize an autonomous approach to validate metabolite identities through data dependent tandem mass spectral acquisition. Additionally, we employ a systems biology approach to cross reference metabolic data with genomic and proteomic data to look for changes that occur systems-wide. For pathways that are difficult to identify, our goal is to elucidate them with additional cell growth experimental protocols using stable isotope labeled substrates in both stress and non-stress conditions. Global isotope metabolomics is employed to identify these pathways in an unbiased manner. We have used these methods to elucidate nitrate assimilation pathways in *Psuedomonas* strains RCH2, N2A2 and N2E2.<sup>1</sup> We are currently performing similar experiments in a dual labeled system with *Bacillus cereus* ATCC 14579 to identify altered assimilation pathways using fully labeled <sup>13</sup>C lactate and <sup>15</sup>N nitrate under metal stress, and sulfur metabolomics using *Desulfovibrio vulgaris* Hildenburough (DvH) with <sup>34</sup>S labeled sulfate.

Abstract. Bioinformatics has become essential part of analyzing large global metabolomics datasets. The XCMS Online<sup>2</sup> platform significantly decreases the time required to process raw liquid chromatography mass spectrometry (LC-MS) data for retention time alignment, feature detection and statistical analysis of dysregulated features. In a typical metabolomics workflow, the accurate mass of each feature is matched with potential candidates from a database of metabolites. Data dependent tandem mass spectra are often used to validate these identities, which can be done using an autonomous workflow<sup>3</sup> or by manual interpretation of pooled quality control samples. This is followed by analysis of the metabolic pathways they are involved in to determine how they are affecting the whole system as a whole. We have recently developed a streamlined method to easily identify these aberrant pathways directly from the raw metabolomic data using a predictive pathway analysis time. Dysregulated pathways can be further understood with respect to upstream gene and protein expression processes by correlating genomic and

proteomic data, also in an automated approach<sup>5</sup>. Resulting overlaps can be easily visualized using the newly developed Pathway Cloud Plot, where the statistical significance (*p*-value) of the perturbed pathways are plotted versus the percent overlap of the identified dysregulated metabolites in the total identified metabolites of a given pathway. Additional information about the pathway size is indicated by the radius of the bubble. This novel cloud plot allows for an easy visual interpretation of perturbed metabolic pathways in the entire system. *Desulfovibrio vulgaris* Hildenborough was subjected to both nitrate stress and exposure to mercuric chloride and preliminary results indicate alterations in nitrate assimilation and sulfate reduction processes respectfully. Further analysis would be useful to gain more insight into how these processes are changing the system.

In some instances, metabolic pathways can be obscured by multiple enzymatic reactions that utilize the same substrates and/or yield the same end products. To elucidate these pathways, global isotope metabolomics can be employed using substrates that contain stable isotopes. Labeled starting materials are used to culture microbial strains, allowing them to be metabolized into the system until it reaches a steady state. Cell cultures must be quickly pelleted by centrifugation and flash frozen with liquid nitrogen to maintain the metabolic profile of the system without applying additional stress to prevent alterations from fast-acting enzymatic and signaling processes. The global isotope metabolome is tracked throughout a biological system by performing LC-MS on metabolite extracts and by isotope pattern ratio analysis of metabolite features between the labeled and unlabeled samples and between the stressed and unstressed samples. This comparative analysis provides information on energy consumption, biosynthesis and salvage processes that are not easily identified without looking at how the stable isotope is incorporated into the system. The major advantage of this approach is that it is unbiased and therefore able to detect novel processes that could not be achieved by looking at a targeted set of metabolites.<sup>1</sup> Here we demonstrate isotope analysis on *Bacillus cereus* exposed to a metal mixture using <sup>13</sup>C-lactate and <sup>15</sup>N-nitrate and preliminary results on DvH cultured with <sup>34</sup>Ssulfate in the presence of mercuric chloride.

#### References

- 1. Kurczy, M.E. et al. Global Isotope Metabolomics Reveals Adaptive Strategies for Nitrogen Assimilation. *ACS Chemical Biology* **11**, 1677-1685 (2016).
- 2. Tautenhahn, R., Patti, G.J., Rinehart, D. & Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **84**, 5035-5039 (2012).
- 3. Benton, H.P. et al. Autonomous metabolomics for rapid metabolite identification in global profiling. *Analytical chemistry* **87**, 884-891 (2014).
- 4. Li, S.Z. et al. Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput. Biol.* 9, 11 (2013).
- Huan, T.F., E.M.; Rinehart, D.; Johnson, C.H.; Ivanisevic, J.; Benton, H.P.; Fang, M.; Aisporna, A.; Hilmers, B.; Poole, F.L.; Thorgersen, M.P.; Adams, M.W.W.; Krantz, G.; Fields, M.W.; Robbins, P.D.; Niedernhofer, L.J.; Ideker, T.; Majumder, E.L.; Wall, J.D.; Rattray, N.J.W.; Goodacre, R.; Lairson, L.; Siuzdak, G. A Cloud-Based Systems Biology Platform Guided by Metabolomics. *Nature Methods* In Review (2017).

Funding: ENIGMA, Scientific Focus Area Program at Lawrence Berkeley National Laboratory for the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under contract number DE-AC02-05CH11231, and the National Institutes of Health.

#### Temporal Variability and Microbial Activity in Groundwater Ecosystems

Heidi J. Smith<sup>1</sup>\*(heidi.smith@montana.edu), Anna Zelaya<sup>1</sup>, Isaac Miller<sup>1</sup>, Dominique Joyner<sup>2</sup>, Terry C. Hazen<sup>2</sup>, **Matthew W. Fields**<sup>1</sup>, Adam P. Arkin<sup>3</sup>, and Paul D. Adams<sup>3</sup>

<sup>1</sup>Montana State University, Bozeman; <sup>2</sup>University of Tennessee, Knoxville; <sup>3</sup>Lawrence Berkeley National Laboratory, Berkeley CA

#### http://enigma.lbl.gov/

**Project Goals:** A fundamental goal in the field of microbial ecology is to link the activity and structure of microbial populations and communities to processes occurring within an ecosystem. This project aims to identify the drivers of community structure and succession by identifying the metabolically active fraction of microbial communities from both pristine and contaminated groundwater habitats at the Field Research Center (FRC) at Oak Ridge National Laboratory (ORNL). It is hypothesized that community function is independent of phylogeny and that functionality will be altered as a result of environmental perturbations. The use of geochemically distinct wells in combination with the enumeration and sequencing of translationally-active microorganisms, activity assays, carbon utilization profiles, and geochemical measurements will allow for the elucidation of the mechanisms shaping community structure and function in terms of turnover of natural organic matter and major contaminants (*e.g.*, NO<sub>3</sub><sup>-</sup>). Multiple assay comparisons will be used to achieve an accurate characterization of the active fraction of groundwater microbial communities and will ultimately be applied to continuous sediment cores from both contaminated and pristine locations.

Abstract: Saturated subsurface environments are estimated to contain approximately 40% of the prokaryotic biomass on Earth, and due to the complexity of these habitats they support highly diverse microbial communities. In addition, it is estimated that over 98% of the Earth's consumable and available freshwater is in the subsurface as groundwater. However, the factors that determine microbial community assembly, structure, and function in groundwater systems and the impact on water quality and contaminant transport remain poorly understood. Three non-contaminated background wells were sampled for groundwater geochemistry and microbial diversity approximately 3 times a week over a period of three months. Community analysis via ss-rRNA paired-end sequencing and distribution-based clustering revealed temporal differences in richness, diversity, and variability in the groundwater communities. Microbial community composition of a given well was on average >50% dissimilar to any other well at a given time point. Similarities in community structure across wells were observed with respect to the presence of 20 cosmopolitan populations in all samples in all wells; however, wells differed in the relative abundances of these taxa. Similarity percentage (SIMPER) analysis revealed that temporal variability was explained by lowly abundant and transient populations or more highly abundant and frequently present taxa in a sample-dependent manner.

Based on the observations of temporally and spatially variable groundwater communities, we aim to use a combination of methodological approaches to determine how functional groups of microorganisms relate to habitat as well as structure/composition heterogeneity. Specifically, we identified the active fraction of microbial communities and quantified rates of activity from both contaminated and pristine groundwater habitats. Total cell abundances, quantification of translationally-active microorganisms, <sup>3</sup>H-leucine incorporation, carbon utilization profiles, and community sequencing of active microorganisms (SSU rRNA and sorted translationally-active cells) were used to investigate four groundwater wells representing geochemical extremes. Two previously studied background wells and two contaminated wells were sampled November 2016 – January 2017. Contaminants in non-background wells include: radionuclides (U, Sr, and Tc), metals (Sr, Cd, Ba, B, Hg, Cr), volatile organic contaminants (VOCs), and nitrate. Of particular interest for this study is the U-nitrate-pH gradient present within contaminated groundwater wells.

Bioorthogonal non-canonical amino acid tagging (BONCAT) was used to identify translationally-active bacterial and archaeal cells for microscopic evaluation and sequencing. Incubations with additions of amino acid L-azidohomoalanine (AHA) followed with fluorescent tagging of AHA-containing cellular proteins can identify newly synthesized proteins. For contaminated wells, large proportions of the community were identified as translationally-active; however, specific rates of activity were low based on labeled amino acid incorporation. Total cell abundances ranged from  $1.11-2.07 \times 10^5$  cells/mL with 73-84% of the community being translationally active. Rates of <sup>3</sup>H-leucine incorporation for the two contaminated groundwater samples were  $14.5\pm 2.4$  ng C/d and  $45.1\pm 7.4$  ng C/d, respectively.

The objectives of this study are part of a larger program that aims to understand the geospatial relationships between hydrogeology, geochemistry, and microbiology. Ultimately, we aim to apply these methodologies to attached biofilm communities from continuous sediment cores spanning depths from 0-50 m. It is known that free living microbial assemblages are substantially different than attached sediment populations. Due to sampling constraints, groundwater has been more routinely studied as it is difficult to obtain representative sediment samples. Studies that have evaluated species diversity in sediment boreholes have observed significant variation; however, it is not known how species variation relates to variation in function or activity. Additionally, *in situ* measurements from this study will aid in the development a predictive framework for understanding large scale biogeochemical cycling from groundwater environments.

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

### Title: The role of adaptive evolution in shaping the structure and function of model microbial consortia

N. Stopnisek<sup>1</sup>, N.R. Elliott<sup>1</sup>, K.L. Hillesland<sup>2</sup>, P.J. Walian<sup>3</sup>, J. Zhou<sup>4</sup>, L. Wu<sup>4</sup>, A.W. Thompson<sup>5</sup>, S. Turkarlsan<sup>5</sup>, N.S. Baliga<sup>5</sup>, **D.A. Stahl**<sup>1</sup>\*(<u>dastahl@uw.edu</u>), Adam P. Arkin<sup>3</sup> and Paul D. Adams<sup>3</sup>

<sup>1</sup>University of Washington, Seattle, WA; <sup>2</sup>University of Washington, Bothell, WA; <sup>3</sup>Lawrence Berkeley National Lab, Berkeley, CA; <sup>4</sup>University of Oklahoma, Norman, OK and <sup>5</sup>Institute of System Biology, Seattle, WA

#### http://enigma.lbl.gov/

Project Goals: Laboratory based studies of nascent and evolved model microbial consortia are used to develop a predictive - rather than an observational – understanding of environmental processes. The general aim of these studies is to examine the role of adaptive evolution in determining community stability and productivity. Specific project aims include the identification of mutations correlated with improvements in community efficiency (as measured by conversion of substrates to microbial biomass) and community stability, and the determination of the contribution of individual mutations to community efficiency and stability. The project is part of a large research effort within ENIGMA consortium, entitled 'Evolution and Assembly of Communities', which aims to understand the impact of adaptive evolution on system-level processes.

**Abstract:** Syntrophy is a widespread relationship between microbes that plays a central role in the decomposition of carbon in anoxic environments. In these interactions, bacteria ferment organic acids to produce hydrogen, formate or acetate, which are consumed by hydrogeno- and aceto-trophic methanogens. Removal of the byproducts of the fermentation benefits bacteria because the free energy ( $\Delta G$ ) decreases with decreasing concentrations of the byproducts. A number of bacterial species seem to be specialized to ferment organic acids in syntrophic associations with hydrogenotrophic species. The question is how and why this specialization evolved? The question was addressed by examining the evolution of experimentally established obligate syntrophy between the sulfate reducing bacterium Desulfovibrio vulgaris Hildenborough (DvH) and the archaeon Methanococcus maripaludis (Mmp) (1). The cocultures evolution was followed over 1000 generations in two environments, one allowing for the heterogeneous distribution of resources and the other not. Evolved cocultures showed significant improvements in stability, growth yields and growth rates by the 300<sup>th</sup> generation, suggesting a capacity for rapid adaptive evolution of both partners (1). Genome resequencing of 1000<sup>th</sup> generation evolution lines revealed that the loss of a capacity for sulfate respiration, due to lossof-function mutations in the aps and sat genes, was a common evolutionary outcome for 13 out of 21 replicate evolution lines (2). Genome resequencing also revealed the emergence and coexistence of distinct evolved strains of DvH and Mmp, all apparently contributing to

significant improvements in growth of the cocultures (2). These observations led to further questions such as: i) what combinations of mutations are responsible for the observed improvements? ii) at what stage of the evolution did these mutations arise? and iii) what are the benefits of these mutations to the individual and to the community?

To address the first two questions, we sequenced the genomes of 72 clonal isolates from three 1000 generation evolution lines and obtained metagenomes from the earlier time points in the evolution (prior to the 1000<sup>th</sup> generation). The analysis of mutations in clonal isolates showed that all strains were enriched in mutations in genes related to signal transduction, energy production/conversion, and nutrient transport. Interestingly, all 36 DvH clonal isolates represent unique genotypes and only few common mutations were found. Hypothetical porin (DVU0799), UTP-glucose-1-phosphate uridylyltransferase (galU, DVU1283) and long-chain fatty acid transporter protein (DVU1260) were among the few genes mutated in all clonal isolates. Physiological and biochemical characterization of these mutated genes suggested they contribute to improved inter-species cooperation by facilitating the direct transfer of nutrients/metabolites by promoting aggregation and increasing membrane permeability. Mutations in regulatory genes (transcript regulators and histidine kinases) associated with energy metabolism, lactate utilization and motility suggested that functional modification of these central processes also played important roles in adaptive evolution. For example, mutations in Mmp clonal isolates were concentrated in amino acid transport systems (MMP0166, MMP0419 and MMP1511) and in the regulator of archaellum (MMP1718).

Genotypic information is now being complemented by detailed phenotypic characterization of earlier time points in evolution and of selected clonal isolates grown in mono- and co- culture. Preliminary analysis of metagenomes from the earlier times of evolution suggests an early emergence of common mutations, followed by rapid and repeated gain and loss of mutations as these simple communities evolved to become more stable and productive. Improvement in co-culture growth was significant even if the pairings were from clones isolated from different evolution lines. Notably, growth of evolved DvH clonal isolates in monoculture also showed significant improvements in yield when grown in the same medium, suggesting that one common adaptive feature of adaptive evolution involved changes in maintenance energy requirements.

#### References

- 1. Hillesland, KL; Stahl, DA. PNAS, 10.1073/pnas.0908456107
- Hillesland, KL; Lim, S; Flowers, JJ; Turkarslan, S; Pinel, N; Zane, GM; Elliott, N; Qin, Y; Wu, L; Baliga, NS; Zhou, J; Wall, JD; Stahl, DA. *PNAS*, 10.1073/pnas.1407986111

ENIGMA (<u>http://enigma.lbl.gov</u>) at LBNL supported by Office of Biological and Environmental Research US Dept of Energy Contract No: DE-AC02-05CH11231
## Understanding the thermodynamic Foundations of microbial Growth Efficiencies in the Lab and Field

Frederick von Netzer<sup>1\*</sup> (fcrvn@uw.edu), Drew Gorman-Lewis<sup>1</sup>, Everett Shock<sup>2</sup>, Serdar Turkarslan<sup>3</sup>, Christina E. Arens<sup>3</sup>, Anne W. Thompson<sup>3</sup>, Nitin S. Baliga<sup>3</sup>, Aifen Zhou<sup>4</sup>, Jizhong Zhou<sup>4</sup>, Alexander Aaring<sup>5</sup>, Romy Chakraborty<sup>5</sup>, Ji Won Moon<sup>6</sup>, Dwayne Elias<sup>6</sup>, Dominique C. Joyner<sup>6,7</sup> Terry C. Hazen<sup>6,7</sup>, Heidi Smith<sup>8</sup>, Matthew Fields<sup>8</sup>, Farris Poole<sup>9</sup>, Michael W. W. Adams<sup>9</sup>, Hans Carlson<sup>5</sup>, Adam Deutschbauer<sup>5</sup>, David Vuono<sup>10</sup>, Kelley Meinhardt<sup>1</sup>, **David A. Stahl**<sup>1</sup>, Adam P. Arkin<sup>5</sup> and Paul D. Adams<sup>5</sup>

<sup>1</sup>1University of Washington, Seattle; <sup>2</sup>Arizona State University, Tempe; <sup>3</sup>Institute for Systems Biology, Seattle; <sup>4</sup>University of Oklahoma, Norman; <sup>5</sup>Lawrence Berkeley National Laboratory, Berkeley; <sup>6</sup>Oak Ridge National Laboratory, Oak Ridge; <sup>7</sup>University of Tennessee, Knoxville, <sup>8</sup>Montana State University, Bozeman; <sup>9</sup>University of Georgia, Athens; and <sup>10</sup>Desert Research Institute, Reno

#### http://enigma.lbl.gov

Project Goals: A key element of microbial growth and therefore microbial community assembly is how microbes partition the available resources between energy required for maintenance and growth. Using microcalorimetry and thermodynamic modelling, we gained a quantitative proxy for microbial growth efficiency under different growth conditions. As the next step, we aim to apply our methods to address the ecological framework guiding the partitioning of denitrification pathways at the Oak Ridge National Laboratory Field Research Site (FRC); through the establishment of microbial activity assays, isotope fractionation analysis as well as establishment of mass balances and stoichiometry for representative nitrate respiring isolates.

The assembly of microbial communities is determined by many factors, with the environment setting the stage via availability of electron acceptors, donors and carbon sources as well as with physical/chemical parameters such as temperature and salinity. In this framework microbes have to adapt to either stable or dynamic conditions and to either compete or share resources for survival. Also, bacteria need to balance constantly the division of available energy between maintenance of basic cellular functions and growth. Therefore, microbes with the most favorable ratio between maintenance and growth requirements should be more competitive compared to microbes with higher energy demands for maintenance and growth.

Maintenance energy levels as a proxy for microbial competitiveness are usually measured in chemostats near zero growth conditions. Here, we attempt to capture maintenance energies via microcalorimetry and a metabolite-dependent thermodynamic model as a quantitative proxy for microbial growth efficiency. Microcalorimetry offers a direct and highly sensitive method to assess the enthalpy-related terms of microbial growth in relation to the potential energy supplied by growth substrates. This in turn allows for a quantitative description of growth and maintenance in thermodynamic terms under different growth conditions, based on the comparison of the metabolite profile at the start and end of growth.

We have examined the influence of temperature stress, simulated environmental dynamics and adaptation to salt stress on the growth efficiencies of different strains of *Desulfovibrio vulgaris* Hildenborough and *D. alaskensis* G20. These analyses quantified the cost of maintenance (survival) in relationship to increasingly suboptimal growth conditions, the cost of regulation in a fluctuating environment and the reduction in maintenance costs realized through adaptive evolution.

We aim to apply our growth efficiency and maintenance proxy to field conditions by addressing the ecological framework guiding the partitioning of denitrification pathways at the Oak Ridge National Laboratory Field Research Site (FRC). This site has a long contamination history with nitrate among a large variety of other contaminants. Thermodynamic modeling based on mass balances and stoichiometries of site-relevant nitrate respiring isolates will be associated with in situ measurements of nitrate respiration and nitrate stable isotope fractionation analyses to develop a predictive framework for microbial community assembly and activity.

As the next step, we are cultivating available nitrate respiring isolates (*Rhodanobacter, Aci-dovorax* and *Pseudomonas* spp.) under different C/N ratios and the application of the thermody-namic model established with metabolite profiles, we aim to record the energy requirements under different biogeochemical conditions. These profiles can then out into context to actual biogeochemical field conditions as characterized by nitrate specific acetylene-block activity assays and the metabolic history as traced by nitrate stable isotope fractionation.

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

## Quantitative Measurements of Cellulase Display in the Model Gram+ Microbe *Bacillus subtilis* Define Determinants Required for Enzyme Display and Activity

Grace L. Huang,<sup>1,2\*</sup> (gLHuang@chem.ucla.edu), Ye Seong Kim,<sup>1</sup> and Robert T. Clubb<sup>1-3</sup>

<sup>1</sup>Department of Chemistry and Biochemistry; University of California – Los Angeles, Los Angeles, CA, USA. <sup>2</sup>UCLA-DOE Institute of Genomics and Proteomics; University of California – Los Angeles, Los Angeles, CA, USA. <sup>3</sup>Molecular Biology Institute; University of California – Los Angeles, Los Angeles, CA, USA.

#### http://www.biochemistry.ucla.edu/biochem/Faculty/Clubb/Clubb Lab/Main.html

#### **Project Goals:**

Lignocellulose is a promising feedstock from which to sustainably produce biocommodities, but its recalcitrance to hydrolysis limits its use. One strategy to overcome this problem is to use consolidated bioprocessing (CBP) microbes to directly convert biomass into biochemicals [1]. We seek to develop a robust heterologous saccharolytic enzyme display system to give Gram+ bacteria cellulolytic activity and enable their use in CBP [2]. Here we discuss our efforts to define the determinants that affect heterologous surface protein display in *B. subtilis*, an industrially used model Gram+ microbe. We report quantitative measurements of cellulase display via the noncovalent LysM module. Utilizing five parent strains, we show that genetic elimination of eight proteases improves display levels. However, protein display is inefficient with only 28% of secreted cellulases bound to the cell surface, at much lower numbers than when cellulases are added *ex vivo*. Also, *B. subtilis* appears to shed cell wall enzymes as it enters stationary phase, perhaps caused by cell wall turnover. These quantitative measurements form the foundation for genetic engineering of *B. subtilis* for increased stability of displayed protein systems.

#### **References:**

- 1. Lynd, L.R., et al., *Consolidated bioprocessing of cellulosic biomass: an update.* Curr Opin Biotechnol, 2005. **16**(5): p. 577-83.
- 2. Huang, G.L. and R.T. Clubb, *Progress Towards Engineering Microbial Surfaces to Degrade Biomass*, in *Biomass*, J. Tumuluru, Editor. 2016, InTech. In press.

This work was supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under award number DEFC02-02ER63421.

## Approaches for Evaluating the Production Potential of High Volume Products in Microbial Systems

Matthew R. Long<sup>1,2</sup>\* (mrlong2@wisc.edu), Tony [WenZhao] Wu<sup>1,2</sup>, Christos T. Maravelias<sup>1,2</sup>, and Jennifer L. Reed<sup>1,2</sup>

<sup>1</sup>Department of Chemical and Biological Engineering, UW-Madison, Madison, WI 53706 <sup>2</sup>DOE Great Lakes Bioenergy Research Center, Madison, WI 53706

Project Goals: Metabolic models can be used to engineer biofuel production strains, where models can find bottlenecks in metabolic pathways, identify important regulatory interactions, and suggest perturbations to force a microorganism to produce or utilize more of a compound of interest. Metabolic modeling will be used to improve conversion of sugars derived from lignocellulosic biomass into specialty biofuels using different microbial platforms, including *Saccharomyces cerevisiae* and *Zymomonas mobilis*. Here we compared different chemical products (including biofuels) by evaluating their chemical production potential based on model-predicted maximum theoretical yields and productivities.

In order to identify the most promising candidates for next-generation biofuels and bioproducts, it is critical to evaluate the production potential of different chemical products in a variety of host organisms. Genome-scale metabolic models offer the ability to efficiently screen a wide variety of chemicals for production in different biological systems; however, traditional constraint-based modeling approaches have focused on calculating the maximum theoretical yields. This value often occurs at an unacceptably low biomass growth rate which results in low productivities. In order to alleviate this problem, a new approach has been developed which instead identifies the maximum theoretical productivity of a chemical production strain in a reactor. The productivity depends upon the type of reactor (*e.g.* chemostat or batch) and the mode of operation (*e.g.* continuous or induced).

This productivity analysis was applied to a large database of High Production Volume (HPV) chemicals totaling more than 3,500 chemicals. Productivity was evaluated in both *Escherichia coli* and *Saccharomyces cerevisiae* with the inclusion of heterologous reactions for non-native metabolites from reaction databases (*e.g.* KEGG and MetaCyc). This analysis shows differences between the maximum possible production rate and the maximum potential productivity. These differences could affect the screening criteria which would otherwise over- or under-estimate the true chemical production potential for different chemical products. Furthermore, the productivity captures differences in reactor types and conditions, allowing for development of strains in the most productive system. Further methods for identifying genetic strategies (*i.e.* growth coupling via gene knockouts) for achieving the maximum potential productivity have also been developed and can be used to design strains which achieve high productivities.

*Funding statement*. This work was funded by the U.S. Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02- 07ER64494).

## Functional analysis of copper and silver storage sites and their role in metal homeostasis in *Chlamydomonas*

Kristen Holbrook\*<sup>1</sup> (kholbrook@chem.ucla.edu), Daniela Strenkert,<sup>1</sup> Bonnie Chu Brown,<sup>1</sup> Stefan Schmollinger,<sup>1</sup> Jennifer Pett-Ridge,<sup>2</sup> Carolyn Larabell<sup>3</sup> and **Sabeeha Merchant**<sup>1</sup>

<sup>1</sup> Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA, <sup>2</sup>Lawrence Livermore National Laboratory, Livermore, CA, <sup>3</sup>Lawrence Berkeley National Laboratory, Berkeley, CA

Project Goals: Processes regulating metal bioavailability and homeostasis impact cellular functions in all forms of life and their dysfunction contributes to reduced primary productivity as well as disease in livestock and humans. This project seeks to understand the mechanisms of Cu homeostasis and how its toxic analog Ag influences these processes in the model alga *Chlamydomonas reinhardtii*. We will investigate factors that control Cu storage and bioavailability. In so doing, this project will contribute broadly to the understanding of metal metabolism in the context of photosynthesis as well as the environmental impact of heavy metals.

Copper  $(Cu^{+})$  is an essential cofactor required for protein function in various cellular pathways, ranging from DNA synthesis to respiration. However,  $Cu^+$  uptake and distribution is highly regulated because (1) the reactivity of Cu<sup>+</sup>, which makes it useful in biology, can also lead to cytotoxicity, (2) the interplay between pathways for metabolizing  $Cu^+$  and other essential metals has the potential for broad disruption of metabolism if even one metal is present at too low or high a level and (3) pathways required for the uptake of  $Cu^+$  can also lead to cellular accumulation of toxic metals in the environment. The advent of analytical techniques has revealed considerable detail concerning Cu metabolism, but mechanistic and regulatory details of intracellular distribution and storage of Cu<sup>+</sup> are under-investigated. The Merchant laboratory has developed the green alga Chlamydomonas reinhardtii as a powerful eukaryotic reference organism for fundamental discovery in the field of trace metal homeostasis, including mechanisms of elemental sparing (in face of deficiency or other stress), pathways of metal assimilation and distribution (especially for Cu and Fe), and most recently, the visualization of intracellular compartmentalization sites for Cu, Fe and Mn. These storage sites are dynamic: they sequester excess metals to prevent cytotoxicity but the essential elements are bio-available upon transfer of the organism to a situation of deficiency. During Zn or Cu limitation, Chlamydomonas accumulates Ag, a process that is influenced by expression of a Cu transporter, CTR2. The goal of this project is to understand how nutritionally essential  $Cu^+$  and its toxic analog  $Ag^+$  are sequestered and stabilized and how the metals are trafficked into and out of intracellular compartments.

Supported by a cooperative agreement with the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421.

## Design, Characterization, and Emerging Applications of Highly Symmetric Protein Nanostructures

Kevin Cannon<sup>1,2</sup>\* (kacannon@ucla.edu), Josh Laniado,<sup>1,2</sup> Yuxi Liu,<sup>1,2</sup> Julien Jorda,<sup>1,2</sup> Brendan Ahmer,<sup>1,2</sup> **Robert Clubb**,<sup>1,2</sup> and **Todd Yeates**<sup>1,2</sup>

<sup>1</sup>UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA; <sup>2</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, CA

http://yeateslab.mbi.ucla.edu/

Project Goals: A primary goal of our work is to develop robust strategies for building novel materials and nanoscale structures using protein molecules as the building blocks. Several recent successes have been demonstrated in creating novel protein cages and ordered 2-D arrays. Our ongoing efforts are focused on improving design strategies to increase the currently low success rates, and on exploring applications of the designed protein materials in energy-related applications, including organizing sequentially acting enzymes. Current efforts toward achieving these goals will be described.

Motivated by the rich diversity of protein molecules that have evolved by nature to form complex and highly symmetric supramolecular structures, recent engineering efforts in the field of protein design have exploited symmetry to create novel self-assembling protein structures of types unseen in biology. Such materials range from finite cages or shells to essentially unbounded two-dimensional and three-dimensional arrays (i.e. crystals) to linear or tubular filaments. Recent work in the Yeates laboratory and in collaborating groups has led to the successful design and characterization of several cage structures exhibiting tetrahedral, octahedral, and even icosahedral symmetries which show close agreement with design specifications. For certain design approaches, a complete list of allowable symmetry combinations that can be used for construction has been articulated, opening a path for creating a rich diversity of geometrically defined protein materials.

Multiple new lines of work are addressing current design challenges. Some of those efforts aim to improve upon the helix fusion strategy of cage design, wherein two naturally oligomeric proteins are fused together by an alpha helical linker at a specified geometry. New design variations are focused on reducing the polymorphism and flexibility of designed structures by optimizing design parameters to create more compact and rigid structures, especially to allow higher order symmetries such as icosahedral to be reliably achieved. Projects toward this goal include utilizing chemical cross-linkers to staple the alpha-helical fusion into a strictly defined geometry, as well as replacing the single helix linker with a coiled coil to make the structures more robust. Promising candidates from these approaches are currently undergoing characterization.

Further work in the lab has been inspired by the large microcompartment shells that exist in many bacterial cells. Our recent experiments have led to the formation of a novel dodecahedral nanocage made from one redesigned microcompartment shell protein. This result has inspired efforts to utilize these shell proteins for further cage designs as well as for gaining further insight into the evolution of symmetric proteins in general.

Following the exciting successes of protein nanoarchitectural designs in the last few years, the lab has also begun to turn its attention to the ways in which these protein nanomaterials can be endowed with properties useful for applications in materials science, energy and medicine. Ongoing efforts include decorating the outside of previously characterized cages with new amino acid sequences that correspond to bioactive peptides or enzymatic recognition tags. Our recent experiments have shown success in using sortase to covalently attach other proteins to the outside of a tetrahedral cage by incorporating a short amino acid sequence that is recognized by sortase at the C-terminal end of one of the cage's component proteins. The success of this method could make it possible to attach a wide range of other proteins to the outside of cages, and the availability of up to 4 unique termini exposed on the surface could lead to applications where multiple enzymes that act in sequence are attached to the same cage in order to increase the pathway flux.

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421.

## Development of metabolomic methods for investigating metabolic regulation of the MEP pathway in *Zymomonas mobilis* and *Escherichia coli*

Julia I. Martien,<sup>1,2</sup> David M. Stevenson,<sup>1,2</sup> Nagendra Kurumbang,<sup>1,2</sup> Robert Landick,<sup>1,2</sup> **Daniel Amador-Noguez**<sup>1,2</sup>\* (amadornoguez@wisc.edu)

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

Project Goals: The mission of the Great Lakes Bioenergy Research Center is to perform research that generates technology to convert cellulosic biomass to advanced biofuels. In alignment with this goal, we have developed a method to examine the metabolic regulation of the MEP pathway using metabolite quantification with LC-MS. Understanding the metabolic response of the MEP pathway to environmental and genetic changes will help to guide targeted metabolic engineering for high-yield production of advanced biofuels such as isoprene and isoprene-derived alcohols.

The methyl-erythritol-phosphate (MEP) pathway responsible for producing isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) in most bacteria (Fig 1). These metabolites play an integral role in cellular metabolism as they are the building-blocks for biosynthesis of isoprenoids, a diverse class of compounds encompassing quinones, carotenoids, hopanoids, and various secondary compounds. Additionally, IPP and DMAPP are direct precursors to isoprene, a high-value chemical which can be used for the production of synthetic polymers, or converted to high-grade fuels such as methylbutenol. In recent years, efforts have been made to metabolically engineer bacteria for over-production of isoprene and isoprene-derived alcohols via the MEP pathway. However, directed engineering has been limited by an incomplete understanding of *in vivo* regulation of the MEP pathway. Our work aims to understand the metabolic regulation of the MEP pathway in both *E. coli* and the emerging biofuel-producer *Zymomonas mobilis*. We utilize a metabolomic approach to monitor metabolic changes in the MEP pathway, and connected pathways of central metabolism, in response to environmental and genetic perturbations.

In order to interrogate the metabolic regulation of the MEP pathway, we have developed a method using LC-MS to quantify all eight intermediates of the MEP pathway: 1-deoxy-Dxylulose 5-phosphate (DXP), 2-C-methyl-D-erythritol 4-phosphate (MEP), cytidine diphosphatemethlyerythritol (CDP-ME), cytidine diphosphate-methylerythritol 2-phosphate (CDP-MEP), methylerythritol 2,4-cyclodiphosphate (MEcDP), hydroxyl-methrylbutanol diphosphate (HMBDP), IPP and DMAPP as well as downstream products such as geranyl pyrophosphate (GPP) and farnesyl pyrophosphate (FPP) (Fig 1). This method also allows for quantification of a broad range of central metabolites including intermediates of glycolysis, the pentose phosphate pathway, and the TCA cycle as well as amino acids, nucleotides, and electron carriers such as NAD(P)H. With this method, it is possible to monitor depletion and accumulation of intermediates of the MEP pathway in response to changes in nutrient conditions or abiotic stressors, and to identify interactions between the MEP pathway and the rest of central metabolism. To date, we have examined the effects of growth stage, nitrogen depletion, aerobic growth, and exposure to ethanol on the MEP pathway in the context of central metabolism. We have observed a dramatic effect of exposure to oxygen on the abundance of intermediates in the MEP pathway in *Z. mobilis*, indicative of a bottleneck at the final two reductive steps (Fig 1).

In addition to environmental perturbations, we are interested in examining the metabolic consequences of genetic changes to the MEP pathway. With our LC-MS method and engineered strains of *E. coli* and *Z. mobilis* generated by the Landick Lab, we are able to quantify the effects of over-expressing or silencing genes of the MEP pathway on intermediate metabolites of the pathway, as well as central carbon metabolism. To date, we have performed metabolomic analysis on *E. coli* and *Z. mobilis* strains containing plasmids that overexpress the first five genes of the MEP pathway, ending with the production of MEcDP, and have observed a significant increase in the abundance of MEcDP upon induction of the plasmids. This overexpression is accompanied by changes in central metabolism that may reveal key carbon, energy, and redox inputs that limit the activity of the MEP pathway.



Figure 1. The seven enzymatic steps of the MEP pathway.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02- 07ER64494).

#### **Optimizing MLG Production for Improved Biofuel Crop**

Sang-Jin Kim <sup>1,2,§</sup>, Starla Zemelis-Durfee <sup>1,2,§</sup>\*(<u>zemeliss@msu.edu</u>), Jacob Jensen <sup>1,4,§</sup>, Curtis Wilkerson <sup>1,3,4</sup>, and **Federica Brandizzi** <sup>1,2</sup>

<sup>1</sup> Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI 48824; <sup>2</sup> MSU-DOE Plant Research Lab, Michigan State University, East Lansing, MI 48824; <sup>3</sup> Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824; and <sup>4</sup> Department of Plant Biology, Michigan State University, East Lansing, MI 48824

<sup>\$</sup> These authors contributed equally to this work

\* Presenting Author

https://www.glbrc.org/research/plants

http://brandizzilab.prl.msu.edu/

#### **Project Goals:**

The goal of this project is to improve biomass feedstock by increasing the levels of mixedlinkage glucan (MLG), a non-cellulosic polysaccharide, in biofuel crops. Since MLG is composed of only glucose, we expect to generate plants with an improved C6 to C5 ratio that would be more easily fermentable by yeast, consequently generating an enhanced biofuel feedstock. MLG is synthesized by CSLF and CSLH proteins. In previous studies increasing the levels of MLG in barley plants using a constitutive promoter led to growth penalties, demonstrating the choice of promoter and specific tissue location to overexpress the genes of interest in storage vegetative tissues is important to obtain healthy MLG-rich plants. To overcome the growth penalty caused by the accumulation of MLG we plan to optimize the expression of CSFL6 and modify stress response signaling pathways, allowing plants to adapt to MLG overproduction while maintaining healthy plant performance. This research is promising to advance significantly the digestibility of biofuel feedstock produced from grasses.

#### Abstract:

The focus of our work is to improve biofuel grasses with the fewest inputs resulting in desirable bioenergy feedstock at a low cost. For improved biofuels we aim to increase the amount of a cell wall component, mixed-linkage glucan (MLG). MLG is an important cell wall polysaccharide containing  $\beta$ -D-glucosyl residues with both (1,3) and (1,4) linkages making up 20% of cell wall

in grasses<sup>5</sup>. To do this we must understand how MLG is synthesized by CSLF and CSLH proteins<sup>2,3</sup>. Within the CSLF family, CSLF6 has been reported to be highly expressed in the model grass Brachypodium and is exclusively found in monocotyledons<sup>3</sup>. From our RNA seq data we also found that CSLF6 is highly expressed in early endosperm development of Brachypodium when MLG synthesis is most active<sup>4</sup>. Using Brachypodium, we have characterized a CSLF6 homologue, BdCSLF6, and established its functional requirements in vitro and in vivo<sup>5</sup>. Expressing our YFP tagged BdCSLF6 in tobacco, a species that does not produce MLG, we established that BdCSFL6 is localized to the Golgi membranes and is capable of producing MLG. We have also shown production of MLG when BdCSLF6 is expressed in Pichia, which supports that BdCSLF6 alone produces both  $\beta$ -(1:3, 1:4) linkages of MLG. In addition, we also have confirmed the protein topology of BdCSLF6 and proven that the catalytic domain of the enzyme is exposed to the cytosol, supporting that MLG is secreted into the Golgi compartment<sup>5</sup>. By analyzing our transgenic Brachypodium over-expressing YFP fused CSLF6 by a constitutive promoter, we also observed BdCSLF6 is localized to the Golgi but also to the endoplasmic reticulum (ER)<sup>5</sup>. These plants display an increased amount of MLG (~30%) when compared to WT (Bd21-3), yet have a stunted growth phenotype. To avoid growth penalties, we have identified several tissue and developmental specific promoters through RNA seq analysis from Brachypodium and confirmed by RNA *in situ* hybridization and GUS reporter system<sup>1,6,7,8</sup>. The tissue specific promoters selected will drive expression of BdCSLF6 in pith parenchyma cells that are functionally capable of holding increased amounts of MLG. In addition, we plan to modify stress response signaling pathway genes in these BdCSLF6 lines allowing the plants to cope better to the accumulation of MLG. We anticipate that engineering plants to contain increased amounts of MLG in tissue amenable to store MLG will allow for an easily digestible biofuel feedstock without compromising growth.

#### References

1. Anna Karlgren et al. (2009) Non-radioactive *in situ* Hybridization Protocol Applicable for Norway Spruce and a Range of Plant Species. Journal of Visualized Experiments.

2. Burton RA, Collins HM, Kibble NA, et al. Over-expression of specific HvCslF cellulose synthase-like genes in transgenic barley increases the levels of cell wall (1,3;1,4)-beta-d-glucans and alters their fine structure. *Plant Biotechnol J* 2011; **9**(2): 117-35.

3. Burton RA, Wilson SM, Hrmova M, et al. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. *Science* 2006; **311**(5769): 1940-2.

4. Guillon F, Bouchet B, Jamme F, et al. Brachypodium distachyon grain: characterization of endosperm cell walls. *Journal of Experimental Botany* 2011; **62**(3): 1001-15.

5. Kim S-J, Zemelis S, Keegstra K, Brandizzi F. The cytoplasmic localization of the catalytic site of CSLF6 supports a channeling model for the biosynthesis of mixed-linkage glucan. *The Plant Journal* 2015; 81(4): 537-47.

6. Philip Brewer et al. (2006) *In situ* hybridization for mRNA detection in *Arabidopsis* tissue sections. Nature Protocols. Vol 1. No.3.

7. Roche. Manuals for DIG labeling, Colorimetric detection, and Hybridization of RNA probes. <u>http://www.roche-applied-science.com</u>

8. Simon Malcomber et al. (2003) Kellogg Lab in situ hybridization protocol using DIG-labeled probes.

Funding statement.

*This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).* 

#### **GLBRC Targeted Metabolomics: Enabling the Science of Lignocellulose Bioconversion**

Alan J. Higbee <sup>1,2,3</sup> \* (ahigbee@wisc.edu), Jason D. Russell,<sup>1,4</sup>, David H. Keating,<sup>3,5</sup> Trey K. Sato,<sup>3</sup> Daniel R. Noguera,<sup>3,6</sup> Samantha Austin,<sup>3,6</sup> Arne Ulbrich,<sup>1,3,7</sup>, Mick M. McGee,<sup>3</sup> Mike Westphall,<sup>1</sup> **Timothy J. Donohue**, <sup>3,8</sup> and **Joshua J. Coon**,<sup>1,9,10</sup>

<sup>1</sup>Laboratory of Professor Joshua J. Coon, University of Wisconsin, Madison; <sup>2</sup>University of Wisconsin Genome Center, Madison, Wisconsin; <sup>3</sup>DOE Great Lakes Bioenergy Research Center, Madison, Wisconsin; <sup>4</sup>Morgridge Institute for Research, Madison, Wisconsin; <sup>5</sup>Synata Bio, Warrenville IL; <sup>6</sup> Department of Civil and Environmental Engineering, University of Wisconsin, Madison; <sup>7</sup>Dow Chemical Company, Midland, MI; <sup>8</sup> Department of Bacteriology, University of Wisconsin, Madison; <sup>9</sup> Department of Chemistry, University of Wisconsin, Madison; <sup>10</sup> Department of Biomolecular Chemistry, University of Wisconsin, Madison

https://www.glbrc.org/

coonlabs.com

Project Goals: "GLBRC targeted metabolomics" refers to a number of assays that have been developed to enable and guide research focused on identifying and overcoming limitations that reduce the efficiency of lignocelluose conversion to biofuels. Measurement of inhibitory as well as growth promoting compounds, and of compounds arising from microbial metabolism found in lignocellulosic hydrolysates aims to elucidate the causes of the inhibition of microbial utilization of xylose in these feedstock materials and to identify other growth limiting factors. These methods have also been applied to understanding metabolism of inhibitors by various microbes. Measurement of intracellular metabolites and cofactors associated with conversion of sugars to biofuels are used to inform researchers of the metabolic consequences of these inhibitors, the efficacy of strain improvement efforts and to guide future genetic augmentations by identifying potential points of restricted metabolic throughput.

Abstract: The Great Lakes Bioenergy Research Center (GLBRC) in collaboration with the research group of Dr. Joshua J. Coon (Coon Laboratories) has developed a number of assays for specific compounds of interest aimed at enabling efforts to improve the efficiency of lignocellulose deconstruction and conversion to biofuels. During anaerobic fermentation in hydrolysates prepared from Ammonia Fiber Expansion (AFEX®)-pretreated corn stover (ACSH) specialized yeast and bacterial strains are capable of complete conversion of glucose to ethanol, but convert xylose much less efficiently. Xylose can be completely converted during growth in laboratory medium, indicating that components of ACSH inhibit xylose utilization. To identify the mechanism underlying the reduced efficiency of xylose conversion we have developed methods for characterizing the chemical composition of ACSH and other hydrolysates. To understand the effects of ACSH inhibitors on cell physiology, we use a combination of assays to

determine most of the metabolic intermediates associated with bioconversion of sugars to fuels. These methods are also used to investigate the consequences of genetic changes and strain improvement efforts primarily aimed at developing strains that convert xylose to biofuels effectively in ACSH.

A combination of ion exchange chromatography and ion pairing chromatography coupled to tandem mass spectrometric detection gives measurements of most of the critical components of central energy metabolism in cellular extracts. HILIC chromatography with tandem mass spectrometry detection is used to measure amino acid abundances in cellular extracts and also in highly complex and concentrated extracellular media such as lignocellulosic hydrolysates. Intracellular products associated with xylose assimilation are determined using methoximation/trimethylsilylation and GC-MS, enabling assessment of the functionality of this pathway as engineered into S. cerevisiae with the goal of achieving viable rates of xylose bioconversion in a known ethanologen with no endogenous capability to assimilate and utilize xylose. Products of lignin and carbohydrate decomposition referred to as "lignotoxins" in hydrolysates (and their metabolites after microbial action) are measured using reversed phase chromatography routinely coupled to tandem mass spectrometry for targeted analysis of 36 known components. If the ability to detect and potentially identify unknown components is also desired, full scan high resolution / accurate mass spectrometry with fast polarity switching and data dependent MS/MS can be used allowing simultaneous determination of known compounds and nontargeted analysis. Lignotoxin analysis has enabled several lines of research. Headspace solid phase microextraction (HS-SPME) with GC-MS is used for determination of volatile and semivolatile fermentation products and components of hydrolysates and that are not easily measured by other means such as acetaldehyde, furfural and acetamide. Levels of isobutanol production in microbial cultures are routinely measured using headspace sampling GC-MS. Recently we have also developed a HILIC-UHPLC-MS/MS method for determination of most of the compounds in the methylerythritol 4phosphate (MEP) pathway leading to production of isoprenoids. Many of these methods employ stable isotope labeled internal standards (SILIS) to give the most accurate and robust absolute quantitation possible.

*This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).* 

#### Extracellular Products Mediate Bacterial Synergism in Cellulose Degradation

Camila Carlos,<sup>1,2\*</sup> (ccarlos@wisc.edu), Andrea Lubbe,<sup>4</sup> Emily T Beebe,<sup>3</sup> Brian G. Fox,<sup>3</sup> Trent R. Northen,<sup>4</sup> and **Cameron R. Currie**<sup>2</sup>

<sup>1</sup>Department of Bacteriology, University of Wisconsin–Madison, WI 53706
<sup>2</sup>Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin–Madison, WI 53726
<sup>3</sup>Department of Biochemistry, University of Wisconsin–Madison, WI 53706
<sup>4</sup>US Department of Energy Joint BioEnergy Institute, Emeryville, CA 94608; Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA

#### **Project Goals:**

## 1) Investigate the links between microbial biodiversity and plant biomass decomposition rates

## 2) To understand the molecular bases of microbial interactions during plant biomass decomposition

It has been demonstrated that microbial species richness influences ecosystem level processes such as litter decomposition, CO<sub>2</sub> flux and nitrogen cycling, however it is unclear how species interactions and dynamics affect these ecosystem services [1]. Understanding how microbial species coexist and interact during plant biomass decomposition is essential for the establishment of links between biodiversity and carbon cycling. Natural microbial communities often harbor thousands of species, making it difficult to identify the functions and interactions of individual community members. To overcome these limitations, some researchers have utilized synthetic microbial consortia as simplified models of natural microbial communities. This approach allows manipulation of species diversity and taxonomic composition as well as control of environmental variables. Using this bottom-up approach, Tiunov and Scheu [2] manipulated the species diversity of communities of fungal decomposers (ranging from 1 to 5 species) and identified that facilitative interaction plays a major role in decomposition processes.

In order to simplify the richness and interactions in decomposer communities, we established soil-derived microbial consortia enriched on pretreated switchgrass. Five bacterial genera were consistently isolated from these consortia: (A) *Cellulomonas*, (B) *Cellulosimicrobium*, (C) *Ensifer*, (D) *Pseudomonas* and (E) *Ochrobactrum*. To investigate the social interactions among these isolates, we manipulated the species richness of 31 bacterial consortia representing all the possible combinations of isolates (from mono-cultures to five species in combination) and assessed their growth and decomposition rates on different carbon sources: glucose, polyglutamic acid, xylan and cellulose.

Interestingly, we found that facilitative interactions among members of the consortia emerged on complex substrates and, in general, multispecies consortia outperformed the decomposition and productivity rates obtained by mono-cultures (Figure 1). These results have important implications on our understanding of microbial symbiotic interactions and microbial coexistence in nutrient-poor environments.

Genome sequencing of the five strains revealed that the increase in the CAZy domains (or lignocellulolytic enzymes) repertoire of the multispecies consortia do not explain the observed performance of the consortia.

We further analyzed the most cellulolytic pair of strains (Cellulomonas-Pseudomonas) to understand the molecular bases of the synergism during cellulose degradation Figure 1B). Enzymatic activity profiling of secreted proteins also revealed that there are no synergistic interactions between the enzymes secreted by different species. These results indicate that exometabolites are likely to mediate the coexistence of these species during plant biomass decomposition. We utilized untargeted GC-MS to analyze the extracellular metabolites of the monoand co-cultures to identify the metabolites involved in facilitative interactions. Vitamins and growth factors were identified as potential mediators of microbial interactions during plant biomass decomposition. One example is the nicotinic acid which is detected in the mono-culture of Pseudomonas, but showed decreased peak area in the co-culture, indicating its consumption by Cellulomonas. Moreover, the co-culture of Cellulomonas and Pseudomonas was found to produce high amounts of some organic acids of industrial interest (included in the list created by the Biotechonology for Biofuels: [3]), such as malic acid and phenyllactic acid. Our findings represent an important step toward understanding the molecular bases of microbial interactions during plant biomass decomposition.



**Figure 1.** Effect of species richness and composition on total growth and degradation rate of various plant biomass polymers. Expected values were calculated by an additive model of the OD600 of monocultures growing at the same polymer.

#### References

- Keitt TH, Addis C, Mitchell D, Salas A, Hawkes CV. (2016) Climate change, microbes, and soil carbon cycling. In: Marxsen J, Liebig J Climate Change and Microbial Ecology: Current and Future Trends. Norwich, UK: Caister Academic Press; pp. 220.
- 2. Tiunov AV and Scheu S. (2005) Facilitative interactions rather than resource partitioning drive diversityfunctioning relationships in laboratory fungal communities. Ecol Lett 8: 618-25.
- 3. du Preez JC (2016) Editorial: chemicals and bioproducts from biomass. Biotech Biofuels 9:233.

#### Funding statement

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

## Quantification of monolignol ferulate conjugate in Zip-lignin poplar by stimulated Raman scattering microscopy

Wei Shen<sup>1,2\*</sup> (weishen@msu.edu), John Ralph<sup>1,3</sup>, Shawn D. Mansfield<sup>1,4</sup>, and Shi-You Ding<sup>1,2</sup>

<sup>1</sup>Department of Energy Great Lake Bioenergy Research Center, Michigan State University, East Lansing, MI, 48824 USA; <sup>2</sup>Department of Plant Biology, Michigan State University, East Lansing, MI, 48824 USA; <sup>3</sup>Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA; <sup>4</sup>Department of Wood Science, University of British Columbia, Vancouver, BC, Canada

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

# Project Goals: To quantify and localize the monolignol ferulate conjugates (zip-lignin) in genetically modified poplar cell walls by *in situ* chemical imaging using non-destructive hyperspectral Stimulated Raman Scattering (hsSRS) microscopy, therefore provide deeper understanding of cell wall deconstructability affected by lignin chemistry.

GLBRC has demonstrated that Zip-lignin poplar contains more abundant ester linkages by incorporating monolignol ferulate conjugates in lignin backbones. Ester linkages can be easily hydrolyzed under basic condition, which results in increased cell wall digestibility after mild alkaline pretreatment. Zip-lignin also shows increased digestibility under other oxidative pretreatment, suggesting there are complex chemical and topological factors affecting the zip-lignin poplar cell wall features. Among these factors the amount of ferulate moieties incorporated into lignin backbones are of particular interest. Up to this point, they are generally estimated by measuring the ferulate derivatives released through reactions that specifically cleave lignin  $\beta$ -ethers but leave the  $\gamma$ -esters intact. However, the low amount of ferulate derivatives released through this process makes the accurate quantification problematic. In addition, these destructive methods suffer from only delineating bulk structure, i.e., from materials that have been finely ground and thus lose all ultrastructural information, which is of critical importance to understand how cell wall deconstructability is affected by lignin chemistry. In this work, we develop an *in situ* chemical imaging system based on hyperspectral stimulated Raman scattering microscopy with high spectral and spatial resolution. Using this platform, hyper-spectra are taken between 1500 – 1700 cm<sup>-1</sup>, new algorithm is developed to analyze the fine chemistry of the  $\alpha$ - $\beta$  carbon double bonds that present in all type of lignin. We found a  $\sim 30$  cm<sup>-1</sup> shift in their Raman spectrum due to the conjugation between carbon double bond and the carbon oxygen double bond in carbonyl group, and this shift is specific to zip-lignin and has been verified through reduction of ferulate to its corresponding alcohol. By multivariate curve resolution of the hyperspectral images, we uncovered a spatially distinct distribution of zip and native lignin in plant cell wall. These results contribute to a deeper understanding of zip lignin formation process and its effect on biomass deconstruction processes.

The Great Lakes Bioenergy Research Center is a U.S. Department of Energy Bioenergy Center supported by the DOE Basic Energy Research Office of Science (DE-FC02-07ER64494).

#### Using Genetic and Bioreactor Engineering to Produce Oleaginous Bacteria

Kimberly C. Lemmer<sup>1</sup>\*(kim.lemmer@wisc.edu), Weiping Zhang<sup>1,5</sup>, Samantha J. Langer<sup>1</sup>, Alice C. Dohnalkova<sup>8</sup>, Dehong Hu<sup>8</sup>, Rachelle A. Lemke<sup>4</sup>, Jeff S. Piotrowski<sup>1</sup>, Paul D. Hutchins<sup>1,3,6</sup>, Jason D. Russell<sup>1,2,3</sup>, Joshua J. Coon<sup>1,3,6,7</sup>, Galya Orr<sup>8</sup>, Daniel R. Noguera<sup>1,5</sup>, **Timothy J. Donohue**<sup>1,4</sup>

<sup>1</sup>DOE Great Lakes Bioenergy Research Center, <sup>2</sup>Morgridge Institute for Research, <sup>3</sup>Genome Center of Wisconsin, <sup>4</sup>Department of Bacteriology, <sup>5</sup>Department of Civil and Environmental Engineering, <sup>6</sup>Department of Chemistry, <sup>7</sup>Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI; <sup>8</sup>Pacific Northwest National Laboratory, Environmental Molecular Sciences Laboratory, Richland, WA.

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

**Project Goals:** In order to generate knowledge needed to produce next generation biofuels or fuel precursors, the GLBRC identified a need to increase microbial synthesis of anabolic compounds under low  $O_2$ /anaerobic conditions. Our studies address this key research priority for GLBRC by focusing on the synthesis of lipids and fatty acids under low  $O_2$ /anaerobic conditions in *R. sphaeroides*, a facultative microbe that has several novel features relevant to production of advanced biofuel precursors. Our strategy is to use the insight we gain to develop approaches to improve the yield of fatty acids or other hydrocarbons in this or possibly other microbes being studied in GLBRC. Thus, we predict that the new information gained from this project will increase our ability to modify bacteria and possibly other cells to advance a systems biology understanding for biofuel and bio-based products.

Lipids from microbes offer a promising source of renewable alternatives to petroleum-derived compounds. In particular, oleaginous microbes are of interest because they accumulate a large fraction of their biomass as lipids. In this study, we investigate whether it is possible to exploit the native metabolic and regulatory pathways of non-oleaginous bacteria to increase lipid production to oleaginous levels. We use *Rhodobacter sphaeroides*, a facultative bacterium that has a native ability to increase phospholipid membrane content under low  $O_2$  conditions. We screened a *Rhodobacter sphaeroides* Tn5-mutant library for insertions that increased fatty acid content at high  $O_2$  and identified ten high-lipid (HL) mutants for further characterization. We found that the genetic lesions in these mutants did not disrupt pathways known to impinge on fatty acid accumulation. Instead, we found that these HL mutants exhibited changes in their cell envelopes, including sensitivity to drugs that target this region, changes in shape, and ability to secrete lipids, with two HL mutants accumulating ~60% of their total lipids extracellularly. Analysis of the lipid secretions suggests that inner membrane or periplasmic components, including quinones, could also be secreted by these mutants as value-added coproducts. We used one of the highest lipid secreting strains to grow high-density cultures in a fed-batch bioreactor,

and produced 1.3 g/L fatty acids, with lipid content comparable to that of oleaginous microbes, but having the characteristic that the majority of the lipids were excreted. Thus, by combining a single genetic alteration with bioreactor engineering we have converted *R. sphaeroides* into an oleaginous bacterium. Based on the properties of these HL mutants, we conclude that alterations of the cell envelope can be used as novel approach to increase microbial lipid production and secretion, and may be applicable to other organisms.



This work was supported by DOE Great Lakes Bioenergy Research Center grant (DOE Office of Science BER DE-FC02-07ER64494) to TJD. Microscopy was performed at the Environmental Molecular Science Laboratory (EMSL), a DOE Office of Science user facility sponsored by the Office of Biological and Environmental Research and located at PNNL.

#### A Cellulase Family Reunion: Observing and Predicting the Structural Changes Accompanying the Evolution of GH5 Enzyme Specificity

Kirk Vander Meulen<sup>1</sup>\* (kirk.vandermeulen@wisc.edu), Evan Glasgow<sup>1</sup>\* (eglasgow@wisc.edu), Taichi Takasuka<sup>2</sup>, Christopher Bianchetti<sup>3</sup>, Sam Deutsch<sup>4</sup>, and **Brian G. Fox<sup>1</sup>** 

<sup>1</sup>Great Lakes Bioenergy Research Center, University of Wisconsin – Madison; <sup>2</sup>Hokkaido University, Sapporo, Hokkaido, Japan; <sup>3</sup>University of Wisconsin – Oshkosh; <sup>4</sup>Joint Genome Institute, Walnut Creek, CA.

## Project Goals: Synthesize and screen a subfamily of polysaccharide-degrading enzymes, determine the sequence and structural determinants of substrate specificity, predict functions of unknown enzymes, and model the evolution of enzymatic activities.

We have performed a large-scale analysis of the diverse activities exhibited by members of subfamily 4 in glycoside hydrolase family 5 (GH5\_4), a subfamily notable for the range of substrate specificities among its members. Using high-throughput cell-free expression, we tested for the specificities exhibited by 237 GH5\_4 enzymes. Of the activities tested, lichenase activity was the most prevalent (present in 85% of enzymes), followed by xylanase (70%) and mannanase (44%). Two loop regions, each connecting a TIM-barrel core beta-strand with its respective C-terminal alpha-helix, contribute to the geometry of the GH5 hallmark binding channel, and each plays a critical role in determining activity and specificity. One loop is primed for contacts with the positive subsites of a substrate oligosaccharide, and correspondingly an absence of all activities is highly correlated with the absence of a channel-exposed tryptophan. The second critical loop is located both across and upstream from this channel position, and is correspondingly known to form contacts with substrate negative subsites. We have found that in one entire clade of enzymes, both loops are truncated considerably, affecting a striking change in binding channel geometry.

DOE Great Lakes Bioenergy Research Center and Joint Genome Institute are supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contracts DE-FC02-07ER64494 (GLBRC), and DE-AC02-05CH11231(JBEI/JGI), respectively.

#### Using N and O Isotopes to Determine the Source of Microbial N<sub>2</sub>O Production

Nathaniel E. Ostrom<sup>1</sup>, Joshua A. Haslun<sup>1</sup>\* (haslunjo@msu.edu), Peggy H. Ostrom<sup>1</sup>, and Eric L. Hegg<sup>1</sup>

<sup>1</sup>Michigan State University, East Lansing

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

Project Goals: The Great Lakes Bioenergy Research Center supports research that generates technology for the conversion of cellulosic biomass to biofuels. However, agricultural production of cellulosic biomass is associated with emissions of greenhouse gases such as N<sub>2</sub>O and CO<sub>2</sub>. N<sub>2</sub>O is particularly problematic because it has a 100-year global warming potential 300 times that of CO<sub>2</sub> and is the leading cause of stratospheric ozone depletion. Currently the relative importance of microbial processes contributing to N<sub>2</sub>O flux is poorly understood. Stable isotopes of N and O have emerged as a tool to discriminate between pathways responsible for N<sub>2</sub>O flux. The goals of our project are to (1) determine the sources of microbial N<sub>2</sub>O production in soil communities using stable isotopes, (2) elucidate the biochemical pathways of N<sub>2</sub>O production associated with specific naturally occurring enzymes and (3) develop methodology for real-time source determination via stable isotope analysis using laser spectroscopy.

Site preference (SP), the difference in the isotope ratios between the central and terminal N atoms in nitrous oxide (N<sub>2</sub>O), has emerged as a robust discriminator of the microbial origins of N<sub>2</sub>O. SP has emerged as a conservative tracer of N<sub>2</sub>O production, differentiating N<sub>2</sub>O produced bacterial nitrification and denitrification. In contrast, the  $\delta^{15}$ N and  $\delta^{18}$ O of N<sub>2</sub>O are not conservative tracers of N<sub>2</sub>O production. A number of variables (e.g. electron donor abundance, nitrogen substrate and growth rate) markedly influence the  $\delta^{15}$ N and  $\delta^{18}$ O of N<sub>2</sub>O. Importantly, the influence of these variates on SP has not been documented. Thus, we investigated the influence of these variates on SP with different bacterial denitrifiers. Specifically, we examined the effect of the electron donor source (citrate or succinate) and concentration (10, 1, 0.1, and 0.01 mM) on SP in *Pseudomonas chlororaphis subsp. chlororaphis* and *Pseudomonas chlororaphis subsp. aureofaciens*. There was no net isotope effect (no shift in SP) observed across the conditions evaluated (ANOVA, *p*>0.05). These data confirm that SP is a conservative tracer of N<sub>2</sub>O production. The observation that SP differed slightly between our two denitrifiers suggests that changes in the predominance of different bacterial species may account for some of the variation in SP for denitrifiers reported in the literature.

In addition to evaluating the influence of different factors on SP, we are now beginning to evaluate how differences in enzymatic mechanisms contribute to variation in SP. We are specifically interested in nitric oxide reductases (NORs), which are responsible for converting nitric oxide (NO) to N<sub>2</sub>O. We have chosen to examine if shifts in SP occur for bacterial

cytochrome *c*-dependent NOR, bacterial quinol-dependent NOR, and hydroxylamine oxidoreductase.

Trace gas-isotope ratio mass spectrometry (TG-IRMS) is the current the gold standard for isotopic measurements of N<sub>2</sub>O; however the sporadic nature of N<sub>2</sub>O production, along with the costs and analytical time associated with TG-IRMS create challenges to its large scale application. Isotopic analysis of N<sub>2</sub>O *via* laser spectroscopy offers both reduced analytical time, *in situ* analysis, and continuous real-time measurements of SP. We show that SP determinations from the LGR are comparable to those generated by traditional TG-IRMS. Further, we are able to measure SP at atmospheric concentrations without pre-concentration, a first in spectroscopic analysis of SP.

#### Funding statement

This work was funded by the National Science Foundation's Earth Sciences Instrumentation and Facilities program (grant #1456430) and the Department of Energy Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494). The authors declare that they have no conflict of interest.

#### **Transcriptional Analysis of Flowering Time in Switchgrass**

Carl-Erik Tornqvist<sup>1\*</sup> (citornqvist@wisc.edu), Brieanne Vaillancourt<sup>3</sup>, Jeongwoon Kim<sup>3,4</sup>, **C. Robin Buell<sup>3</sup>**, **Michael D. Casler<sup>1,2</sup>**, and **Shawn Kaeppler<sup>1</sup>** 

<sup>1</sup>Department of Agronomy and DOE Great Lakes Bioenergy Research Center, University of Wisconsin – Madison, Madison, Wisconsin; <sup>2</sup>USDA Dairy Forage Research Center, Madison, Wisconsin; <sup>3</sup>Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, Michigan; <sup>4</sup>Current address: Monsanto Company, Chesterfield, Missouri

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

Project Goals: The project is intended to develop tools to streamline and improve efficiency for development of switchgrass as a sustainable biofuel crop. We use genomic information from maize and other species to identify genes responsible for traits that make switchgrass a more sustainable biofuel crop and to develop DNAbased selection methods to speed up the process of developing improved switchgrass cultivars. One of the project's objectives is to identify genes involved in the control of flowering time, a trait correlated with biomass yield. The work described below details one of the approaches we used to investigate flowering time in switchgrass.

Over the past two decades, switchgrass (*Panicum virgatum*) has emerged as a priority biofuel feedstock. The bulk of switchgrass biomass is in the vegetative portion of the plant, therefore, increasing the length of vegetative growth will lead to increase in overall biomass yield. The goal of this study was to gain insight into the control of flowering time in switchgrass, to assist in developing cultivars that have extended vegetative phases through delayed flowering. RNA-sequencing was used to assess genome-wide profiles of transcription across time, between switchgrass genotypes belonging to the two main ecotypes: upland, typically early flowering, and lowland, typically late flowering. Leaves and tissues enriched for the shoot apical meristem (SAM) were collected in a developmental series from emergence through anthesis for RNA extraction. The morphology of the SAM was also tracked to determine the stage of meristem transition and to choose samples that flanked the SAM transition for sequencing and expression analyses. The analyses revealed differential expression patterns between early- and lateflowering genotypes for known flowering time orthologs. In addition, based on expression patterns, many of the identified genes could be classified as putative promoters or repressors of flowering. The candidate genes presented here will be further supported by functional confirmation in switchgrass or the model grass, Brachypodium, and may be used to guide switchgrass improvement through marker assisted breeding and/or transgenic or gene editing approaches.

This work is funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

#### Optimization of Isobutanol Production by Zymomonas mobilis

Indro Ghosh<sup>1</sup>\* (inghosh@wisc.edu), and Robert Landick<sup>1</sup>

<sup>1</sup>Great Lakes Bioenergy Research Center, University of Wisconsin-Madison

https://www.glbrc.org/research/conversion

#### **Project Goals:**

We are optimizing the production of the 'Next-Gen' biofuel isobutanol by *Zymomonas mobilis*, a bacterial species capable of highly efficient fermentation. By expressing a novel set of genes, we have demonstrated production of isobutanol by *Z. mobilis*. To fully harness the rapid and efficient Glucose metabolism of *Z. mobilis*, we are eliminating metabolic bottlenecks in the conversion of Glucose to Isobutanol. Specifically, we will (1) exploit an engineered link between isobutanol production and growth in *Escherichia coli* to estimate best ratio of enzyme expression, (2) measure levels of intermediate metabolites in *E. coli* and *Z. mobilis* to identify bottlenecks, (3) adjust expression levels of pathway genes, and (4) disrupt competing metabolic pathways to maximize intracellular carbon diversion towards isobutanol. Applying this strategy of identifying bottlenecks before remedying them in an iterative fashion will greatly improve titers of isobutanol produced by *Z. mobilis* and help develop an industrially relevant producer strain.

Efficient production of isobutanol by microbes on an industrial scale is of great interest. Isobutanol has been produced previously in *E. coli* and *Saccharomyces cerevisiae* using five genes; an overproduced valine precursor:  $\alpha$ -ketoisovalerate (using *alsS, ilvC,* and *ilvD*) is converted to isobutanol (using *kivd,* and *adhA*). Predictions based on measured enzyme kinetics indicate that efficient isobutanol bioproduction will occur only when bottlenecks within the pathway are eliminated, and when synthesis of enzymes imposes minimal burden to cells. Optimizing enzyme expression levels will accomplish both of these requirements and allow efficient microbial production of isobutanol.

We have demonstrated production of isobutanol by *Z. mobilis* by expressing enzymes (*kivd*, *adhA*) that convert naturally produced  $\alpha$ -ketoisovalerate into isobutanol. To increase titers of produced isobutanol, we are expressing additional genes (*alsS, ilvC, ilvD*) to convert pyruvate to isobutanol in *Z. mobilis*, something that is complicated by the frequent rejection of foreign genes by the innate immune systems of *Z. mobilis*. We are also testing multiple strategies to inactivate *Z. mobilis* Pdc, in order to redirect carbon flux away from the naturally produced ethanol, towards isobutanol. These approaches will hypothetically enable high titer isobutanol production from *Z. mobilis*.

Expression of an optimal ratio of enzymes in *Z. mobilis* will be aided by metabolic intermediate level measurements and additional studies in *E. coli*. A preliminary OptSSeq<sup>1</sup> experiment has

indicated that expression of *ilvC* is essential for efficient pathway function in *E. coli*, in line with expectations based on enzyme kinetics. Upon expressing an isobutanol producing cassette in *Z. mobilis*, we will measure levels of enzymes produced, and also levels of intermediate metabolites. Significant metabolite accumulations will be indicative of bottlenecks, and will guide the design of subsequent cassette iterations to remedy identified bottlenecks. An optimized strain developed using these strategies will be an industrially relevant isobutanol producer.

#### **Publications**

 Ghosh IN, Landick R (2016) OptSSeq: High-Throughput Sequencing Readout of Growth Enrichment Defines Optimal Gene Expression Elements for Homoethanologenesis. ACS synthetic biology 5:1519-1534

This work was funded by the DOE Great Lakes Bioenergy Research Center. (DOE BER Office of Science DE-FC02-07ER64494)

#### Process Systems Engineering for Biofuels and Bio-based Chemicals

Rex T. L. Ng<sup>1,2</sup>\*(<u>rextl.ng@wisc.edu</u>), Wangyun Won,<sup>1,2</sup> Kirti M. Yenkie,<sup>1,2</sup> and Christos T. Maravelias<sup>1,2</sup>

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

https://www.glbrc.org/research/deconstruction

Project Goals: The goal of our research is to develop methods and tools for the synthesis and evaluation of biomass-to-fuels strategies employing pretreatment and conversion technologies developed within the Great Lake Bioenergy Research Center (GLBRC). Specifically, we design new processes, develop models for their evaluation, and eventually identify technological bottlenecks and economic drivers that have to be addressed. Additionally, we develop a general framework for the synthesis of separation and purification networks for bio-based products based on their physical properties. Finally, we explore optimization methods for the design and operation of biofuel supply chain networks with regional depots.

**Process Synthesis and Analysis.** We develop an integrated strategy that utilizes a mixture of  $\gamma$ -valerolactone (GVL), water, and toluene as a solvent containing dilute sulfuric acid as a catalyst for the production of ethanol from lignocellulosic biomass<sup>1</sup>. Specifically, cellulose and hemicellulose fractions are first hydrolyzed into sugars using solvent mixtures and catalyst, and the sugars are then co-fermented into ethanol over engineered yeast strains. We design separation subsystems to (1) effectively recover GVL and toluene for reuse in biomass hydrolysis, (2) recover sugars and GVL from the residual biomass, and (3) recover lignin and humins for heat and power generation. To minimize utility requirements, we conduct heat integration, which allows us to meet all heating requirement using biomass residues. Then, we perform a range of system-level analyses to identify the major cost and technological drivers. We found that the proposed strategy results in lower minimum ethanol selling price ((\$ 2.95 per gallon of gasoline equivalent (GGE<sup>-1</sup>)) than those for the GVL/water-based non-enzymatic ethanol production (\$ 3.37 GGE<sup>-1</sup>)<sup>2</sup> and the enzymatic/biochemical ethanol production (\$ 3.27 GGE<sup>-1</sup>)<sup>3</sup> strategies.

**Design of Separation Networks for Bioproducts**. Separation of bioproducts is a major challenge that can contribute to more than 70% of the total production costs<sup>4,5</sup>. Based on the physical properties of the desired chemical and other components in the bioreactor effluent, there can be multiple feasible options for product recovery<sup>6,7</sup>. These options are composed of several alternative technologies performing similar tasks and their suitability for a particular chemical varies based on the difference in key parameters such as separation efficiency, cost or amount of added separating agents, biomass titer, and desired product purity. To develop new separation strategies, we propose an optimization-based framework that allows us to identify the key cost drivers and critical technologies. This framework provides significant insights for technology selection and

assists in making an informed decision regarding technologies that should be used for a given set of input/output specifications.

**Biofuel Supply Design and Operation**. To ensure biomass supply meets biofuel demand, it is necessary to have an effective supply network. Towards this end, the concept of regional biomass processing depot<sup>8</sup>, where biomass is pretreated and/or densified to a higher density intermediate, has been introduced to improve the performance of supply network in terms of costs and emissions. We develop optimization models for the design and operation planning of biofuel supply chains with regional depots that account for seasonal biomass supply<sup>9,10</sup>. Unlike previous approaches which assume predetermined depot and biorefinery locations, we treat the locations of depots and/or biorefineries as continuous optimization decisions. The proposed models account for technology selection and capacity planning decisions, as well as auxiliary decisions such as harvesting site and biomass feedstock selection, biomass allocation to depots and biorefineries, and inventory planning. Furthermore, the proposed models can be extended to handle different features and practical considerations.

#### References

- Motagamwala, A. H.; Won, W.; Maravelias, C. T.; Dumesic, J. A. An Engineered Solvent System for Sugar Production from Lignocellulosic Biomass Using Biomass Derived γ-Valerolactone. *Green Chem.* 2016, 18 (21), 5756.
- (2) Han, J.; Luterbacher, J. S.; Alonso, D. M.; Dumesic, J. A.; Maravelias, C. T. A Lignocellulosic Ethanol Strategy via Nonenzymatic Sugar Production: Process Synthesis and Analysis. *Bioresour. Technol.* 2015, 182, 258.
- (3) Humbird, D.; Davis, R.; Tao, L.; Kinchin, C.; Hsu, D.; Aden, A.; Schoen, P.; Lukas, J.; Olthof, B.; Worley, M. Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover.
- (4) Kiss, A. A.; Grievink, J.; Rito-Palomares, M. A Systems Engineering Perspective on Process Integration in Industrial Biotechnology. J. Chem. Technol. Biotechnol. 2015, 90 (3), 349.
- (5) Brandt, K.; Schembecker, G. Production Rate-Dependent Key Performance Indicators for a Systematic Design of Biochemical Downstream Processes. *Chem. Eng. Technol.* **2016**, *39* (2), 354.
- (6) Wu, W.; Yenkie, K.; Maravelias, C. T. A Superstructure-Based Framework for Bio-Separation Network Synthesis. *Comput. Chem. Eng.* 2017, *96*, 1.
- (7) Yenkie, K. M.; Wu, W.; Clark, R. L.; Pfleger, B. F.; Root, T. W.; Maravelias, C. T. A Roadmap for the Synthesis of Separation Networks for the Recovery of Bio-Based Chemicals: Matching Biological and Process Feasibility. *Biotechnol. Adv.* **2016**.
- (8) Eranki, P. L.; Bals, B. D.; Dale, B. E. Advanced Regional Biomass Processing Depots: A Key to the Logistical Challenges of the Cellulosic Biofuel Industry. *Biofuels, Bioprod. Biorefining* **2011**, *5*, 621.
- (9) Ng, R. T. L.; Maravelias, C. T. Design of Cellulosic Ethanol Supply Chains with Regional Depots. *Ind. Eng. Chem. Res.* **2016**, *55* (12), 3420.
- (10) Ng, R. T. L.; Maravelias, C. T. Design of Biofuel Supply Chains with Variable Regional Depot and Biorefinery Locations. *Renew. Energy* **2017**, *100*, 90.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

## Suppression of CINNAMOYL-CoA REDUCTASE increases the level of monolignol-ferulates in maize lignins

Rebecca A. Smith<sup>1,2</sup>\* (<u>rasmith29@wisc.edu</u>), Cynthia L. Cass,<sup>3</sup> Mona Mazaheri,<sup>1,4</sup> Rajandeep S. Sekhon,<sup>4</sup> Marlies Heckwolf,<sup>4</sup> Heidi Kaeppler,<sup>4</sup> Natalia de Leon,<sup>1,4</sup> Shawn D. Mansfield,<sup>5</sup> Shawn M. Kaeppler,<sup>1,4</sup> John C. Sedbrook,<sup>3</sup> Steven D. Karlen,<sup>1,2</sup> and **John Ralph**<sup>1,2</sup>

<sup>1</sup>Department of Energy Great Lakes Bioenergy Research Center, the Wisconsin Energy Institute, University of Wisconsin-Madison, Madison, WI 53726, USA; <sup>2</sup>Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706, USA; <sup>3</sup>Department of Energy Great Lakes Bioenergy Research Center, School of Biological Sciences, Illinois State University, Normal, IL 61790, USA; <sup>4</sup>Department of Agronomy, University of Wisconsin-Madison, Madison, WI 53706, USA; and <sup>5</sup>Department of Wood Science, University of British Columbia, Vancouver, BC, Canada

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

# Project Goals: To determine how to maximize the utilization of monolignol ferulate conjugates in cell wall lignification, where they introduce readily cleavable ester linkages ('zips') into the lignin polymer backbone in ways that significantly improve biomass-processing energetics.

One of the most promising alternatives to fossil fuels involves generating ethanol (or secondgeneration biofuels) from plant cell wall polysaccharides (sugars), such as cellulose and hemicelluloses. The problem with this strategy is that the production of cellulosic ethanol is not yet as efficient or cost-effective as fossil fuels. The main reason for this is the harsh treatments required to remove the recalcitrant cell wall phenolic polymer lignin in order to access the cell wall sugars. For this reason, decreasing or altering lignin content in potential biofuel crops has received a lot of attention from the scientific community. Lignin, though detrimental to biofuel processing, is essential to the support and defense of the plant body and decreases in lignin content are therefore usually associated with pendant stems, increased susceptibility to pathogens and reduced biomass yield. The digestibility of the cell walls can be improved by introducing labile ester bonds, which are broken under weak base treatment at room temperature, into the lignin backbone. The FERULOYL-CoA MONOLIGNOL TRANSFERASE (FMT) enzyme, which is naturally found in many plants, uses feruloyl-CoA and monolignols to synthesize the ester-linked monolignol ferulate conjugates. A mutation in the first lignin-specific biosynthetic enzyme, CINNAMOYL-CoA REDUCTASE (CCR), results in an increase in the pool of feruloyl-CoA. Maize (Zea mays) has a native putative FMT enzyme, and its ccr mutants produce an increased pool of feruloyl-CoA that can be used for conversion to monolignol ferulate conjugates. The increase in conjugates correlates with an improvement in the digestibility of maize stem rind tissue.

#### **Publications**

1. "Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone" Wilkerson, C. G.; Mansfield, S. D.; Lu, Fachuang; Withers, S.; Park, J.; Karlen, S. D.; Gonzales-Vigil, E.; Padmakshan, D.; Unda, F.; Rencoret, J; Ralph, J *Science*, **2014**, *344*, 90–93.

2. "Monolignol-ferulate conjugates are naturally incorporated into plant lignins" Karlen S.D., Peck M.L., Zhang C., Smith R.A., Padmakshan D., Helmich K.E., Free H.C.A., Lee S., Smith B.G., Lu F., Sedbrook J.C., Sibout R., Grabber J.H., Runge T.M., Mysore K.S., Harris P.J., Bartley L.E., and J. Ralph. Science Advances, **2016**, 2, e1600393

The Great Lakes Bioenergy Research Center is a U.S. Department of Energy Bioenergy Center supported by the DOE Basic Energy Research Office of Science (DE-FC02-07ER64494).

#### Designing highly specific protein-based small molecule biosensors

Srivatsan Raman, sraman4@wisc.edu Great Lakes Bioenergy Research Center, University of Wisconsin-Madison

Genetically-encoded small molecule biosensors are emerging as a powerful tool in synthetic biology for high-throughput phenotyping of metabolic pathway variants. Dynamic change in concentration of certain metabolites can serve as a direct readout of enzymatic activity, function of an operon or effects of regulatory variants on pathway flux. Real time measurement of metabolites allows us to dynamically regulate pathways in response to changes in the metabolic state. Microbial allosteric transcription factors (aTF) are widely used as small molecule biosensors in synthetic biology. The aTFs bind to a wide variety of small molecules such as sugars, phenolics, polycyclic aromatics, alkanes and other industrially useful molecules. Our inability to design a biosensor for a desired molecule is a major hurdle to further expanding the use of biosensors. We currently rely almost exclusively on natural biosensors. Here, we describe a general approach to make designer biosensors for new molecules by redesigning natural aTF specificity. We evaluate tens of thousands of computational design candidates with a high-throughput screen to identify these new allosteric biosensors. We redesigned the lac repressor to respond to four new molecules – gentiobiose, fucose, lactitol and sucralose with activity and specificity comparable to wild-type lac repressor for IPTG. We are currently working on building a suite of biosensors for high-value bioproducts, next-generation biofuels, environmental pollutants and redox cofactors. Biosensors would facilitate strain engineering, bioconversion and environmental bioprospecting.

#### Exploring RNA-Seq Expression Data With GxSeq

Nicholas Thrower,<sup>1,2</sup>\* (throwern@msu.edu), Matthew Larson,<sup>1,2</sup> and Curtis Wilkerson,<sup>1,2</sup>

<sup>1</sup>Michigan State University, East Lansing; <sup>2</sup>Great Lakes Bioenergy Research Center, East Lansing, Michigan

https://gxseq.glbrc.org

Project Goals: The Great Lakes Bioenergy Research Center is performing basic research to identify new technologies for the conversion of cellulosic biomass to biofuel. One goal is to develop biomass crops with desirable traits leading to more efficient fuel production. Understanding the genetic background for these desirable traits is possible due to advances in high throughput sequencing platforms. RNA-Seq or whole transcriptome sequencing is one of these advances and is becoming a standard for gene expression studies. To assist with storage and analysis of this sequence data, we created a custom web based application called GxSeq. The project began as a tool for investigating gene expression profiles within de-novo assemblies of novel organisms. Over time it has grown to include analyzing and interacting with fully annotated and sequenced genomes.

We have developed a web application (GxSeq) designed to facilitate RNA-Seq expression analysis and its variants. This software provides an array of interactive tables and visualizations customized to the needs of our researchers. The software also provides these researchers with a standard tool to access their data independent of the choice of upstream analysis tools. This reduces the overall complexity of data analysis and so reduces the workload for individual research groups allowing them to focus on experimentation. This software is developed using agile software methodologies, which allows us to quickly adapt the software to the specific needs of individual researchers. By having a close collaboration with researchers and integrating new requests in an expedient manner we are able to accommodate the rapidly changing nature of genomic data analysis. Additionally, this tool allows for center-wide access to large-scale datasets and information derived from analysis of these datasets.

Exploring RNA-Seq data often involves comparison of expression levels between genes or groups of genes. GxSeq has several tools designed to assist in this exploratory analysis including a tabular expression view, dynamic expression charts, k-means clustering and dynamic gene correlation tables. The tabular view can generate a matrix of counts or a ratio from user-selected samples. Interactive clustering allows exploration of gene expression patterns and dynamic gene correlation tables enable deeper exploration of the gene network. Tabular results are displayed alongside functional gene annotations in searchable, sortable tables that can be downloaded for further investigation.

It is also common to explore genes in context of the chromosomal sequence. This genomic view displays annotations and experimental data in a browsable interface. Interactive elements are

intuitive to use and allow dragging along the genome, selecting sequence regions and viewing annotation details. The URL is updated to reflect any changes allowing users to share or save the current view.

The combination of genomic sequence and gene expression data in GxSeq enables researchers to more effectively study the genetic background of desirable biomass traits.

This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494)

### Comparative Studies of Diverse Feedstocks and Identification of High Levels of Hemicellulose and Degradation Inhibitors That Impact Microbial Biofuel Synthesis

**Yaoping Zhang**<sup>1\*</sup> (yzhang8@wisc.edu), Donna Bates,<sup>1</sup> Yury Bukhman,<sup>1</sup> David Cavalier,<sup>2</sup> Lawrence G. Oates,<sup>1</sup> Rebecca Garlock Ong,<sup>2, 3</sup> and Trey Sato<sup>1</sup>

Great Lakes Bioenergy Research Center, <sup>1</sup>University of Wisconsin-Madison, Madison, WI, <sup>2</sup>Michigan State University, East Lansing, MI, <sup>3</sup>Department of Chemical Engineering, Michigan Technological University, Houghton, MI.

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

Project Goals: A variety of biomass feedstocks could be used for producing bio-based chemicals and biofuels. However, chemical compositions of various plant biomass, and even same feedstocks grown under different conditions, can vary widely. We previously found that Saccharomyces cerevisiae were significantly impaired for anaerobic fermentation of hydrolysates produced from pretreated switchgrass harvested from a drought year compared to non-drought years, as well as from pretreated corn stover from the same drought year. Compositional and functional analyses determined that high concentrations of pyrazines and imidazoles generated from excess soluble sugars in drought switchgrass during pretreatment partially contributed to this microbial inhibition. We then expanded our studies to five different feedstocks and would like to know how the levels of lignocellulose-derived inhibitors varied and how their impact on microbial response and fermentation performance, especially for xylose utilization. To avoid interannual variability, we have used lignocellulosic hydrolysates derived from five different plant feedstocks harvested from the same location and timeframe, as well as pretreated in an identical manner. This will allow us to identify feedstock-specific differences.

To investigate how much the known compounds and lignocellulose-derived inhibitors vary in different feedstocks and how these variations affect microbial response and biofuel synthesis, hydrolysates produced from five different feedstocks, including corn stover, switchgrass, miscanthus, sorghum, and mixed prairie, were compared by chemical composition analysis, chemical genomics, and microbial fermentation. To avoid the variability of different growth conditions, all feedstocks were harvested in the same year (Year 2014) from the same location, and then were pretreated with the same Ammonia Fiber Expansion method (AFEX). Hydrolysates were produced and then used to study microbial responses using chemical genomics and microbial fermentation.

Two ethanologens, *Saccharomyces cerevisiae* and *Zymomonas mobilis*, showed poor xylose utilization in corn stover hydrolysate compared to other feedstock hydrolysates. Chemical genomics also revealed a larger microbial stress response to corn stover hydrolysate. Comparison of the hydrolysate composition revealed the highest levels of several aromatic compounds in corn stover hydrolysate, including coumaric acid, coumaroyl amide, ferulic acid, and feruloyl amide. Supplementation of these inhibitors into a non-inhibitory hydrolysate inhibits

xylose utilization. These results indicate that elevated concentrations of hydroxycinnamic acid derived compounds inhibit xylose utilization in corn stover hydrolysate.

Chemical analysis of untreated feedstocks revealed that high concentrations of hydroxycinnamic acid-derived inhibitors originate from untreated corn stover and AFEX pretreatment. Furthermore, the levels of these inhibitors in the untreated biomasses and their hydrolysate varied both between feedstocks and between harvest years for the same feedstock. Year 2014 corn stover has the highest xylan/glucan content ratio, perhaps indicating joint regulation of the deposition of these compounds within the cell wall. Our results indicate that high hemicellulose content in year 2014 corn stover result in high levels of these inhibitors found in its hydrolysate, causing the poor microbial fermentation performance, especially xylose utilization. Although a higher hemicellulose content in feedstocks would be expected to contribute to greater ethanol yields, it may be linked to higher levels of inhibitors in hydrolysates and reduced microbial fermentation performance and xylose utilization.

#### References

Ong, R. G., A. Higbee, S. Bottoms, Q. Dickinson, D. Xie, S. A. Smith, J. Serate, E. Pohlmann, A. D. Jones, J. J. Coon, T. K. Sato, G. R. Sanford, D. Eilert, L. G. Oates, J. S. Piotrowski, D. M. Bates, D. Cavalier, and Y. Zhang. 2016. Reaction products of drought stressed bioenergy crops inhibit microbial biofuel production. Biotechnology for Biofuels 9:237.

*This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).*
# Self-Assembled Gold Nanoparticle Film for Nanostructure-Initiator Mass Spectrometry with Passive On-Line Salt Fractionation

Todd A. Duncombe<sup>1\*</sup> (taduncombe@lbl.gov), Paul Adams<sup>1</sup>, Anup Singh<sup>1,2</sup>, Trent Northen<sup>1</sup>

<sup>1</sup>Joint BioEnergy Institute, Emeryville, California, USA; <sup>2</sup>Sandia National Laboratories, Livermore, California, USA

# Project goals: Sensitive and robust array-based mass spectrometry tools are essential for performing metabolic screens. This work aims at developing a novel analytical platform to perform high-throughput analysis of complex samples using an on-line fractionation process followed by laser desorption ionization mass spectrometry.

Surface Assisted Laser Desorption Ionization (SALDI) is a method of ionization for mass spectrometry that uses a nanostructured medium to absorb energy from an incident laser and transfer that energy onto a target sample. The transferred energy ionizes and desorbs the target sample such that it can be injected into a mass analyzer for the charge to mass ratio of the molecules to be detected. In contrast to Matrix Assisted Laser Desorption Ionization (MALDI), SALDI does not require the addition of matrix ions to facilitate the LDI process. Thus eliminating deleterious 'matrix effects' such as ionization suppression of the molecules of interest, high background, or matrix ions with a similar mass to charge ratio obscuring the detection. SALDI-MS platform that has, in the past, leveraged laser-resonant wet etched-silicon nanostructures and initiator molecules for high sensitivity detection of adsorbed small molecules. lipids, and peptides in LDI-MS. Previously developed, nanostructure-initiator mass spectrometry (NIMS), required fabricating a nanostructured-silicon wafer using hydrofluoric acid etching followed by the manual coating of an initiator molecule. Due to the lowrepeatability of both steps – a wide degree variability is observed from one NIMS surface to the next. In this work, we created an improved NIMS platform utilizing fluorinated Au nanoparticles assembled on a surface for the analysis of spotted samples.

We present the self-assembly of fluorinated Au nanoparticle films as a massproducible fabrication methodology for generating nanostructure-initiator mass spectrometry substrates. The nanostructured surfaces enable the direct high sensitivity detection of peptides (20 femto mol) and other small molecules using laser desorption ionization. Further, through a photolithographic liftoff technique we can realize micropatterned fluorinated Au nanoparticle films. These micropatterns create a discrete wettability patterns, allowing us to passively fractionate hydrophobic molecules of interest from high-salt background environments for robust and predictable mass spectrometry.

# References

Northen, T.R., et. al., Clarthrate nanostructures for mass spectrometry, *Nature*, **449**, 1033 (2007).

This work conducted was supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-FC02-07ER64494 and through contract DE-AC02-05CH11231, respectively.

#### Methods for high-throughput massively parallel soil ecofunctional gene analysis

James R. Cole<sup>1</sup>\* (colej@msu.edu), Jiarong Guo,<sup>1</sup> Leo Tift,<sup>1</sup> James M. Tiedje,<sup>1</sup> and **Phil Robertson**<sup>1</sup>

<sup>1</sup>Michigan State University, East Lansing, Michigan 48824

#### **Project Goals:**

To improve the sustainable development of bioenergy, our project characterizes beneficial microbes in three important bioenergy crop systems and their impacts on critical biogeochemical processes, especially the nitrogen cycle, and ultimately explores ways to manage such microbiomes for improved biofuel sustainability.

#### http://www.glbrc.org

In analysis of biofuel crop soil microbiomes we are interested in targeting both genes providing a taxonomic/phylogenetic analysis of soil community microbial members and an analysis of ecofunctional genes involved in microbe-plant interactions, including genes coding for ecological processes important to plant and soil health. We are targeting these genes through both primer-targeted (PCR-based) and untargeted (shotgun) sequencing methods and analysis.

Standard taxonomic analysis has relied on amplification and sequencing of SSU rRNA genes. This method is well developed and inexpensive but has several drawbacks. First, the SSU rRNA genes are very slowly evolving and have limited resolving power for taxa below the level of genus; second, the SSU gene is often present in multiple copies per cell, presenting difficulties in calculating relative abundance of different organisms. Because there are fewer evolutionary constraints, many protein-coding core genes are more rapidly evolving than the SSU rRNA gene yet share many of the same advantages as a taxonomic marker, e.g., are present in all organisms and are unlikely to be horizontally transferred. In addition, most are normally present in a single copy. For our studies we have chosen the *rplB* gene coding for ribosomal protein L2. Examining completed bacterial genome sequences, we found that, as expected, the *rplB* gene shows more change than the SSU rRNA gene so is better able to resolve at the lower ranks, which are more ecologically relevant (Fig 1).



**Figure 1. A.** All completed genomes in Pseudomonadales are included in pairwise comparison; **B.** All completed genomes in Pseudomonas are included. The dashed line is y = x. Size of dots indicates number of genome pairs that share the same SSU rRNA gene identity and *rplB* identity.

However, it can be difficult to use these protein-coding core genes as markers because the known PCR primers have phylogenetic biases. To circumvent this problem, we have used the Xander

tool previously developed for this project (Wang et al., 2015. Xander: employing a novel method for efficient gene-targeted metagenomic assembly. *Microbiome* 3:32) to assemble and annotate core phylogenetic marker genes beyond the SSU rRNA gene. This Xander tool uses a Combined Weighted Assembly Graph (CWAG) that combines nodes from a standard De Bruijn graph representation of shotgun data with the states (nodes) of a Hidden Markov Model (HMM) for the gene(s) of interest. The HMM information adds weights to the graph edges and allows us to search for the gene(s) of interest using standard graph-theory path-finding algorithms on the CWAG. We used Xander to assemble *rplB* genes from 21 soil metagenomic data sets, seven replicates from each of three biofuel crops. We found that the *rplB* gene provided much better separation of the three biofuel crops than SSU rRNA genes isolated from the same metagenomic datasets (Fig 2).



**Figure 2.** Comparison of SSU rRNA gene and *rplB* in beta diversity analysis (ordination) using large soil metagenomes. Both genes show microbial community in corn (C) rhizosphere is significantly different from those in *Miscanthus* (M) and switchgrass (S). Additionally, *rplB* (both nucleotide and protein) separate communities of *Miscanthus* and switchgrass.

Many important ecofunctional genes are relatively rare and present in only a small fraction of bacterial cells. Bulk metagenomics either fail to accurately quantify changes in these genes or require excessive sequencing depth at greatly increased cost for sequencing and data processing. Protein-coding genes are, in general, less conserved than structural RNA genes, meaning that often no single probe or primer pair is able to target a gene's full range of diversity. We have been developing and testing an efficient high-throughput primer design tool to develop multiple PCR primer sets for each targeted gene. Our tool helps with the design of multiple primers from potentially large reference sets of 3000 sequences or greater. We cast the problem as a variant of the well-known "maximum coverage problem" from computer science. Since this problem has no practical exact solution, we use a "greedy" algorithm to choose a set of primer pairs from the candidates that maximizes the diversity covered by the primer sets. It requires an aligned set of reference sequences as input, with an optional phylogenetic tree for diversity weighting, an optional amplicon size range, and a maximum number of primer pairs to be developed per gene.

During tool testing, we developed new primer sets for nitrogen cycling genes (*amoA*, *nifH*), recalcitrant carbon degradation genes (*cutC*, *cntN*), antibiotic resistance genes (*tet\_sul2*, *tetA-G*), an integrase gene involved in mobile elements (*int11*) and a microbial gene involved in reducing plant stress (*acdS*). We have experimentally validated a set of three non-degenerate primer pairs targeting *cntN* sequences. The sequencing results showed adequate sensitivity, satisfactory amplicon size, and 99% PCR efficiency with the three primers used. This tool is also being employed in the design of primers targeting ACC deaminase for investigations into the role of disease suppressive soil microbial communities in promoting plant health. The current tool is already in use by several research groups. We have developed the tool to be "KBase ready" and intend to help integrate this functionality both into our FunGene and into KBase.

# Associative Nitrogen Fixation across a Nitrogen Input Gradient

Sarah S. Roley,<sup>1,2</sup> David S. Duncan,<sup>3,4</sup> Di Liang,<sup>1,2,5</sup> Aaron Garoutte,<sup>1,6</sup> Randall D. Jackson,<sup>3,4</sup> James M. Tiedje,<sup>1,5,6</sup> and **G. Philip Robertson**\*<sup>1,2,5</sup> (robert30@msu.edu)

<sup>1</sup>Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI; <sup>2</sup>WK Kellogg Biological Station, Michigan State University, Hickory Corners, MI; <sup>3</sup>Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI; <sup>4</sup>Department of Agronomy, University of Wisconsin-Madison, Madison, WI; <sup>5</sup>Department of Crop, Soil and Microbial Sciences, Michigan State University, East Lansing, MI; <sup>6</sup>Center for Microbial Ecology, Michigan State University, East Lansing, MI

https://www.glbrc.org/research/sustainability

# **Project Goals:**

The overall project goal is to ascertain controls on microbial nitrogen (N) cycling processes in the switchgrass rhizosphere. More specifically, we aim to determine how associative nitrogen fixation (ANF) rates change with N availability, through the growing season, and across soil types. This information will help us estimate the annual N inputs from ANF. If the ANF contributions are substantial, it could mean that switchgrass can be grown with lower fertilizer N inputs, ultimately improving its economic and environmental sustainability.

# Abstract text

Associative nitrogen fixation (ANF) is the process by which microbes convert dinitrogen gas  $(N_2)$  to ammonia  $(NH_3)$  in a loose association with plants. ANF has been studied primarily in tropical systems, where it can contribute substantial amounts of N to sugar cane<sup>1</sup> and other grasses<sup>2</sup>. In temperate ecosystems, ANF has been documented in several species, but far less is known about its response to external N inputs, its contribution to ecosystem N inputs, or how its rates change with the growing season.

Switchgrass (*Panicum virgatum*) is considered a candidate bioenergy crop, because of its wide native range, its genetic diversity, its large size, and its ability to grow with minimal external inputs<sup>3</sup>. Switchgrass yields are often non-responsive to N fertilizer addition, even with annual harvest and removal of N in biomass<sup>4</sup>. N<sub>2</sub>-fixing microbes can be found in the switchgrass rhizosphere<sup>5</sup> and there is evidence that switchgrass takes up fixed N<sup>6</sup>, but it is unknown if ANF is sufficient to meet switchgrass' annual N needs. In addition, the controls on ANF in switchgrass are not well understood.

To examine the controls on ANF in switchgrass, we measured ANF via in vitro  ${}^{15}N_2$  incubations of switchgrass rhizosphere soil and roots<sup>7</sup>. We obtained the soil and roots from switchgrass plots grown in fertile mollisols in Wisconsin, USA, and from moderately-fertile alfisols in Michigan, USA. At both sites, the switchgrass was grown under each of 3 fertilizer levels: unfertilized, 56 kg N ha<sup>-1</sup> yr<sup>-1</sup> (the recommended N fertilizer amount) and 196 kg N ha<sup>-1</sup> yr<sup>-1</sup>. We measured ANF

4 times, at relevant stages of plant phenology, including 1) prior to fertilizer addition; 2) 2-3 weeks after fertilizer addition; 3) at switchgrass flowering; and 4) after plant senescence.

As expected, we found that both soil and root ANF declined with N addition; the highest ANF rates occurred in the unfertilized plots. In soils, these patterns were only evident after fertilizer was applied and disappeared by plant senescence, but in roots, those patterns persisted in all time periods. This suggests that soil fixation is responsive to rapid changes in soil N availability, while root fixation is influenced by the legacy of soil N availability, perhaps as manifested through changes in the microbial community. Indeed, soil ANF was correlated with inorganic N concentration and N mineralization, but root fixation was not correlated with any metrics of N availability, except fertilizer addition rate.

We also expected that the highest ANF rates would occur during mid-growing season, when plant N demand is highest. Instead, root ANF was highest post-senescence. This surprising pattern suggests that root diazotrophs are taking advantage of a surge of carbon (C) available post-senescence, and they may not be mutualistic or may only be seasonally cooperative.

To estimate whether ANF could meet switchgrass' annual N needs, we scaled the soil and root rates to the full growing season. We found that in the unfertilized plots, root fixation could account for  $0.71 \pm 0.2$  kg N ha<sup>-1</sup> yr<sup>-1</sup> and soil fixation could account for  $47 \pm 7.5$  kg N ha<sup>-1</sup> yr<sup>-1</sup>. In comparison, the net amount of N removed in the annual harvest was  $51.1 \pm 10.1$  kg N ha<sup>-1</sup> yr<sup>-1</sup>. The combination of root and soil ANF is thus equivalent to 93% of switchgrass' annual N needs and potentially explains the lack of yield response to N fertilizer addition.

## References

1. Urquiaga S, Xavier RP, Morais RF, Batista RB, Schultz N, Leite JM, Maia e Sá J, Barbosa KP, Resende AS, Alves BJR, Boddey RM. 2011. Evidence from field nitrogen balance and 15N natural abundance data for the contribution of biological N2 fixation to Brazilian sugarcane varieties. Plant and Soil 356: 5-21.

2. de Morais, R. F., D. M. Quesada, V. M. Reis, S. Urquiaga, B. J. R. Alves, and R. M. Boddey. 2012. Contribution of biological nitrogen fixation to Elephant grass (Pennisetum purpureum Schum.). Plant and Soil **356**:23-34.

3. McLaughlin SB, Kszos LA. 2005. Development of switchgrass (Panicum virgatum) as a bioenergy feedstock in the United States. Biomass & Bioenergy 28: 515-535.

4. Parrish DJ, Fike JH. 2005. The biology and agronomy of switchgrass for biofuels. Critical Reviews in Plant Sciences 24: 423-459.

5. Bahulikar RA, Torres-Jerez I, Worley E, Craven K, Udvardi MK. 2014. Diversity of Nitrogen-Fixing Bacteria Associated with Switchgrass in the Native Tallgrass Prairie of Northern Oklahoma. Applied and Environmental Microbiology 80: 5636-5643.

6. Morris DR, Zuberer DA, Weaver RW. 1985. Nitrogen fixation by intact grass soil cores using (N-2)-N-15 and acetylene reduction Soil Biology & Biochemistry 17: 87-91.

7. Gupta VVSR, Kroker SJ, Hicks M, Davoren CW, Descheemaeker K, Llewellyn R. 2014. Nitrogen cycling in summer active perennial grasses in South Australia: non-symbiotic nitrogen fixation. Crop and Pasture Science 65: 1044-1056.

*This work was funded in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and DOE OBP Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830).* 

Utility of a Sorghum bicolor RNA atlas for improving energy sorghum stem composition

Brian McKinley, Bill Rooney, Jeremy Schmutz, Curt Wilkerson, John Ralph and John Mullet

Photoperiod sensitive energy sorghum hybrids accumulate ~20-40 Mg of dry biomass per hectare during ~200 day growing seasons. Approximately 80% of energy sorghum's harvestable biomass is located in plant stems that are 4-5 meters in length and ~20-40 mm in diameter. The composition of sorghum stems has been characterized by NIR, HPLC, and NMR at different stages of plant development, at various times during the growing season, and in well watered and water-limited environments. The analysis showed a wide range of stem biomass composition, variation in the ratio of cell wall to non-structural carbohydrates, and dynamic changes in composition during development. Sorghum stems accumulate sucrose, glucose, fructose, starch, and mixed linkage glucans that can be efficiently converted to a wide range of biofuels and bioproducts at low cost. Stem structural carbohydrates, especially from *bmr*-genotypes with modified lignin chemistry, can also be converted to products following pretreatments. Further improvement in energy sorghum stem composition through pathway engineering requires in depth knowledge of target genes, pathways, and their regulation. To obtain this information, an RNA Atlas of gene expression profiles from sorghum tissues and developmental states was collected in part through a JGI Community Sequencing Project. Results from transcriptome analysis of sorghum stems enabled the identification of >200 genes involved in cell wall biosynthesis, and key genes/pathways that regulate the accumulation of sucrose in sorghum stems. The database is being used to identify promoters for engineering stem composition in a tissue specific manner at optimal times during sorghum development.

Partitioning Nitrous Oxide (N<sub>2</sub>O) Emissions from Ammonia Oxidizing Bacteria (AOB) and Ammonia Oxidizing Archaea (AOA) in Corn and Switchgrass Ecosystems

Di Liang<sup>1,2,3\*</sup>(liangdi@msu.edu) and G Philip Robertson<sup>1,2,3\*</sup>

<sup>1</sup>Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI; <sup>2</sup>WK. Kellogg Biological Station, Michigan State University, Hickory Corners, MI; <sup>3</sup>Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI

https://www.glbrc.org/research/sustainability

# **Project Goals:**

Our goal in this project is to understand the relative importance of ammonia oxidizing bacteria (AOB) and ammonia oxidizing archaea (AOA) as sources of nitrate and nitrifierderived nitrous oxide (N<sub>2</sub>O) in potential bioenergy cropping systems. In particular we aim to understand how AOA and AOB respond to nitrogen fertilizer in switchgrass vs. corn cropping systems. We use 1-octyne, a recently developed and tested chemical inhibitor of AOB (Taylor *et al.*, 2013, Taylor *et al.*, 2015) to distinguish activities and emissions from AOA vs. AOB.

# Abstract:

Nitrous oxide (N<sub>2</sub>O) is a potent greenhouse gas with a global warming potential 300 times higher than CO<sub>2</sub>. The atmospheric N<sub>2</sub>O abundance has increased from a preindustrial baseline of 270 to 329 ppb<sub>v</sub> as of 2016 (<u>http://cdiac.ornl.gov</u>, accessed on December 20, 2016), largely due to human activities. N<sub>2</sub>O also reacts with electronically excited oxygen atoms O (<sup>1</sup>D) in the stratosphere to form nitric oxide (NO), which catalyzes ozone depletion (Portmann *et al.*, 2012). Gross anthropogenic N<sub>2</sub>O emissions of 6.2 Tg N<sub>2</sub>O-N yr<sup>-1</sup> are dominated by agriculture (Davidson and Kanter, 2014). As agricultural soils contribute N<sub>2</sub>O more than any other anthropogenic components, understanding sources of N<sub>2</sub>O from agricultural soils is critically important for developing N<sub>2</sub>O mitigation practices (Paustian *et al.*, 2016).

Soil N<sub>2</sub>O emissions mainly result from two microbial processes including denitrification and nitrification. Nitrification converts ammonia (NH<sub>3</sub>) to nitrite (NO<sub>2</sub><sup>-</sup>) and nitrate (NO<sub>3</sub>), with N<sub>2</sub>O as a byproduct (Robertson and Groffman, 2015). Nitrification is performed mainly by two taxa: ammonia oxidizing bacteria (AOB) and ammonia oxidizing archaea (AOA). Recent studies have confirmed that both AOA and AOB are able to produce N<sub>2</sub>O (Stieglmeier *et al.*, 2014). However, the relative contribution of AOA and AOB to N<sub>2</sub>O is still unclear. Our understanding of the mechanisms regulating N<sub>2</sub>O production from AOA and AOB is hindered mainly by the difficulties of isolating pure cultured ammonia oxidizers. For example, there is currently only one pure culture of soil AOA (*Nitrososphaera viennensis EN76*) available for physiological studies. In this study, we examined the seasonal responses of AOA and AOB following nitrogen fertilizer application to annual (corn) and perennial (switchgrass) cropping systems. We took soil samples before and 1, 5, 13, 21, and 42 days after fertilization in late fall of 2015 and spring and summer of 2016. Using 1-octyne as a selective inhibitor targeting AOB in short-term (24h) incubations, we were able to separate the  $N_2O$  and  $NO_3^-$  contributed by AOA and AOB, respectively.

Our results show that: 1) In both corn and switchgrass ecosystems, AOB were the main contributors to nitrate accumulation in spring and summer (60-75% of accumulated nitrate) while in late fall AOA and AOB contributed equally; 2) AOB dominated N<sub>2</sub>O emissions (> 80%) in spring and summer in both corn and switchgrass ecosystems, but in late fall AOA and AOB contributed equally to N<sub>2</sub>O in corn while AOB were the major N<sub>2</sub>O contributors (60%) in switchgrass ecosystems; 3) AOB but not AOA produced more nitrate in response to fertilization in both corn and switchgrass; the AOB response lasted 42 days; and 4) both AOA and AOB produced more N<sub>2</sub>O in response to fertilization in corn but not in switchgrass.

We conclude that AOA and AOB have different responses to nitrogen fertilization in different cropping systems. In most seasons more nitrate and  $N_2O$  was produced by AOB than AOA, and AOB also tended to be more responsive to fertilizer additions.

### References

- Ravishankara AR, Daniel JS, Portmann RW. Nitrous oxide (N<sub>2</sub>O): the dominant ozone-depleting substance emitted in the 21st century. *Science*, 2009, 326(5949):123-5, doi: 10.1126/science.1176985.
- Portmann RW, Daniel JS, Ravishankara AR. Stratospheric ozone depletion due to nitrous oxide: influences of other gases. *Philos Trans R Soc Lond B Biol Sci*, 2012, 367(1593):1256-64, doi: 10.1098/rstb.2011.0377.
- 3. Davidson EA and Kanter D. Inventories and scenarios of nitrous oxide emissions. *Environ. Res. Lett*, 2014, 9:105012, doi:10.1088/1748-9326/9/10/105012.
- 4. Paustian K, Lehmann J, Ogle S, Reay D, Robertson GP, Smith P. Climate-smart soils. *Nature*, 2016, 532(7597):49-57, doi: 10.1038/nature17174.
- Robertson, GP and Groffman PM. Nitrogen transformations, 2015, Pages 421-446 in E. A. Paul, editor. Soil microbiology, ecology and biochemistry. Fourth edition. Academic Press, Burlington, Massachusetts, USA.
- 6. Stieglmeier M, Mooshammer M, Kitzler B, Wanek W, Zechmeister-Boltenstern S, Richter A, Schleper C. Aerobic nitrous oxide production through N-nitrosatin hybrid formation in ammonia-oxidizing archaea. *ISME J*, 2014, 8(5):1135-46, doi: 10.1038/ismej.2013.220.
- Taylor AE, Vajrala N, Giguere AT, Gitelman AI, Arp DJ, Myrold DD, Sayavedra-Soto L, Bottomley PJ. Use of aliphatic n-alkynes to discriminate soil nitrification activities of ammoniaoxidizing thaumarchaea and bacteria. *Appl Environ Microbiol*, 2013, 79(21):6544-51, doi: 10.1128/AEM.01928-13.
- Taylor AE, Taylor K, Tennigkeit B, Palatinszky M, Stieglmeier M, Myrold DD, Schleper C, Wagner M, Bottomley PJ. Inhibitory effects of C2 to C10 1-alkynes on ammonia oxidation in two Nitrososphaera species. *Appl Environ Microbiol*, 2015, 81(6):1942-8, doi: 10.1128/AEM.03688-14.

This work was funded in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and DOE OBP Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830).

# Characterization of *Populus* Transgenic Plants Overexpressing PtDUF266A (OXPtDUF266A) and Biofuel Production

Chang Geun Yoo<sup>1\*</sup> (<u>vooc@ornl.gov</u>), Yongil Yang,<sup>1</sup> Yunqiao Pu,<sup>1</sup> Hao-Bo Guo,<sup>2</sup> William Rottmann,<sup>4</sup> Kimberly A Winkeler,<sup>3</sup> Cassandra M Collins,<sup>3</sup> Lee E. Gunter,<sup>1</sup> Sara S Jawdy,<sup>1</sup> Xiaohan Yang,<sup>1</sup> Hong Guo,<sup>2</sup> Gerald A Tuskan,<sup>1</sup> Jin-Gui Chen,<sup>1</sup> Arthur J Ragauskas,<sup>1,2</sup> and **Paul Gilna**<sup>1</sup>

<sup>1</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>University of Tennessee, Knoxville; <sup>3</sup>ArborGen Inc., Ridgeville, South Carolina

http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Determining the function of proteins containing Domains-of-Unknown Function (DUF) is a challenge once their phenotypic value is identified. In this study, one member of *Populus* DUF266 proteins, PtDUF266A, was investigated as a potential candidate for improved biofeedstocks for biofuel production. DUF266-containing proteins have been considered as 'not classified glycosyltransferases (GTnc)' because they are phylogenetically distinct from the GTs. In specific *Populus* transgenic plants overexpressing DUF266-containing proteins (*OXPtDUF266A*) were characterized by diverse analysis approaches including (1) wet chemical bulk analysis for chemical composition, (2) anthrone assay for cellulose content, (3) gel permeation chromatography (GPC) analysis for molecular weights of lignin and cellulose, (4) gene expression analysis for effects on biomass biosynthesis gene regulation, and (5) nuclear magnetic resonance (NMR) analysis for structural information of lignin.

The results showed that the *OXPtDUF266A* were larger than WV94 under the same growing conditions. In the *OXPtDUF266A*, the glucose and cellulose contents notably increased. In addition, degree of polymerization of cellulose in the OXPtDUF266A transgenic plants also increased. However, its cellulose crystallinity index was not changed. Based on gene expression analysis, cellulose biosynthesis-related genes such as *CESA* and *SUSY* were upregulated in the *OXPtDUF266A* transgenic plants. The gene expression analysis results indicate that overexpression of PtDUF266A induced more expression of cellulose biosynthesis genes. Besides the cellulose, physicochemical

properties of lignin in the *OXPtDUF266A* transgenic plants were also analyzed by NMR and GPC analyses. The lignin molecular weights and lignin S/G ratio in the transgenic plants were not notably changed.

In addition, to evaluate the *OXPtDUF266A* transgenic plants as a suitable feedstock for biofuel production, sugar release of the transgenic plants was tested. Compared to the wild-type plants, the *OXPtDUF266A* transgenic plants showed more sugar release (38% higher). Increase of total cellulose content in the *OXPtDUF266A* transgenic plants resulted in an increase of sugar release. Overall, the overexpression of *PtDUF266A* showed great potential as a feedstock for biomass utilization.

# Defined Tetra-Allelic Gene Disruption of the 4-Coumarate:Coenzyme A ligase 1 Gene by CRISPR/Cas9 in Switchgrass Results in Lignin Reduction and S/G Ratio Alteration

Jongjin Park<sup>1,2\*</sup> (<u>jpark@noble.org</u>), Zeng-Yu Wang,<sup>1,2</sup> Chang Geun Yoo,<sup>2</sup> Emma Dempewolf,<sup>1</sup> Yunqiao Pu,<sup>2</sup> Smriti Debnath,<sup>1</sup> Arthur J Ragauskas,<sup>2,3</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>The Samuel Roberts Noble Foundation, Ardmore, Oklahoma; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>University of Tennessee, Knoxville

# http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

The production of biofuels from renewable biomass alleviates the dependence on fossil fuels.

Switchgrass *(Panicum virgatum)*, a C4 perennial grass species, has been developed into a lignocellulosic feedstock for bioenergy. In order to reduce cell wall recalcitrance and improve bioethanol production, RNAi knock-down technology has been used to suppress genes of interests in switchgrass. In recent years, genome editing methods including zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated proteins 9 (Cas9), have been developed to knock-out specific genes by altering the genomic DNA sequences (Bibikova et al., 2003; Boch et al., 2009; Kim and Kim, 2014). We employed the CRISPR/Cas9 system to produce low-lignin switchgrass because the CRISPR/Cas9 system is relatively simple to work with compared to ZFNs and TALENs.

Switchgrass is an outcrossing species with a complex allo-tetraploid genome (2n = 4x = 36), which causes difficulties in producing homozygote knock-out plants.

Lignin is a major component of secondary cell walls and contributes to the recalcitrance problem during fermentation. In order to develop a CRISPR/Cas9 system in switchgrass, we choose to target 4-Coumarate:coenzyme A ligase (4CL), a key enzyme involved in the early steps of the monolignol biosynthesis. We identified three 4CL genes, Pv4CLJ, Pv4CL2 and Pv4CL3 in switchgrass.

qRT-PCR analysis revealed that *Pv4CLJ* transcripts were more abundant in the internode and the node than in the leaf. *Pv4CL2* transcripts were barely detectable in the three different tissues – internode, node, and leaf. *Pv4CL3* was preferentially expressed in leaf. Internode and node are highly lignified tissues comprised of parenchyma and sclerenchyma cells distributed in the interfascicular region and the vascular sheath.

Therefore, *Pv4CL1* was selected as the main target gene. Specific gRNA was constructed to target Pv4CL1. After introducing the construct into switchgrass calli, forty plants were regenerated. After PCR screening and sequencing, four plants (*Pv4cl1-#25, #26, #28, and #29*) were confirmed to have the tetra allelic mutations simultaneously. The *Pv4cl1* knock-out plants showed reddish stems and reduced cell wall thickness, and had a 20 to 29% reduction in total lignin. The reduction in lignin content in the *Pv4cl1* knock-out plants led to a 9% and a 28% increases in glucose and xylose releases, respectively. This study demonstrated that we have established the CRISPR/Cas9 system in switchgrass, and the system has been successfully used to precisely target the selected *Pv4CL1* gene to create switchgrass knock-out plants with reduced recalcitrance.

## References

Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764.

- Boch, J., Scholze, H., Schomack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512.
- Kim, H. and Kim, J.S. (2014) A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321-334.

## Evaluation of Multiple Levers for Overcoming the Recalcitrance of Cellulosic Biomass

Evert Holwerda<sup>1,2\*</sup> (evert.k.holwerda.th@dartmouth.edu), Robert M Worthen,<sup>1,2</sup> Ninad Kothary,<sup>2,5</sup> Michael Balch,<sup>1,2</sup> Mark Davis,<sup>2,3</sup> Brian Davison,<sup>2</sup> Mike Himmel,<sup>2,3</sup> Debra Mohnen,<sup>2,4</sup> Charles Wyman,<sup>2,5</sup> Gerald Tuskan,<sup>2</sup> Lee Lynd,<sup>1,2</sup> **Paul Gilna**<sup>2</sup>

<sup>1</sup>Dartmouth College, Hanover, New Hampshire; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>National Renewable Energy Laboratory, Golden, Colorado; <sup>4</sup>University of Georgia, Athens; <sup>5</sup>University of California, Riverside

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

The primary barrier to economically competitive cellulosic biofuels is the resistance of plant cell walls to deconstruction – termed recalcitrance. Overcoming this barrier may be approached via recalcitrance "levers" drawn from three categories:

- I. Start with nature's best with respect to feedstocks and biocatalysts;
- II. Apply biotechnology to improve plants, enzymes, and microbes;
- III. Augment biological capability with non-biological processing.

Here we evaluate the impact of multiple levers in each of these categories both individually and in combination, including:

- three engineered switchgrass plant lines and their parent lines;
- two natural variant Populus lines;
- solubilization mediated by either fungal cellulase in the presence of yeast (SSF), *Caldicellusiruptor bescii*, or *Clostridium thermocellum*; and
- augmentation of biologically-mediated solubilization using either cosolvent-enhanced liquid fractionation (CELF) or co-treatment (i.e., intermittent milling during fermentation).

Key observations include:

- The extent of increased solubilization observed for engineered/variant plant lines is highly dependent upon which biocatalyst is used. Thus, plant recalcitrance exhibits a strong dependence on the biocatalyst by which it is evaluated.
- The relative effectiveness of biocatalysts at mediating solubilization was *C. thermocellum*> *C. bescii* > fungal cellulose.
- The relative impact of various levers at enhancing solubilization was augmentation (co-treatment or CELF) > choice of biocatalyst > choice of feedstock and feedstock modification.
- Total carbohydrate solubilization in excess of 90% is observed for solubilization of all *Populus* and switchgrass lines tested using either co-treatment- or CELF-augmented *C. thermocellum* fermentation.
- Lignin-rich residues remaining after fermentation and co-treatment showed no evidence of chemical modification, in contrast to fermentation residues following thermochemical pretreatment, and appear promising for further processing into coproducts.
- We infer that *C. thermocellum* cultures are able to solubilize all major chemical linkages in switchgrass and *Populus* given physical access to them.
- Our results provide proof of concept for achieving high carbohydrate solubilization without thermochemical pretreatment and without added enzymes.

# Explaining and Predicting Biomass Recalcitrance with Rigidity Percolation Theory

Erica Gjersing<sup>1,2\*</sup> (<u>Erica.gjersing@nrel.gov</u>), Gerald A Tuskan,<sup>2</sup> Wellington Muchero,<sup>2</sup> Brian Davison,<sup>2</sup> Bryon Donohoe,<sup>1</sup> Robert W Sykes,<sup>1,2</sup> Stephen R Decker,<sup>1</sup> Mark F Davis,<sup>1,2</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>National Renewable Energy Laboratory, Golden, Colorado; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

There remains considerable debate on the roles, organization, and interactions of the various polymers within the plant cell wall and how each contributes to the cell wall recalcitrance. We present a model based on Rigidity Percolation theory demonstrating reduced recalcitrance is dependent on compositional variation of the major plant cell wall polymers, lignin, hemicellulose, pectin, and cellulose, from studies using natural variants and genetically-modified feedstocks. Rigidity percolation theory is a subset of general percolation theory that has been used in amorphous materials science to understand how compositional variation and the connectivity of the network structures affect physical properties. The theory was also evaluated over a range of feedstocks including alfalfa, switchgrass, eucalyptus, *Populus*, and pine. Rigidity percolation theory predicts a dramatic change in the physical properties of a system when the number of floppy modes with the network structure approaches zero. For the biomass samples studied here, the predicted change is observed when the plotting average coordination number versus sugar release indicating deconstruction mechanisms are more effective on plant cell walls composed of more floppy elements and hindered by more rigid elements.

This framework can also be used as a predictive tool to understand how changes in composition can influence biomass recalcitrance. Rigidity percolation theory can be used to predict alternate combinations of cell wall polymer composition that could produce low recalcitrant plant lines while supporting normal plant growth. The effective application of rigidity percolation theory will require

moving beyond the current approach of modifying only individual cell wall polymers as a means of reducing recalcitrance.

# Field Experiments of Nine Switchgrass TOP Lines

Holly Baxter<sup>1,2\*</sup> (<u>hbaxter@utk.edu</u>), Mitra Mazarei,<sup>1,2</sup> Charleson Poovaiah,<sup>1,2</sup> Chunxiang Fu,<sup>2,3</sup> Hui Shen,<sup>2,4</sup> Ajaya Biswall,<sup>2,5</sup> Guifen Li,<sup>2,3</sup> Desalegn Serba,<sup>2,3</sup> Kelsey Yee,<sup>2</sup> Alexandru Dumitrache,<sup>2</sup> Jace Natzke,<sup>2</sup> Miguel Rodriguez,<sup>2</sup> Olivia Thompson,<sup>2</sup> Geoffrey Turner,<sup>2,6</sup> Robert Sykes,<sup>2,6</sup> Steve Decker,<sup>2,6</sup> Mark Davis,<sup>2,6</sup> Jonathan Mielenz,<sup>2</sup> Brian Davison,<sup>2</sup> Steven Brown,<sup>2</sup> Malay Saha,<sup>2,3</sup> Yuhong Tang,<sup>2,3</sup> Debra Mohnen,<sup>2,5</sup> Richard Dixon,<sup>2,4</sup> Zeng-Yu Wang,<sup>2,3</sup> C. Neal Stewart, Jr.,<sup>1,2</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>University of Tennessee, Knoxville; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>Samuel Roberts Noble Foundation, Ardmore, Oklahoma; <sup>4</sup>University of North Texas, Denton; <sup>5</sup>University of Georgia, Athens; <sup>6</sup>National Renewable Energy Laboratory, Golden, Colorado

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Switchgrass (*Panicum virgatum* L.) is a perennial warm-season C4 grass that has been identified as a candidate lignocellulosic bioenergy crop because of its rapid growth rate, nutrient use efficiency, and widespread adaptation throughout eastern North America. Cell wall recalcitrance in switchgrass and other lignocellulosic feedstocks is a major economic barrier for enabling efficient enzymatic, microbial, or chemical breakdown of cell wall carbohydrates into fermentable sugars. Recent research has focused on developing switchgrass lines that are more amenable to the fermentation process, either by genetic engineering or by selection of low-recalcitrant lines through association analyses.

Greenhouse studies have identified several potentially successful transgenic routes for reducing cell wall recalcitrance and/or improving growth in switchgrass. An important validation step, especially for genetically engineered plants, is to perform multi-year field studies, which is a vital goal of BESC. It is well known that the greenhouse is not always predictive of crop performance in the field. Herein we present data from nine BESC "TOP Lines" and appropriate controls from agronomically-relevant University of Tennessee (Knoxville) field studies, in which plants were grown under the USDA APHIS BRS release into the environment permits for two or three field

seasons. Data includes: (1) agronomic performance (morphology and end-of-season biomass), (2) lignin content and composition by high-throughput py-MBMS, (3) sugar release by high-throughput enzymatic assays, (4) ethanol yield by separate hydrolysis and fermentation assays, and (5) incidence of switchgrass rust, caused by the pathogen (*Puccinia emaculata*).

COMT (caffeic acid 0-methyltransferase) down-regulated switchgrass lines grown in the field for three years (2011-2013) had consistently lower lignin levels, reduced S/G ratios, and improved sugar release across all three years. By the end of year three, both transgenic lines produced 36--41% more ethanol than controls and produced equivalent biomass as controls with no difference in switchgrass rust incidence. The MYB4 over-expressing plants grown for three years (2012-2014) had decreased lignin, improved sugar release, and improved ethanol yields of up to 50%, with one line also producing 63% more biomass than the control in year two. The MYB4 transgenic lines were similar to the control in rust susceptibility with the exception of line L1, which did not exhibit any rust symptoms for the duration of the experiments. Some miRNAl56 overexpressing lines, which were grown for three years (2013-2015), had decreased lignin content, reduced syringyl/guaiacyl (S/G) ratios, and improved sugar release. One line produced 25-56% more biomass relative to the control across all three years. Rust susceptibility varied significantly among the different lines and among years. Galacturonosyltransferase4 (GAUT4) downregulated lines grown for three years (2013-2015) had altered S/G ratios and improved sugar release. All three transgenic lines produced significantly more biomass than the control, and one line also had a 23% increase in ethanol yield in year one. Although some GAUT4 lines showed increased rust susceptibility relative to the control in year two, no differences among transgenic lines and the control were observed in year three. One *folylpolyglutamate synthase 1* (FPGSI) down- regulated line had 27% higher biomass relative to the control, whereas a second line produced 7% more ethanol with no change in biomass production in year one. No changes in rust susceptibility were observed between FPGS1 transgenic lines and the control in year two. Several natural variant lines grown for two years (2014-2015) had reduced lignin, increased sugar release, and improved biomass production compared with the control, with one line also showing an increase in ethanol yield in year two. One GAUTI down-regulated line produced 35% more biomass than the control at the end of the first growing season (2015), whereas another line had improved sugar release. One ERF/SHN over-expressing line produced 41% more biomass than the control at the end of the first growing season (2015). Data are currently being collected and analyzed for switchgrass lines with over-expression of GA2-ox and downregulation of CWG, both of which were planted in late spring of 2016 and have just completed their first growing season.

# From Gene to Network: Switchgrass TOP Line RNA-seq Data Analysis Pipeline at BESC

Chunman Zuo,<sup>2,3</sup> Xin Chen,<sup>2,3</sup> Manesh Shah,<sup>2</sup> Debbie Weighill,<sup>2,5</sup> Amy Mason,<sup>1,2</sup> Luis L Escamilla-Trevino,<sup>1,4</sup> Hongli Xie,<sup>1,2</sup> Qin Ma,<sup>2,3</sup> Zeng-Yu Wang,<sup>1,2</sup> Richard A Dixon,<sup>2,4</sup> Gerald Tuskan,<sup>2</sup> Wellington Muchero,<sup>2</sup> Tim Tschaplinski,<sup>2</sup> Susan K Holladay,<sup>2</sup> Richard S Nelson,<sup>1,2</sup> Dan Jacobson,<sup>2</sup> Ying Xu,<sup>2,3</sup> and Yuhong Tang<sup>1,2\*</sup> (<u>ytang@noble.org</u>), and **Paul Gilna**<sup>2</sup>

<sup>1</sup>The Samuel Roberts Noble Foundation, Ardmore, Oklahoma; <sup>2</sup>BioEnergy Science Center (BESC), Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>University of Georgia, Athens; <sup>4</sup>University of North Texas, Denton; <sup>5</sup>University of Tennessee, Knoxville

http://bioenergycenter.org/

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Switchgrass (*Panicum virgatum*) has been used for more than a decade as a model crop for sustainable energy research. At BESC, a transformation pipeline was established to identify genes/pathways that improve bioenergy traits of switchgrass. In addition, evaluations of switchgrass natural variants for enhanced biofuel traits were conducted. Twelve transgenic or natural variant lines were identified as BESC switchgrass TOP Lines that had increased sugar release and normal or enhanced plant growth. Systematic analyses, including those of the transcriptome, were carried out for these plants in order to understand the underlying pathways controlling the improved traits.

RNA-seq, as well as other systematic analyses, was conducted on tissue from greenhouse-grown switchgrass TOP Line plants and their controls just entering reproductive stage 1 (R1). The RNA-seq and other TOP Line analysis data are stored in and retrieved from the BESC LIMS. From data quality control (QC) to gene expression analysis for differential expression, each RNA-seq dataset was processed through a compilation of publically available software. Using only reads mapped to one location of the genome, gene level assemblies were generated with HISAT2 and related programs originally with switchgrass genome assembly V1.1 and then later with genome assembly V3.1. To facilitate gene expression visualization at sample and gene levels, the RNA-seq data also are displayed in a BESC Jbrowse portal along with the switchgrass reference genome V1.1 and V3.1.

For each TOP Line, genes whose expression were different from the respective control levels were selected, using differential analysis software, DESeq, as well as an array-like method once an expression matrix was generated. In order to provide a richer biological context for results interpretation, differentially expressed genes were (via the identification of poplar orthologs) projected into our extensive poplar systems biology network models for further analysis. These networks help to identify changes contributing to the cell wall modification in the switchgrass RNAi TOP Lines, providing further genes to target to improve biofuel traits. Significantly modified pathways for cell wall biosynthesis also were identified using Mapman from differentially expressed gene list of the Top Lines. A selection of these genes is being further analyzed for their potential to reduce wall recalcitrance through additional network analyses and considered for functional studies in plants using transient and stable knockdown or overexpression strategies.

# Gene Expression Differences between *Clostridium thermocellum* Biofilm and Planktonic Cells Lead to Specialized Activities and Growth

Alexandru Dumitrache<sup>1</sup>\* (<u>dumitrachea@ornl.gov</u>), Dawn M Klingeman,<sup>1</sup> Jace Natzke,<sup>1</sup> Miguel Rodriguez,<sup>1</sup> Richard J Giannone,<sup>1</sup> Robert L Hettich,<sup>1</sup> Brian H Davison,<sup>1</sup> Steven D Brown<sup>1</sup> and **Paul Gilna**<sup>1</sup>

<sup>1</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

# http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Concerns around carbon and energy security have encouraged efforts to develop alternative fuel sources. Our goal was to present a fundamental and novel study into adherent biofilms and the non-adherent planktonic cells of the anaerobic thermophile *Clostridium thermocellum*, which is a model biofuel bacterium for direct conversion of lignocellulosic plant biomass. Unlike canonical biofilms that feed from the flux-limited diffusion of soluble nutrients and where non-adherent cells have prime access to labile carbon, *C. thermocellum* utilizes cellulose as the solid attachment surface which provides the only carbon source for energy and growth, thus placing planktonic cells at an inherent disadvantage. Expression studies that compare and contrast the biofilm lifestyle for *C. thermocellum* cells have not been reported, that we are aware of, and this growth mode represents the primary cell type for industrially relevant applications.

This model biofuel bacterium forms biofilms adherent to lignocellulosic feedstocks in a continuous cell-monolayer in order to efficiently break down and uptake cellulose hydrolysates. We developed a novel bioreactor design to generate separate sessile and planktonic cell populations for 'omics studies. Sessile cells had significantly greater expression of genes involved in the catabolism of carbohydrates through glycolysis, pyruvate fermentation and ATP generation by proton gradient; the anabolism of proteins and lipids and the cellular functions critical for cell division consistent with substrate replete conditions. Planktonic cells had notably higher gene expression for flagellar motility and chemotaxis, cellulosomal cellulases and anchoring scaffoldins, and a range of stress-

induced homeostasis mechanisms such as oxidative stress protection by antioxidants and flavoprotein co-factors, methionine repair, Fe-S cluster assembly and repair in redox proteins, cell growth control through tRNA thiolation, recovery of damaged DNA by nucleotide excision repair and removal of terminal proteins by proteases.

Using well-controlled and novel bioreactors, the current study demonstrates that microbial attachment and access to the solid carbon source produced widespread gene expression changes in the bacterium. This study provides insights into population heterogeneity within a constrained system which is of broad interest. Most notably, 59% of the protein coding genome recorded a minimum two-fold change in gene expression between biofilm and planktonic cell populations. The more productive sessile cells focus on both conversion and growth while the planktonic cells are more stressed – and are present in higher numbers at the end of the fermentation. We provided succinct summaries for key physiological insights and detailed information for researchers interested in the finest details. RNA-Seq data were verified by RT-qPCR and supporting proteomics data.

Insight into these cellular adaptations stands to benefit future genetic studies and strain engineering of industrial-ready phenotypes and contributes to our fundamental knowledge about adherent cellulolytic microbes.

## Investigating the Role of Polysaccharide Methylation in the Plant Cell Wall

Abilgail Agyeman,<sup>1,2</sup> Seungwon Jang,<sup>1</sup> Maria J Peña,<sup>1,2</sup> Malcolm O'Neill,<sup>1,2</sup> William S York,<sup>1,2</sup> Breeanna Urbanowicz<sup>1,2\*</sup> (<u>breeanna@uga.edu</u>), and **Paul Gilna**<sup>2</sup>

<sup>1</sup>University of Georgia, Athens; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

One of the core goals of the Plant Biomass Formation and Modification Focus Area in BESC is to understand the molecular mechanisms that control biomass recalcitrance. The thick recalcitrant secondary cell walls of Populus are principally composed of cellulose, glucuronoxylan (GX), and lignin with lesser amounts of pectic polysaccharides and glycoproteins. Interactions among these components are believed to be responsible for the resistance of plant biomass to enzymatic deconstruction to fermentable sugars. The GX present in hardwoods including *Populus* and in mature stems of the model plant Arabidopsis thaliana has a backbone composed of 1.4-linked  $\beta$ -D-xylosyl (Xyl) residues that are often substituted at O-2 with glucuronic acid (GlcA) or 4-O-methyl glucuronic acid (MeGlcA). Recently, we identified a new gene (GXMT1) which belongs to a family in Arabidopsis and encodes a xylan methyltransferase. Plants carrying a mutation in this gene synthesize xylan in which the degree of GlcA methylation is reduced to ~25% of wild-type levels. Our previous results (Urbanowicz et al., 2012), suggests that O-methylation plays a role in the polysaccharide's ability to associate with lignin and or other glycopolymers and thereby impacts biomass recalcitrance. Reducing O-methylation by directed breeding or genetic manipulation has considerable potential to affect the recalcitrance of lignocellulosic biomass by modulating the interactions of glucuronoxylan with the other components of the secondary cell wall without any negative effects on plant development or fitness. However, our knowledge of the enzymes involved in polysaccharide methylation and the role of these non-glycosyl substituents in plant cell wall untrastructure remains enigmatic.

As part of an expanded investigation into the identification and characterization of polysaccharide methyltransferase gene families, we isolated homozygous T-DNA insertion *Arabidopsis* mutant lines for several candidates that contain methyltransferase motifs that are highly expressed during both primary and secondary cell wall formation. Cell wall material from mutant plants was fractionated to enrich for different polysaccharides including pectins, arabinogalactans (AGPs) and glucuronoxylan. Fractions enriched in glucuronoxylan were analyzed from mutants in genes expressed during secondary cell wall formation to quantify structural differences with a focus on xylan *O*-methylation. To identify structural changes in the non-hemicellulosic cell wall glycopolymers we developed a robust the fractionation procedure that yielded high amounts of pure, intact soluble polymers. The neutral and acidic sugar content of the isolated glycopolymers was analyzed using an optimized HPAEC-PAD method, which discriminates methylated verses unmethylated saccharides. To confirm that the changes in methylation were specific, NMR structural characterization was performed. Our discovery and characterization of new methyltransferases extends the portfolio of structural targets that can be modified to increase the economic value of lignocellulosic biomass by modulating biopolymer interactions in the cell walls of biomass crops such as poplar and switchgrass.

## References

Urbanowicz, Breeanna R., et al., "4-O-methylation of glucuronic acid in Arabidopsis glucuronoxylan is catalyzed by a domain of unknown function family 579 protein." *Proceedings of the National Academy of Sciences* 109.35 (2012): 14253-14258.

## Lignin Valorization through Biological Funneling

Gregg T Beckham<sup>1,5\*</sup> (<u>gregg.beckham@nrel.gov</u>), Jeffrey C Cameron,<sup>2</sup> Fang Chen,<sup>3</sup> Richard A Dixon,<sup>3</sup> Bryon S Donohoe,<sup>1</sup> Lindsay D. Eltis,<sup>4</sup> Adam M Guss,<sup>5</sup> Christopher W Johnson,<sup>1</sup> Eric M Karp,<sup>1</sup> Rui Katahira,<sup>1</sup> Kristin Moore,<sup>2</sup> Sandra Notonier,<sup>1</sup> Davinia Salvachua,<sup>1</sup> Derek R Vardon,<sup>1</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>National Renewable Energy Laboratory, Golden, Colorado; <sup>2</sup>University of Colorado at Boulder; <sup>3</sup>University of North Texas, Denton; <sup>4</sup>University of British Columbia, Canada; <sup>5</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

# http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Lignin structure and composition vary significantly across plant species. Regardless of the feedstock type, in many modern biorefinery designs, lignin is typically not utilized for fuels and chemicals production, but instead is slated to be burned for process heat because its inherent heterogeneity and recalcitrance make it difficult to selectively valorize.<sup>1</sup> Indeed, despite many decades of lignin depolymerization research, most catalytic strategies to break down lignin yield a highly heterogeneous slate of aromatic compounds.

In nature, some microbes have evolved catabolic pathways that enable the utilization of lignin-derived aromatic molecules as carbon and energy sources. Aromatic catabolism typically occurs via upper pathways that act as a "biological funnel" to convert heterogeneous lignin-derived substrates to central intermediates, such as protocatechuate or catechol. These compounds subsequently undergo ring cleavage and are further converted, often via the  $\beta$ -ketoadipate pathway, to central carbon metabolism. Recently, we employed a natural aromatic-catabolizing organism, *Pseudomonas putida* KT2440, to demonstrate that these metabolic pathways can be harnessed and engineered to convert both aromatic model compounds and heterogeneous, lignin-enriched streams into value-added compounds.<sup>2-5</sup> To make this concept of biological funneling a reality will require systems-level understanding of the catabolic pathways used by microbes to convert lignin to monomers and then convert these monomers to value-added compounds.<sup>6</sup>

Here, we will present several insights into lignin depolymerization and aromatic catabolism by *P. putida* KT2440 and other aromatic-catabolic microbes. From a time-resolved proteomics experiment in a real-world, lignin-rich substrate with an analysis of both the extracellular and intracellular fractions, we identified multiple enzymes that are exclusively expressed and secreted in the substrate that are known to cleave bonds in lignin. Additionally, from both proteomics and transcriptomics, we have identified multiple enzymes of interest for localization studies to better understand the flow of aromatic carbon from lignin to central carbon metabolism in the context of cellular structure. To date, we have validated a GFP-based labeling approach with a known aromatic catabolic enzyme. A localization study is ongoing from the extracellular enzymes through the upper aromatic catabolic pathways and the  $\beta$ -ketoadipate pathway.

Additionally, we are examining the ability of *P. putida* to catabolize new types of lignin substrates. Specifically, caffeoyl alcohol has been recently shown to be a naturally occurring polymer in seed coats of several plant species.<sup>7-8</sup> These monomers form homogeneous, linear ether-linked lignin (dubbed C-lignin), which represents a promising new avenue for engineering homogeneous plant lignin for more effective valorization approaches. Using a C-lignin extract, we have demonstrated that *P. putida* can depolymerize and catabolize C-lignin-rich substrates, demonstrating the ability to potentially use microbial strategies to valorize these novel lignins. Overall, this work demonstrates that the use of microbial aromatic catabolism coupled to advances in lignin chemistry *in planta* may one day enable an approach to valorize lignin by overcoming its inherent heterogeneity to produce fuels, chemicals, and materials.

### References

- 1. Ragauskas, A. J. et al., Lignin valorization: Improving lignin processing in the Biorefinery. *Science* **344**, 1246843 (2014).
- 2. Linger, J. G. et al., Lignin valorization through integrated biological funneling and chemical catalysis. *Proc. Natl. Acad. Sci.* **111**, 12013-12018 (2014).
- 3. Vardon, D.R. et al., Adipic acid production from lignin. Energy Environ. Sci. 8, 617-628 (2015).
- 4. Salvachúa, D., Karp, E. M., Nimlos, C. T., Vardon, D. R. and Beckham, G. T., Towards lignin consolidated bioprocessing: Simultaneous lignin depolymerization and product generation by bacteria. *Green Chem.* **17**, 4951-4967 (2015).
- 5. Vardon, D.R. et al., cis,cis-Muconic acid: Separation and catalysis to bio-adipic acid for nylon-6,6 polymerization. *Green Chem.* **18**, 3397-3413 (2016).
- Beckham, G. T., Johnson, C. W., Karp, E. M., Salvachua, D. and Vardon, D.R., Opportunities and challenges in biological lignin valorization. *Current Opinion in Biotechnology* 42, 40-53 (2016).
- 7. Chen, F., Tobimatsu, Y., Havkin-Frenkel, D., Dixon, R. A. and Ralph, J., A polymer of caffeyl alcohol in plant seeds. *Proceedings of the National Academy of Sciences* **109**, 1772-1777 (2012).
- 8. Chen, F. et al., Novel seed coat lignins in the Cactaceae: structure, distribution and implications for the evolution of lignin diversity. *The Plant Journal* **73**, 201-211(2013).

# Metabolic Network Modeling of *Clostridium thermocellum* for Systems Biology and Metabolic Engineering

R. Adam Thompson,<sup>2,3</sup> Sergio Garcia,<sup>1,2</sup> Sanjeev Dahal,<sup>2</sup> Intawat Nookaew,<sup>2</sup> Donovan S. Layton,<sup>1,2</sup> Adam M. Guss,<sup>2,3</sup> Dan G. Olson,<sup>2,4</sup> Lee R. Lynd,<sup>2,4</sup> Cong T. Trinh<sup>1,2,3\*</sup> (<u>ctrinh@utk.edu</u>), and **Paul Gilna**<sup>2</sup>

<sup>1</sup>University of Tennessee, Knoxville; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>Bredesen Center for Interdisciplinary Research and Education, Knoxville, Tennessee; <sup>4</sup>Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

*Clostridium thermocellum* is a gram-positive thermophile that can directly convert lignocellulosic material into commercially relevant chemicals such as biofuels and biochemicals. Its metabolism contains many branches and redundancies, which limit the production of biofuels and biochemicals at industrially relevant yields and titers. In order to guide the experimental efforts required to overcome these barriers, we built two models of C. thermocellum metabolism. Through an extensive literature review, we first constructed a model of the core metabolism of *C. thermocellum*. This model was experimentally validated and served to investigate the range of phenotypes of *C. thermocellum* in response to significant perturbation of energy and redox pathways. The results revealed a complex, robust redox metabolism of C. thermocellum. By incorporating experimental data into this core model, we identified redox bottlenecks hindering high-yield ethanol production in C. thermocellum.<sup>1</sup> With the recently published sequence of a genetically-tractable strain C. thermocellum DSM 1313, the KEGG database as a scaffold, and further literature review, we expanded the core model into a genome scale model (iAT601).<sup>2</sup> This model constitutes a knowledge base for the organism, including detailed metabolic information, as well as gene protein reaction association. These features allow us to conduct studies on the impact of secondary metabolisms, isozymes, media composition, and provide a more solid basis for computational strain design. We used several sets of experimental data to train the model, e.g., estimation of the adenosine

triphosphate (ATP) requirement for growth-associated maintenance (13.5 mmol ATP/g DCW/hr) and cellulosome synthesis (57 mmol ATP/g cellulosome/hr). Using our tuned model, we (i) predicted the experimentally observed differences in cell biomass yield based on which cellodextrin species is assimilated,<sup>2</sup> (ii) analyzed the experimentally quantified differences in fermentation profiles (i.e., the ethanol to acetate ratio) between cellobiose- and cellulose-grown cultures, for which we inferred potential regulatory mechanisms to explain the phenotypic differences,<sup>2</sup> (iii) elucidated growth cessation and overflow metabolism in *C. thermocellum* DSM1313 at high cellulose loading, and (iv) designed over 250 genetic modification strategies with the potential to optimize ethanol production, 6,155 for hydrogen production, and 28 for isobutanol production. Our developed genome-scale model iAT60I can serve as a high-quality platform for accurately predicting complex cellular phenotypes under a variety of conditions as well as model-guided rapid strain engineering to produce industrial biofuels and chemicals of interest.<sup>3,4</sup>

## References

- Thompson RA, Layton, D.S., Guss, A.M., Olson, D.G., Lynd, L.R., Trinh, C.T., Elucidating Central Metabolic Redox Obstacles Hindering Ethanol Production in *Clostridium thermocellum. Metab. Eng.* 2015, **32**:207-219.
- Thompson RA, Dahal, S., Garcia, S., Nookaew, I., Trinh, C.T., Exploring Complex Cellular Phenotypes and Model-Guided Strain Design with A Novel Genome-scale Metabolic Model of *Clostridium thermocellum* DSM 1313 Implementing an Adjustable Cellulosome. *Biotechnol. Biofuels* 2016, 9:194.
- 3. Trinh CT, Liu, Y., Conner, D., Rational Design of Efficient Modular Cells. *Metab. Eng.* 2015, **32**:220-231.
- 4. Flowers D, Thompson, R.A., Birdwell, D., Wang, T., Trinh, C.T., SMET: Systematic Multiple Enzyme Targeting-A Method to Rationally Design Optimal Strains for Target Chemical Overproduction. *Biotechnology Journal* 2013,**8**:605-618.

# Modification of Pectin Biosynthesis Leads to Higher Biomass Yield and Saccharification in Bioenergy Feedstock

Ajaya K Biswal<sup>1,2\*</sup> (<u>biswalajaya@ccrc.uga.edu</u>), Melani A. Atmodjo,<sup>1,2</sup> Mitra Mazarei,<sup>2,3</sup> Peter R LaFayette,<sup>1,2</sup> Li Tan,<sup>1,2</sup> Sushree S Mohanty,<sup>1,2</sup> David Ryno,<sup>1,2</sup> Ivana Gelineo-Albersheim,<sup>1,2</sup> Kimberly Hunt,<sup>2,3</sup> Robert W Sykes,<sup>2,4</sup> Erica L Gjersing,<sup>2,4</sup> Geoffrey B Turner,<sup>2,4</sup> Steve R Decker,<sup>2,4</sup> Wayne A Parrott,<sup>1,2</sup> Michael K Udvardi,<sup>2,5</sup> Mark F Davis,<sup>2,4</sup> C. Neal Stewart, Jr.,<sup>2,3</sup> Debra Mohnen,<sup>1,2</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>University of Georgia, Athens; <sup>2</sup>BioEnergy Science Center (BESC), Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>University of Tennessee, Knoxville; <sup>4</sup>National Renewable Energy Laboratory, Golden, Colorado; <sup>5</sup>Samuel Roberts Noble Foundation, Ardmore, Oklahoma

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Lignocellulosic biomass such as switchgrass is a feedstock for biofuel production. However, the critical barrier towards conversion of such biomass to biofuels is its inherent recalcitrance to deconstruction. Switchgrass biomass is rich in cellulose, xylan, and lignin with smaller amounts of pectin, a particularly complex cell wall polysaccharide whose degree of covalent and non-covalent cross-linking with itself and other wall polymers influences wall architecture. The pectin polysaccharides are the most structurally complex of the plant cell wall glycans, consisting of the polysaccharides homogalacturonan (HG) and rhamnogalacturonan I and II. Here we show that reduced expression of three HG-biosynthetic glycosyltransferases, the  $\alpha$ -1,4-galacturonosyl-transferases GAUTI, GAUT4 and GAUT7, impacts both biomass yield and saccharification. All three GAUTs have been proposed to be involved in pectin biosynthesis. We manipulated the expression of GAUT1, GAUT4 and GAUT7 through an RNAi approach and determined the effects on biomass recalcitrance and growth. Glucose release per gram dry biomass was significantly increased by 24-35% in PvGAUT1-KD lines, 10-15% in PvGAUT4-KD lines and 12-22% in PvGAUT7-KD lines compared to controls. Total sugar release was increased by 17-31% in PvGAUTI-KD lines, 14-15% in PvGAUT4-KD lines and 9-14% in PvGAUT7-KD lines compared to controls. Silencing of these three genes leads to 7-43% increased plant height in PvGAUT1-KD lines, 12-17% in PvGAUT4-KD lines and

5-10% in *PvGAUT7-KD* lines, along with 110-230% more tillers compared to wild type. Both the increase in plant height and the greater number of tillers contributed to more aerial biomass in the transgenic lines than control plants. We hypothesize that specific GAUTs synthesize unique HG glycans that function as stand-alone polysaccharides and/or a glycans in wall glycoconjugates that are necessary for native cell wall integrity in grasses. We also hypothesize that reduced amounts of the pectic polymer(s) lead to loosened walls and hence decreased recalcitrance and increased growth in switchgrass.

# Modifying Carbon, Nitrogen, and Electron Metabolism in *Clostridium thermocellum* to Enhance Cellulosic Biofuel Yield and Titer

Adam M Guss<sup>1,2\*</sup> (gussam@ornl.gov), Thomas Rydzak,<sup>1</sup> Beth Papanek,<sup>1,2</sup> Shuen Hon,<sup>1,3</sup> Liang Tian,<sup>1,3</sup> Daniel Amador-Noguez,<sup>1,4</sup> Steven D Brown,<sup>1</sup> Daniel G Olson,<sup>1,3</sup> Lee R Lynd,<sup>1,3</sup> and **Paul Gilna**<sup>1</sup>

<sup>1</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>University of Tennessee, Knoxville; <sup>3</sup>Dartmouth College, Hanover, New Hampshire; <sup>4</sup>University of Wisconsin, Madison

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

A sustainable future will require the development of renewable alternatives to petroleum-derived fuels and chemicals, and one potential solution for the replacement of gasoline involves the conversion of plant biomass into liquid fuels. Due to its native ability to rapidly consume cellulose and its existing ethanol production pathway, *Clostridium thermocellum* is a leading candidate organism for implementing a consolidated bioprocessing strategy for biofuel production, wherein biomass deconstruction and fermentation occur in a single vessel without added enzymes. In addition to producing ethanol, *C. thermocellum* converts cellulose and soluble cellodextrins such as cellobiose to lactate, formate, acetate, H<sub>2</sub>, amino acids, and other products. Therefore, metabolic engineering is required to optimize flux to a single product.

A mutant strain of *C. thermocellum* was constructed to remove major side product formation, resulting in *C. thermocellum*  $\Delta hydG \Delta ldh \Delta pfl \Delta pta-ack$ . This strain no longer produces formate, acetate and lactate; hydrogen production is decreased four-fold; and the ethanol yield is doubled compared with the wild type on cellobiose, crystalline cellulose Avicel, and pretreated biomass. Laboratory strain evolution has allowed for further improvement of growth rate, yield, and titer, resulting in a strain capable of converting crystalline cellulose to ethanol at a titer of 22.5 g/L at

75% of the maximum theoretical yield. Genome resequencing revealed mutations that occurred during strain evolution and suggests mechanisms responsible for the improved phenotypes.

We also introduced four genes from a high ethanol yielding strain of engineered *T. saccharolyticum* that are likely important for ethanol production – adhA, nfnAB, and a mutant adhE – into wild-type *C. thermocellum*. We observed significant improvements to ethanol yield and titer in the resulting strain. Further engineering of this *C. thermocellum* strain by eliminating hydrogen production did not improve ethanol yield, but instead decreased ethanol titer on higher initial substrate concentrations. We hypothesize that we will need to further engineer *C. thermocellum* metabolism to more closely resemble that of *T. saccharolyticum* via an approach that includes both gene deletions and heterologous expression.

While these mutants exhibits higher ethanol yield, amino acids are still produced as end products. Therefore, genetic analysis of nitrogen metabolism is being investigated to understand the mechanism of ammonium assimilation and to devise strategies to prevent production of amino acids as fermentation products. Progress in understanding and altering nitrogen metabolism in *C. thermocellum* and the impact on product formation also will be discussed.

## New Insights on the Lignin Pathway in Grasses

Juan Carlos Serrani-Yarce,<sup>1,2\*</sup> (Juan.Serrani@unt.edu), Jaime Barros-Rios,<sup>1</sup> Luis L Escamilla-Trevino,<sup>1,2</sup> Nancy Engle,<sup>2</sup> Timothy J Tschaplinski,<sup>2</sup> Richard A Dixon<sup>1,2</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>University of North Texas, Denton, Texas; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

# http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

The lignin biosynthetic pathway was thought for many years to be a highly conserved metabolic process across all plant species. This assumption was based on studies on dicotyledonous plants, mainly the model *Arabidopsis thaliana*. Recent studies have shown that monocots and dicots differ in their patterns of lignification, and that these differences can be found even among species within the monocot or dicot clades.

We previously studied the lignin pathway in the leading biomass crop switchgrass (*Panicum virgatum*), and found that the effects of downregulation of some of the genes involved in the earlier steps of the pathway were not consistent with the current understanding of the monolignol pathway (Shen et al., 2013). Importantly, switchgrass plants downregulated in cinnamate 4-hydroxylase (C4H), coumaroyl shikimate 3'-hydroxyalse (C3'H) and caffeoyl CoA 3-0-methyltransferase (CCoAOMT), showed lignin phenotypes inconsistent with the previous literature on dicot plants (Reddy et al., 2005, Schilmiller et al., 2009, Wagner et al., 2011). Also, the second HCT step in the shikimate shunt described *inArabidopsis* and tobacco (Hoffmann et al., 2003, 2004) is unlikely to occur in switchgrass (Escamilla-Trevino et al., 2014). Finally, a recent study (Ha et al., 2016) shows that caffeoyl shikimate esterase (CSE) is critical for normal lignification in *Medicago truncatula* (dicot), poplar (*Populus deltoides*, dicot) and switchgrass (*Panicum virgatum*, monocot), but *Brachypodium distachyon* and corn (*Zea mays*) do not possess orthologs of CSE. Moreover, preliminary results on enzymatic activities and kinetics of recombinant enzymes in *Brachypodium* and switchgrass suggest that the route caffeate  $\rightarrow$ ferulate  $\rightarrow$ feruloyl CoA could function during monolignol biosynthesis.

To study more precisely the early steps of the lignin pathway in grasses, we used the model *B. distachyon,* a diploid monocot with a small genome size and ease of transformation. *Brachypodium* has eight phenylalanine ammonia-lyase (*PAL*) genes, one of which encodes a bifunctional PTAL (phenylalanine/tyrosine ammonia-lyase) (Barros et al., 2016), three *C4H* genes, two hydroxycinnamoyl CoA: shikimate/quinate hydroxycinnamoyl transferases (*HCT*) genes, and one *C3'H* gene. We generated knockdown lines showing strong downregulation for these gene families (*PAL* 43%, *TAL* 80%, *C4H70%, HCT92%,* and *C3'H90%*) and reduced lignin content (*PAL* 19%, *TAL* 25%, *C4H* 12%, *HCT* 9% and *C3'H* 17%). Metabolite profiling analysis of these lines showed significant changes in the levels of some lignin pathway intermediates, particularly a large accumulation of 3-0-coumaroylshikimic acid in the C3'H line, suggesting that the ester route to lignin is functional in *Brachypodium*. However, an alternative route to caffeate must exist, as *Brachypodium* possess only weak extractable esterase activity with caffeoyl shikimate and has no CSE orthologs (Ha et al., 2015). This may involve the activity of a putative 4-coumarate 3-hydroxylase.

### References

- Barros, J., Serrani-Yarce, J.C., Chen, F., Baxter, D., Venables, B.J. and Dixon, R.A. (2016) Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nat Plants*, 2, 16050.
- Escamilla-Trevino, L.L., Shen, H., Hernandez, T., Yin, Y., Xu, Y. and Dixon, R.A. (2014) Early lignin pathway enzymes and routes to chlorogenic acid in switchgrass (*Panicum virgatum L.*). *Plant Molecular Biology*, 84, 565-576.
- Ha, C.M., Escamilla-Trevino, L., Yarce, J.C., Kim, H., Ralph, J., Chen, F. and Dixon, R.A. (2016) An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. *Plant Journal*, 86, 363-375.
- Hoffmann, L., Besseau, S., Geoffroy, P., Ritzenthaler, C., Meyer, D., Lapierre, C., Pollet, B. and Legrand,
- M. (2004) Silencing ofhydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell*, 16, 1446-1465.
- Hoffmann, L., Maury, S., Martz, F., Geoffroy, P. and Legrand, M. (2003) Purification, cloning, and properties ofan acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *Journal ofBiological Chemistry*, 278, 95-103.
- Reddy, M.S., Chen, F., Shadle, G., Jackson, L., Aljoe, H. and Dixon, R.A. (2005) Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (Medicago sativa L.). *Proceedings of the National Academy of Sciences USA* 102, 16573-16578.
- Schilmiller, A.L., Stout, J., Weng, J.K., Humphreys, J., Ruegger, M.O. and Chapple, C. (2009) Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in Arabidopsis. *Plant Journal*, 60, 771-782.
- Shen, H., Mazarei, M., Hisano, H., Escamilla-Trevino, L., Fu, C., Pu, Y., Rudis, M.R., Tang, Y., Xiao, X.,
- Jackson, L., Li, G., Hernandez, T., Chen, F., Ragauskas, A.J., Stewart, C.N., Jr., Wang, Z.Y. and Dixon, R.A. (2013) A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant Cell*, 25, 4342-4361.
- Wagner, A., Tobimatsu, Y., Phillips, L., Flint, H., Torr, K., Donaldson, L., Pears, L. and Ralph, J. (2011) CCoAOMT suppression modifies lignin composition in Pinus radiata. *Plant Journal*, 67, 119-129.

# Pleiotropy Decomposition of 609 Populus Trichocarpa Genotypes

Deborah Weighill<sup>1,2</sup>, Carissa Bleker<sup>1,2</sup>, Priya Ranjan<sup>1</sup>, Nan Zhao<sup>1</sup>, Madhavi Martin<sup>1</sup>, Gerald Tuskan<sup>1</sup>, Wellington Muchero<sup>1</sup>, Tim Tschaplinski<sup>1</sup>, Daniel Jacobson<sup>1,2\*</sup> (jacobsonda@ornl.gov), and **Paul Gilna**<sup>1</sup>

<sup>1</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Pleiotropy is the phenomenon in which a gene affects multiple phenotypes [1]. One can also have a SNP-centric view of pleiotropy as a single SNP affecting multiple phenotypes [2]. While pleiotropy used to be considered an exception to the rules of Mendelian genetics, it has since been proposed to be a common, central property inherent to biological systems [1]. Pleiotropic patterns can be detected in the results of Genome Wide Association Studies (GWASs) as SNPs within genes having multiple significant phenotypic associations. Two main pleiotropic patterns exist within GWAS results. Firstly, Type 1 pleiotropy occurs when a single SNP within a gene is associated with more than one phenotype [2]. Type 2 pleiotropy occurs when two different SNPs within a single gene have different phenotype associations [2].

In this study, we present a method called Pleiotropy Decomposition for the investigation of pleiotropic patterns from GWAS analysis of *Populus trichocarpa*. The method aims to distinguish between different pleiotropic patterns (Type 1 vs Type 2) while also providing intuitive network representations for the exploration of these pleiotropic patterns in the GWAS results.

GWAS analysis using EMMAX [3] was performed on a carefully constructed, non-related population of ~1,000 *Populus trichocarpa* genotypes and phenotype information in the form of untargeted metabolomic profiles for 609 of these genotypes, followed by False Discovery Rate correction. From the resulting set of SNP-phenotype associations, we constructed a gene-phenotype GWAS network (matrix) consisting of genes connected to their associated phenotypes. Our method involved the decomposition of the gene-phenotype network to two bipartite networks through the use
of pleiotropic modules as an intermediate, latent variable. These networks together unravel the underlying pleiotropic structure of genes.

The main intermediate step in pleiotropy decomposition involved the construction of pleiotropic modules, namely, groups of SNPs that are associated with the same sets of phenotypes. This was performed by constructing a GWAS profile for each SNP as a vector of its phenotypic associations, calculating the Proportional Similarity index between the GWAS profiles of all pairs of SNPs, and clustering using MCL [4]. Pleiotropy Decomposition thus involves the separation of the gene-phenotype matrix into a gene-module matrix and a module-phenotype matrix, which can be visualized as a gene-module bipartite network and a module-phenotype bipartite network, respectively. The gene-module matrix gives information regarding the type of pleiotropy exhibited by genes, (pleiotropic structure of genes), and the module-phenotype gives information regarding the specific phenotype associations of the pleiotropic modules that genes are comprised of.

The use of these decomposition matrices as bipartite networks provides a convenient structure for the exploration of GWAS results, while also unraveling the different pleiotropic signatures of genes. These networks allowed for interesting patterns of enrichment to be found within the set of pleiotropic genes, and highlighted the pleiotropic interactions which link functions such as plant defense responses and lignin biosynthesis. Identification of potential pleiotropic genes, as well as the functions those pleiotropic genes might affect should prove to be a useful tool in planning targeted experiments involving genetic modification, as it could give an idea as to the potential other side effects the modification/deletion of a gene of interest might cause.

## References

- 1. Tyler, Anna L., et al. "Shadows of complexity: what biological networks reveal about epistasis and pleiotropy." *Bioessays* 31.2 (2009): 220-227.
- Solovieff, Nadia, et al. "Pleiotropy in complex traits: challenges and strategies." *Nature Reviews Genetics* 14.7 (2013): 483-495.
- 3. Kang, Hyun Min, et al. "Variance component model to account for sample structure in genome-wide association studies." *Nature genetics* 42.4 (2010): 348-354.
- 4. Van Dongen, Stijn. "Graph clustering via a discrete uncoupling process." *SIAM Journal on Matrix Analysis and Applications* 30.1 (2008): 121-141.

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

# Pyruvate Decarboxylase: Rationally Evolving Thermostable Enzymes for Metabolic Engineering

Deanne W. Sammond<sup>1, 2\*</sup> (<u>Deanne.Sammond@nrel.gov</u>), Daehwan Chung,<sup>1,2</sup> Bryon Donohoe,<sup>1,2</sup> Markus Alahuhta,<sup>1,2</sup> Nicholas S. Sarai,<sup>1</sup> Michael E Himmel,<sup>1</sup> Vladimir Lunin,<sup>1,2</sup> Adam M Guss,<sup>2</sup> Yannick J Bomble,<sup>1,2</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>National Renewable Energy Laboratory, Golden, Colorado; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

#### http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Strategies for the biological production of chemicals have focused on metabolically engineering organisms to produce target chemicals near theoretical yields. Streamlining metabolic pathways by eliminating branched points is a common approach to increase chemical yields, generally achieved by removing native metabolic enzymes. Alternatively, introducing a non-native enzyme can provide a metabolic "short cut," bypassing branched points. Here we aim to introduce the mesophilic enzyme, pyruvate decarboxylase (PDC) into a thermophilic organism to introduce just such a metabolic short cut in order to increase the production of ethanol. We have developed an *in silico-in vitro-in vivo* platform to rationally evolve the mesophilic PDC for enhanced thermostability. Importantly we find that cooperative effects for the thermostabilizing mutations are critical in our design of an enzyme with a significantly enhanced lifetime at elevated temperatures.

PDC converts pyruvate to acetaldehyde through a non-oxidative decarboxylation mechanism, a reaction that is part of the fermentative process under anaerobic conditions. PDC is found in mesophilic organisms, yet to date has not been identified in thermophiles.<sup>1</sup> The metabolic routes in fermentative thermophilic organisms from pyruvate to ethanol are branched and produce a number of additional products including acetic acid, formic acid and lactic acid. Expression of PDC in a fermentative thermophile would bypass metabolic branched points, channeling more pyruvate directly to acetaldehyde for conversion to ethanol. Additionally, PDC uses a non-oxidative

mechanism and therefore will not consume the limited supply of redox cofactors such as NADH, NADPH, or ferredoxin.

PDC is a 260 kDa homotetrameric enzyme with two active sites buried at each dimer interface. A thermostable PDC must include stabilization of the monomer to resist denaturation at high temperatures and stabilization of the interactions between the individual units to maintain the active complex. PDC from the mesophile *Zymonomas mobilis* has an optimum growth temperature of 30°C, while the optimum growth temperature of the host, *Caldicellulosiruptor hydrothermalis*, is 65°C.<sup>2</sup> Our aim is to identify stabilizing mutations located in the different regions of PDC, defining these regions as specific sub-domains including the dimer and tetramer interfaces, the surface and the core of the protein. Multiple protein design approaches are required to identify mutations located in each of these regions. We utilize the protein design software, Rosetta, for our *in silico* mutagenesis.<sup>3,4</sup> Rosetta offers the benefit of having many distinct algorithms designed to identify thermostabilizing mutations, allowing us to target the various regions of the PDC.

## References

- 1. M. S. Eram, K. Ma, Decarboxylation of pyruvateto acetaldehyde for ethanol production by hyperthennophiles. *Biomolecules* **3**, 578-596 (2013).
- 2. M. L. Miroshnichenko et al., Caldicellulosiruptor kronotskyensis sp. nov. and Caldicellulosiruptor hydrothennalis sp. nov., two extremely thermophilic, cellulolytic, anaerobic bacteria from Kamchatka thermal springs. *Int. J. Syst. Evol. Microbiol.* **58**, 1492-1496 (2008).
- 3. B. Kuhlman, D. Baker, Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* 97, 10383-10388 (2000).
- 4. K. W. Kaufmann, G. H. Lemmon, S. L. Deluca, J. H. Sheehan, J. Meiler, Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987-2998 (2010).

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

# Rational Methyltransferase Expression in *Escherichia coli* for Transformation of New Organisms

Lauren Riley<sup>1,2,3\*</sup> (<u>rileyla@ornl.gov</u>), Beth Papanek,<sup>1</sup> Kaela O'Dell,<sup>1,2</sup> Irenee Payne,<sup>1</sup> Carly Duffy,<sup>1</sup> Bob Schmitz,<sup>4</sup> Jan Westpheling,<sup>2,4</sup> Adam Guss,<sup>1,2,3</sup> and **Paul Gilna**<sup>1,2</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>BioEnergy Science Center, Oak Ridge, TN; <sup>3</sup>University of Tennessee, Knoxville; <sup>4</sup>University of Georgia, Athens

# http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

Many organisms naturally possess complex physiological traits that are of interest for biotechnology research. The ability to easily harness these traits in their native host could usher in a new era of biotechnology where synthetic biology is routinely applied to these non-standard organisms. However, many of these organisms are unable to be bioengineered due to a lack of available genetic tools. The development of genetic tools is limited in part by the inability to efficiently transform DNA into these organisms. One of the major barriers to successful transformation of bacteria is native DNA restriction-modification systems. DNA restrictionmodification systems act as a bacterial immune system to cut DNA that is methylated differently than the host and are typically comprised of methylation and restriction subunits. To prevent host death, restriction enzymes and the cognate DNA methylases recognize the same target sequence. Therefore, in order to overcome restriction, DNA needs to be methylated in the same manner as the host organism prior to transformation. In order to determine the sites targeted for restriction in these strains, methylome analysis was initially performed for 17 organisms in collaboration with the Department of Energy Joint Genome Institute. This information was used to choose methyltransferases for expression in E. coli using a new system for multiple gene integration in *E. coli*. The gene integration system utilizes serine bacteriophage integrases, which enable a single, unidirectional recombination event between two specific DNA sequences, attB and attP, for stable insertion of DNA into the E. coli chromosome. This process of mimicking host methylation patterns successfully allowed for transformation of the type strain of Clostridium thermocellum, strain ATCC 27405. The native methyltransferase gene Cthe0519

and a bi-functional Phi3TI methyltransferase were expressed from the *E. coli* chromosome. Plasmid DNA isolated from this methylating *E. coli* strain was then transformed into *C. thermocellum*. A similar approach is being used to demonstrate genetic transformation of other phylogenetically and metabolically diverse organisms. This system will allow for rapid transformation of new organisms to help further bioengineering research.

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

# Signaling Between Switchgrass and Fungal Endosymbionts in the Genus Serendipita

Prasun Ray<sup>1,2\*</sup> (<u>pray@noble.org</u>), Yingqing Guo,<sup>1</sup> Liang Sun,<sup>1</sup> Yuhong Tang,<sup>1</sup> Kelly D. Craven,<sup>1</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>The Samuel Roberts Noble Foundation, Ardmore, Oklahoma; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

*Serendipitales* (previously *Sebacinales*) are root-associated symbiotic fungi with growth promoting abilities. This association is termed mycorrhiza (Gk. mykós, "fungus," and riza, "roots"). *Serendipita vermifera* has been shown to be tremendously beneficial in enhancing biomass yield and drought tolerance in switchgrass, an important bioenergy crop. Considering their proven beneficial impact on plant growth and their apparent ubiquity, *Serendipitales* fungi should be considered as a previously hidden, but amenable and effective microbial tool for enhancing plant productivity and stress tolerance. Unfortunately, the agronomic utility of these fungi are currently hampered by the paucity of strains available, the large majority being isolated from Australian orchids. Of particular relevance to this study, our group has addressed this constraint by isolating the first North American strain of *Serendipita* named as *Serendipita vermifera* ssp. *bescii* from Oklahoma that is presumably better adapted to the specific agro-climatic conditions of the Southern Great Plains region of the central United States (Ray and Craven, U.S. Patent pending).

Development of mycorrhizal symbiosis involves the differentiation of both symbionts to create novel symbiotic interfaces within the root cells. Further, recent evidence indicates that the plant and mycorrhizal fungi perceive each other prior to their physical interaction. While this perception is real, underlying factors facilitating this signal perception are currently unknown. The aim of this project was to explore the signals used by the symbiotic partners for the development of the symbiosis. The overarching goal is to design crop-fungus combinations for optimal and sustainable performance by identifying these key signaling components. To accomplish this, we investigated the transcriptome profile of the symbiotic interface - i.e., switchgrass roots at different stages of fungal colonization by the use of RNA-Seq technology.

*Panicum virgatum* Cultivar Alamo (AP13) seedlings were colonized *in vitro* with two different strains of *Serendipitaceae* namely *Serendipita vermifera* strain MAFF305830 (Orchid, Australia, 1988) and *Serendipita vermifera* ssp. *bescii* (Switchgrass, USA, 2014). Plants were harvested at three different stages of colonization namely (a) pre-colonization, (b) early stage colonization (c) late stage colonization; RNA was extracted from roots by standard protocol. Stranded RNA-Seq library(s) were created and quantified by qPCR. Sequencing was performed using Illumina HiSeq 2500. Raw fastq file reads were filtered and trimmed using the JGI QC pipeline. Following trimming, reads under the length threshold were removed. Filtered reads from each library were aligned to the reference genome using HISAT version 0.1.4-beta. DESeq2 (version 1.8.1) was subsequently used to determine which genes were differentially expressed between pairs of conditions at adjusted p < 0.05.

We report that switchgrass seems to perceive fungal symbiont prior to physical contact. Following that, the plant defense system reorganizes to accommodate the fungal symbiont. Consequently, the nutrient transport machinery is adjusted for the fungal symbiont. Two fungal strains differentially induced plant gene expression, demonstrating unique responsiveness of plant genotype to these fungal symbionts. Understanding such responsiveness will facilitate to design crop-fungus combinations for optimal sustainable agricultural practices.

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

## Switchgrass Fermentation by Thermophilic Microbiomes

Xiaoyu Liang<sup>1,2\*</sup> (<u>Xiaoyu Liang TH@dartmouth.edu</u>) Xiongjun Shao,<sup>1,2</sup> Evert K Holwerda,<sup>1,2</sup> Dan Olson,<sup>1,2</sup> Sean Murphy,<sup>1,2</sup> Liang Tian,<sup>1,2</sup> Tom L Richard,<sup>3</sup> Jason M Whitham,<sup>1</sup> Dawn M Klingeman,<sup>1</sup> James G Elkins,<sup>1</sup> Steven D Brown,<sup>1</sup> Lee R Lynd,<sup>1,2</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>Dartmouth College, Hanover, New Hampshire; <sup>3</sup>The Pennsylvania State University, University Park

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

The study of lignocellulose-fermenting microbiomes can inform and enable development of industrial processes based on defined microbial cultures. Yet there have been few fundamental studies of lignocellulose fermentation under the conditions anticipated for industrial processes, including conversion of high solids (> 100 g/L). Motivated by this perspective, we have initiated study of efficient switchgrass fermentations by thermophilic anaerobic lignocellulolytic consortia of microbiomes.

Triplicate semi-continuous, anaerobic cultures were operated at 55°C on mid-season harvested switchgrass with no added organic nutrients for more than 19 months to obtain steady-states at solids concentrations from 30g/L to 150g/L and residence times (RT) from 20 to 3.3 days. Cultures were fed semi-continuously by replacing  $1/10^{\text{th}}$  of the fermentation broth at regular time intervals. Undiminished fractional carbohydrate solubilization was observed over a 5-fold range of feedstock loading. Stable methanogenesis with minimal accumulation of organic acids was observed under all conditions, which was something of a surprise at RT=3.3 days. Yet more surprising, initial experiments indicated that fractional carbohydrate solubilization by pure cultures of *Clostridium thermocellum* was at least as high as that obtained with lignocellulose-fermenting microbiomes at low solids loading. However, solubilization of higher substrate loadings was dependent on the cooperative action of mixed community microbiomes.

16S rDNA and metagenomics analysis was conducted to characterize the microbiome. At low (30 g/L) solids loading, Firmicutes dominated each of the cultures, representing between 54--96% of the sequence reads, and these populations increased with reduced residence time. Species present in the consortia were also matched to closely related cultured species. Sixty percent of the microbiome was shown to be potentially cellulolytic, with the most abundant matching to *Clostridium clariflavum*. Plate culture isolation further revealed that *C. clariflavum* was one of the major cellulolytic species existing in the mixed culture consortia.

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

# The Effect of Lignin and Hemicellulose Removal on Switchgrass Deconstruction by *Clostridium thermocellum*

Ninad Kothari<sup>1,2\*</sup> (<u>nkoth001@ucr.edu</u>), Charles Cai,<sup>1,2</sup> Rajeev Kumar,<sup>1,2</sup> Charles E. Wyman,<sup>1,2</sup> and **Paul Gilna**<sup>2</sup>

<sup>1</sup>University of California, Riverside; <sup>2</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

# http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

When produced from lignocellulosic biomass, ethanol is a sustainable transportation fuel with very low net emissions of greenhouse gases and its use will reduce our heavy dependence on petroleum. Ethanol is a high-octane fuel and can be further upgraded to drop-in fuels such as butanol, jet fuel, gasoline, and diesel. Ethanol made from corn starch and cane sugar currently dominates the alternative fuels market. However, production from lignocellulosic biomass has the promise for larger-scale impact at a low cost, and commercial cellulosic ethanol projects are now starting operations. These conventional lignocellulosic biomass conversions to ethanol employ major operations for size reduction, pretreatment, enzyme production, enzymatic hydrolysis, fermentation, and product recovery. Unfortunately, the high doses of fungal cellulolytic enzymes required to achieve high sugar yields necessary for commercial success are too costly to support large expansion of this nascent industry.

Consolidated bioprocessing (CBP) eliminates this expensive separate enzyme production step and simplifies the process by combining enzyme production, enzymatic hydrolysis, and fermentation operations.

*Clostridium thermocellum*, a leading CBP microorganism, produces a complex cellulosome that hydrolyzes the polysaccharides in biomass and then ferments their breakdown products into ethanol and other metabolites. Although the *C. thermocellum* cellulosome can realize about 48% glucan conversion from milled switchgrass without pretreatment, these yields are still too low to be

economically attractive. Therefore, pretreatment may be necessary to enhance biological deconstruction by *C. thermcoellum*. A variety of pretreatment technologies can prepare lignocellulosic biomass for high yields from CBP by achieving distinctive changes in the solids' compositional and structural characteristics. However, because the influence of variation in the composition of pretreated biomass solids on deconstruction by *C. thermcoellum* is not understood– hydrothermal, dilute acid, dilute alkali, and co-solvent enhanced lignocellulosic fractionation (CELF) pretreatments were applied at varying conditions to milled switchgrass. Dilute acid and hydrothermal are considered to be leading pretreatments that reduce biomass recalcitrance by removing most of the hemicellulose and some of the Klason lignin (K-lignin). Dilute alkali pretreatments, on the other hand, remove a large part of K-lignin but only some of the hemicellulose. By comparison, the CELF invented at the University of California, Riverside employs tetrahydrofuran (THF) as a miscible co-solvent in combination with an aqueous dilute acid solution to realize high solubilization and recovery of hemicellulose sugars as well as greatly enhanced K-lignin removal from lignocellulosic biomass compared to dilute acid alone.

Each of these pretreatment technologies were optimized to maximize total sugar yields from pretreatment (Stage 1) combined with wild-type *C. thermocellum* biological deconstruction (Stage 2). Biomass deconstruction by *C. thermocellum* was further compared to that realized by application of various loadings of conventional fungal enzymes to the same pretreated solids.

The translation of sugar deconstruction by *C. thermocellum* into metabolite production was also analyzed. The pretreatments were performed over a range of temperatures and times, and the resulting solids were washed thoroughly prior to CBP. *C. thermocellum* fermentations were performed at a 5 g/L glucan loading in a 50 mL working volume incubated at 60°C with a 180 rpm shaking speed, and fungal enzymes mediated enzymatic hydrolysis was performed at 50°C and 150 rpm. The results showed that enzymatic hydrolysis at high and expensive enzyme loadings of  $\geq$ 65 mg protein/g glucan was required to realize the high levels of polysaccharide deconstruction achieved by *C. thermocellum* without added enzymes. This work also showed that physical removal of lignin from switchgrass had a greater impact on biological deconstruction by *C. thermocellum* than lignin relocation and/or hemicellulose removal. Moreover, removal of lignin and hemicellulose simultaneously by CELF pretreatments resulted in 100% sugar release from switchgrass when combined with CBP. At the same time, metabolites production by *C. thermocellum* followed the trend observed for deconstruction of the polysaccharides in the different pretreated solids.

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

# The Unique Mechanism of the Dominant Multi-Component Cellulase from *Caldicellulosiruptor bescii*

Roman Brunecky,<sup>1</sup> Bryon S. Donohoe,<sup>1</sup> John M Yarbrough,<sup>1</sup> Ashutosh Mittal,<sup>1</sup> Brian R Scott,<sup>2</sup> Hanshu Ding,<sup>2</sup> Larry E Taylor, II,<sup>1</sup> Jordan F Russell,<sup>3</sup> Daehwan Chung,<sup>1,4</sup> Jan Westpheling,<sup>3,4</sup> Sarah A Teter,<sup>2</sup> Michael E Himmel,<sup>1,4</sup> Yannick J Bomble<sup>1,4\*</sup> (<u>yannick.bomble@nrel.gov</u>,) and **Paul Gilna**<sup>4</sup>

<sup>1</sup>National Renewable Energy Laboratory, Golden Colorado; <sup>2</sup>Novozymes, Inc., Davis, California; <sup>3</sup>University of Georgia, Athens; <sup>4</sup>BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee

## http://bioenergycenter.org

Project Goals: The BioEnergy Science Center (BESC) focuses on fundamental understanding and elimination of biomass recalcitrance. BESC's approach to improve accessibility to the sugars within biomass involves (1) improved plant cell walls for rapid deconstruction and (2) multi-talented microbes for converting plant biomass into biofuels in a single step [consolidated bioprocessing (CBP)]. Biomass research works with two potential bioenergy crops (switchgrass and *Populus*) to develop improved varieties and to understand cell wall biosynthesis pathways. We test large numbers of natural variants and generate specific modified plants samples. BESC's research in deconstruction and conversion targets CBP manipulating thermophilic anaerobes and their cellulolytic enzymes for improved conversion, yields, and titer. Enabling technologies in biomass characterization, 'omics, and modeling are used to understand chemical and structural changes within biomass and to provide insights into mechanisms.

The cellulase, CelA, from the thermophile, *Caldicellulosiruptor bescii*, is one of the most active cellulose degrading enzymes known to date. In the saccharification of Avicel, a common cellulose standard, CelA outperforms mixtures of benchmark fungal exo- and endoglucanases. Unlike the secretomes of cellulolytic fungi, which typically comprise multiple, single catalytic domain enzymes for biomass degradation, some bacterial systems employ an alternative strategy that utilizes multicatalytic domain cellulases. Additionally, CelA is extremely thermostable and highly active at elevated temperatures, unlike commercial fungal cellulases. However, the activity of CelA seems to be diminished when acting on biomass, yet the barriers responsible for this loss of activity are not yet clear. Many of the factors negatively affecting digestion of lignocellulosic materials by *C. bescii* enzyme cocktails containing CelA appear to be significantly different from the performance barriers affecting fungal cellulases. Here, we explored the activity and degradation mechanism of CelA on a variety of pretreated substrates to better understand how the different bulk components of biomass, such as xylan and lignin, impact its performance. Notably, we have determined that lignin content, but not cellulose crystallinity, is an impediment to the cellulolytic activity of CelA.

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

# High throughput phenotyping and quantitative genetics to understand productivity and drought traits in the model C4 grass Setaria

Ivan Baxter<sup>1,3</sup>\*, Maximilian Feldman<sup>1</sup>, Patrick Ellsworth<sup>4</sup>, Asaph Cousins<sup>4</sup>, Andrew D.B. Leakey<sup>2</sup>, Darshi Banan<sup>2</sup>, Rachel Paul<sup>2</sup>, and **Thomas P. Brutnell<sup>1</sup>** 

<sup>1</sup>Donald Danforth Plant Science Center, St Louis; <sup>2</sup>University of Illinois, Urbana-Champaign; , <sup>3</sup>USDA-ARS, St Louis, and <sup>4</sup>Washington State University.

# http://foxmillet.org/

Genetically tractable model systems closely related to bioenergy grasses need to be developed to drive the crop improvement required for large scale, ecologically sustainable bioenergy production. *Setaria viridis* is an ideal candidate C<sub>4</sub> panacoid grass. The overarching objectives of this large, collaborative project are to utilize genomic, computational and engineering tools to begin the genetic dissection of drought response in *S. viridis*. This will be achieved through: 1) Quantitative trait and association genetics; 2) novel controlled environment and field phenotyping combined with molecular and chemical profiling; 3) development of metabolic and gene networks; 4) development of transformation technologies; 5) reverse genetic testing of candidate genes.

Phenotyping has become the rate-limiting step in using large-scale genomic data to understand and improve agricultural crops. Our project has used the Bellwether Phenotyping platform for controlled-environment plant growth and automated, multimodal phenotyping to study how plant biomass traits change temporally in response to water availability and identify the genetic loci underlying those changes. We have analyzed two independent genetically structured populations of *Setaria* sp.: an interspecific *S. italica* x *S. viridis* recombinant inbred line population and to two grow outs of a *S. viridis* natural diversity panel. We developed Plant Computer Vision (PlantCV) as an open-source, platform independent quantitative image analysis community resource and have used it to quantify height, biomass, water-use efficiency, color, plant architecture, and near-infrared traits. We have identified QTLs for height, biomass and water use efficiency in both populations. We have also conducted elemental profiling and isotopic studies of the same plants to link biochemical and phenomic approaches to understanding the response of plants to water deficit.

*This research was funded through under Prime Agreement No. DE-SC0008769 from Department of Energy to the Donald Danforth Plant Science Center* 

# Utilizing *Setaria viridis* as a Model for Molecular Characterization of Jasmonate-Mediated Growth and Defense Responses

Christine Shyu\*, Kimberly Maxson-Stein, Indrajit Kumar, Dustin Mayfield-Jones and **Thomas P. Brutnell** Donald Danforth Plant Science Center, St. Louis, MO 63132

\*Presenter. Contact e-mail: CShyu@danforthcenter.org

Project Goals: This project aims to utilize molecular, genetic, and genomic resources in *Setaria viridis* to dissect the jasmonate signaling network in panicoid grasses. Outcomes from this research provide opportunities to engineer bioenergy feedstocks for improved stress resistance without compromising growth.

The plant hormone jasmonate (JA) and its derivatives control many important agricultural traits from growth and development to defense against biotic and abiotic stresses. Though JA signaling components have been studied in the dicot model Arabidopsis, mechanisms underlying JAmediated growth and defense in grasses have not been elucidated. Setaria viridis is a model panicoid grass with a rapid life cycle, short plant stature and rapidly developing genetic and molecular toolsets. It is closely related to bioenergy feedstocks such as sorghum, switchgrass, and Miscanthus, making it an ideal model for gene discovery and molecular studies. The core JA signaling pathway consists of JASMONATE ZIM-DOMAIN (JAZ) repressors that interact with COI in the presence of bioactive JA. Upon COI-JAZ interaction, JAZs are ubiquitinated and targeted for degradation, resulting in activation of downstream transcription factors (TFs) that regulate JA-responsive gene expression. In this study, Setaria viridis COI-JAZ interactions were examined and unique interaction partners were identified. Cas9-mediated genome editing of SvCOIs with different JAZ-interaction patterns were generated and preliminary phenotypes were observed. To identify downstream TF interaction partners of JAZ proteins, a wound treatment was applied to the S. viridis leaf and transcriptional responses measured over a developmental gradient. One hundred and eight RNA-seq libraries were generated from control and wounded leaf segments across time. Clustering analysis of the RNA-seq data coupled with co-expression analyses led to identification of novel basic helix-loop-helix TFs that are candidates for regulating JA signaling outputs. Outcomes from this research provide insight into the dynamics and complex regulation of JA responses in grass systems, and provide opportunities to engineer bioenergy crops for enhanced stress resistance without compromising growth.

# Quantitative Trait Loci for Leaf Carbon Isotopic Signature and Transpiration Efficiency in the C<sub>4</sub> grass *Setaria*

Asaph B. Cousins<sup>1\*</sup> (acousins@wsu.edu), Patrick Z. Ellsworth<sup>1</sup>, Max Feldman<sup>2</sup>, Patricia V. Ellsworth<sup>1</sup>, Ivan Baxter<sup>2</sup>, and **Tom Brutnell**<sup>2</sup>

<sup>1</sup>Washington State University, Pullman, WA 99163, <sup>2</sup>Donald Danforth Plant Science Center, St. Louis, MO 63132

# http://foxmillet.org

**Project Goals:** The objectives of this project are to 1) identify the gene regulatory and metabolic networks important for adaptation to low water availability and high-density plantings and to gain a greater understanding of the physiology underlying these adaptations and 2) develop technologies to precisely control gene insertion and replacements events for large scale engineering of pathways in model and target feedstocks.

**Abstract:** Leaf carbon isotopic composition ( $\delta^{13}C_{\text{leaf}}$ ) in C<sub>4</sub> plants is a proposed proxy for transpiration efficiency (TE) suitable for high throughput phenotyping. To test this, we studied the C<sub>4</sub> plants S. viridis (A10) and S. italica (B100) to determine the relationship between TE and  $\delta^{13}$ C<sub>leaf</sub>. We found that  $\delta^{13}$ C<sub>leaf</sub> differentiated between well-watered and water-limited plants and correlated with plant- and leaf-level measures of TE (Ellsworth et al., in review). Following this experiment, we used a population of 217 recombinant inbred lines (RIL) derived from A-10 and B-100 to identify quantitative trait loci (QTL) relating to  $\delta^{13}$ C<sub>leaf</sub> and TE under both well-watered and water-limited conditions. The experiment was conducted in the Bellweather Phenotyping Facility at the Danforth Center where plant size and water use was measured daily. Leaf  $\delta^{13}$ C was significantly more negative in the water-limited treatments (P < 0.0001). Additionally, TE and  $\delta^{13}$ Cleaf had co-localized QTL on chromosome 7 and 9. Additionally, the additive effects associated with these QTL were in opposite directions corroborating the negative relationship between TE and  $\delta^{13}$ Cleaf commonly reported in C<sub>4</sub> plants. Plant size and transpiration also had the same co-localized QTL as TE and  $\delta^{13}$ Cleaf. Two unique QTLs were found for TE that were not present for either plant size or transpiration on chromosome 5 and 9, suggesting these two QTLs are independent of either biomass production or transpiration. Having identified the same QTLs for  $\delta^{13}$ C<sub>leaf</sub> and TE, we are

moving closer to finding a genomic control of TE and to using marker-assisted approaches to select for TE in C<sub>4</sub> plant breeding programs.

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

Functional Genomics of Lipid Accumulation in *Rhodosporidium toruloides* 

Dominic Pinel<sup>1#</sup>, Samuel T. Coradetti<sup>2#</sup>, Gina Geiselman<sup>1</sup>, **Rachel B. Brem<sup>2,3</sup>, Adam P. Arkin**<sup>\*1,4,5</sup> (aparkin@lbl.gov), and Jeffrey M. Skerker<sup>1,4</sup>

<sup>1</sup>Energy Biosciences Institute, UC Berkeley, Berkeley, CA; <sup>2</sup>Buck Institute for Research on Aging, Novato, CA; <sup>3</sup>Department of Plant and Microbial Biology, UC Berkeley, Berkeley, CA; <sup>4</sup>Department of Bioengineering, UC Berkeley, Berkeley, CA; <sup>5</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA <sup>#</sup>These authors contributed equally to this project.

# Project goals: to map genotype to phenotype on a genomic scale in the oleaginous yeast *Rhodosporidium toruloides*.

Wild isolates of the basidiomycete yeast *Rhodosporidium toruloides* can accumulate more than 70% of their biomass as neutral lipid in the form of triacylglycerol, and produce high concentrations of carotenoids. R. toruloides is also remarkable for its ability to utilize a wide spectrum of plant cell wall sugars and lignin derivatives and its natural tolerance of components of biomass hydrolysates that often inhibit growth of other species. These properties make R. toruloides an attractive host for sustainable production of hydrocarbon chemicals and fuels from lignocellulosic feedstocks. Although the *R. toruloides* research community is growing rapidly, relatively sparse biochemical data limits current engineering efforts. To enable the rapid interrogation of gene function on a genomic scale, we developed a 300,000+ member strain library of barcoded R. toruloides mutants via high efficiency Agrobacterium tumefaciens mediated transformation. With this mutant set we established methods for massively parallel assignment of phenotypes to genes in a given condition, by tracking the relative abundance of barcoded mutant strains in growth competition experiments and physical enrichments. We used this technique to identify a large cohort of genes (many with no homology to well-characterized genes in other systems) with roles in central carbon metabolism, nucleation and maturation of lipid bodies, protein trafficking, and nutrient sensing. Single-gene targeted deletion experiments validated the roles of these genes in the phenotypes inferred from our genomic approach. Together, these data provide a comprehensive understanding of the genetics of lipid metabolism in *R. toruloides*, and will serve as a foundation for the rational engineering of improved production strains of this fungus.

Funding statement: this project was funded by a grant from the Department of Energy Office of Biological and Environmental Research.

# Genomics-Assisted Breeding for Leaf Rust (*Melampsora*) Resistance in Shrub Willow (*Salix*) Bioenergy Crops

**Lawrence B. Smart**<sup>1\*</sup> (lbs33@cornell.edu), Chase R. Crowell<sup>2</sup>, Craig H. Carlson<sup>1</sup>, Fred E. Gouker<sup>1</sup>, Mariami Bekauri<sup>2</sup>, Ali Cala<sup>2</sup>, and Christine D. Smart<sup>2</sup>

<sup>1</sup>Horticulture Section, <sup>2</sup>Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Geneva, NY 14456

http://willow.cals.cornell.edu

**Project Goals:** We will leverage recent significant investments in genomics and genetics resources for *Salix* and *Melampsora* by the DOE and USDA and also develop new systems genetics tools to characterize willow leaf rust diversity and map genes for resistance in willow. We will use ITS-LSU sequencing and genotyping-by-sequencing to characterize 200 single uredinial isolates of *Melampsora* infecting *Salix purpurea* cultivars across the Northeast. The pathogenicity of those isolates will also be tested. We will accurately map QTL for resistance to *Melampsora* in *S. purpurea* in both a segregating F<sub>2</sub> linkage mapping population and association mapping populations. We will map eQTL controlling gene-level and coordinated gene network transcriptomic responses to *Melampsora* colonization by performing RNA-Seq on leaves of F<sub>2</sub> *S. purpurea* progeny inoculated with *Melampsora*. We will map and characterize major genes for resistance to *Melampsora* in different *Salix* species hybrids using a common parent mapping approach.

Abstract: Shrub willow (*Salix* spp.) is a proven, high-yielding perennial woody crop that can be grown on underutilized or marginal agricultural land, but which faces a major long-term threat of yield losses due to leaf rust, caused by Melampsora spp. Although Melampsora has been described taxonomically, species diversity within a single site and regionally is unknown. Approximately 200 single pustule isolates were collected from MI, NY, PA, VT, and WV, and are being characterized by rDNA sequencing. Preliminary data show that both M. americana and *M. paradoxa* are present in the shrub willow growing region. Further investigation of Melampsora spp. diversity throughout the Northeast will be conducted using genotyping-bysequencing, a genome wide high-throughput SNP identification method, providing thousands of markers for population studies targeting un-methylated coding sequences. Additionally, we explored disease resistance by surveying association and F<sub>2</sub> linkage mapping populations (Fig. 1) of S. purpurea for rust severity in September 2015 with the aim of identifying QTL and tightlylinked SNPs for marker-assisted selection. Preliminary data show a significant QTL for rust severity that maps to a common locus on Chr01 in both the linkage and association populations. GWAS identified a significant QTL on Chr02, while linkage mapping identified QTL on Chr05 and Chr10. There is also good evidence that major, qualitative resistance genes to Melampsora can be introgressed from diverse Salix species through hybridization. In order to map rust resistance genes across diverse species, we have developed eight species hybrid populations

produced by crossing reference female (94006) and male (94001) *S. purpurea* genotypes with six different *Salix* species. These have been established in a replicated field trial, will be genotyped by GBS, and phenotyped for rust incidence to map and characterize resistance loci. The parents of these mapping populations have been challenged with three isolates of *M. americana* and one isolate of *M. paradoxa* and display a range of resistance, including some progeny that are more resistant than the *S. purpurea* parents. This project will generate markers tightly linked to both major resistance genes and to QTL involved in quantitative resistance that will be developed into tools for early marker-assisted selection in the introgression of major resistance genes, while also selecting long-term for durable resistance in improved cultivars of this long-lived perennial bioenergy crop.



**Figure 1.** Frequency distribution of rust severity (log-transformed) in the  $F_2$  mapping population. Parents and grandparents of the  $F_2$  progeny are highlighted within bins along the frequency distribution. Log-transformed rust severity (%) was subsequently used for mapping.

*This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2015-67009-23957.* 

# Genomics and Phenomics to Identify Yield and Drought Tolerance Alleles for Improvement of Camelina as a Biofuel Crop

**John M. Dyer**<sup>1</sup>\* (john.dyer@ars.usda.gov), Hussein Abdel-Haleem,<sup>1</sup> Daniel Schachtman,<sup>2</sup> Yufeng Ge,<sup>2</sup> Toni Kutchan,<sup>3</sup> Noah Fahlgren,<sup>3</sup> Sheeja George,<sup>4</sup> and Russ Gesch<sup>5</sup>

<sup>1</sup>USDA-ARS, US Arid-Land Agricultural Research Center, Maricopa, AZ; <sup>2</sup>University of Nebraska, Lincoln, NE; <sup>3</sup>Donald Danforth Plant Science Center, St. Louis, MO; <sup>4</sup>University of Florida, Quincy, FL; and <sup>5</sup>USDA-ARS, North Central Soil Conservation Research Laboratory, Morris, MN

Project Goals: Plant oils represent renewable sources of energy-dense hydrocarbons that can be used for biofuels, but a major challenge is to produce these oils in non-food crops that have high yields and can grow under marginal conditions. Our goal is to improve the suitability of camelina as a non-food bioenergy crop. The major objectives are to: 1) Develop and apply automated, non-destructive high-throughput phenotyping (HTP) protocols to evaluate the phenotypic diversity of a camelina panel consisting of 250 accessions, grown under well-watered and water-limited conditions; 2) Discover alleles/genes controlling morphological, physiological, seed, and oil yield traits using genome-wide association studies (GWAS); and 3) Identify, test, and validate useful germplasm under diverse environments and marginal production areas. These studies will enable the discovery of new genes associated with crop yield and stress tolerance, and identify high-yielding cultivars that are suitable for certain geographical regions.

The aim of many breeding programs is to increase seed yield under stress conditions. Abiotic stress tolerance and seed yield are complex quantitative traits that are controlled by genetic, physiological and environmental factors making the identification of the underlying genetic structure difficult. Genomics-based approaches, including GBS and genome-wide association studies, offer great potential for elucidating the networks that underpin these complex traits. A major challenge, however, is that hundreds to thousands of closely related plant lines need to be phenotyped to have the statistical power to identify marker/trait associations. Therefore, these approaches require robust methods for high-throughput phenotyping, which is currently recognized as a major bottleneck for genomics-based crop improvement.

Several research groups, including ours, are developing and advancing the use of cutting-edge, high-throughput phenotyping technologies in both the field and greenhouse (Andrade-Sanchez et al., 2014; Fahlgren et al., 2015; Pauli et al., 2016). In this project, we will use both approaches to study a large and diverse population of camelina accessions grown under well-watered and water-limited conditions. GBS and GWAS will then be used to discover useful gene/alleles that influence crop performance and seed yields in both favorable and water-limited environments.

Greenhouse-based HTP will be conducted at the Donald Danforth Plant Science Center using a custom controlled-environment Conviron growth house that is fully integrated with a LemnaTec

Scanalyzer 3D<sup>HT</sup> conveyor and imaging system that provides quantitative, non-destructive, and time-series phenotyping of up to 1,140 small to medium-sized plants. Environmental controls include temperature, relative humidity, light intensity, and water stress treatments. Imaging systems include visible (VIS) spectrum (color RGB cameras) and near-infrared (NIR) reflectance, along with a stand-alone saturating pulse-induced chlorophyll fluorescence (PSII) system. VIS imaging is used to measure leaf and shoot area to estimate above ground biomass, plant color, leaf morphology and shoot architecture, and time-series enable calculation of growth rates. NIR reflectance is used to measure the water status of plant shoot tissues. The PSII imaging system is used to measure the maximum quantum efficiency of photosystem II, which can be a useful measure of plant stress and as an estimate of photosynthetic performance.

The Danforth Center already has significant experience in analyzing camelina using this system. For example, the camelina panel was phenotyped using four replicates of each camelina accession (1,000 plants), from 5-45 days after planting. Imaging was done starting at day 5 after planting and continued for 40 days, resulting in ~279,000 images, which were analyzed using PlantCV, an open-source image analysis software developed at the Danforth Center. To determine conditions suitable for water stress treatments, a pilot experiment was conducted for two accessions, Licalla (high oil content genotype) and SRS933 (genotype used as sequence reference), with water treatments from 0-100% field capacity (10% steps), from 8-45 days after planting. The tested genotypes were able to survive at 20% field capacity or higher.

The camelina panel will also be grown in two different field locations including Maricopa, AZ and Scottsbluff, NE, under well-watered and water limited conditions. In preparation for these studies, the camelina panel was recently planted in Maricopa for seed multiplication. A pilot study is also being conducted using 10 genotypes, cultivated under well-watered and water-limited conditions, and will be analyzed using both conventional and HTP methodologies. The HTP system in Maricopa features a LeeAgra AvengerPro spray rig carrying various sensors for measuring canopy temperature, canopy height, spectral reflectance, and RGB and infrared imaging. LiDAR technology is also being developed for 3D canopy reconstruction. The Nebraska platform includes a manually operated cart-based system with complimentary sensors to those in Maricopa. Both phenotyping platforms include GPS systems and onboard weather stations. Data will be georeferenced and analyzed using custom algorithms and pipelines.

## References

1. Andrade-Sanchez et al., (2014). Development and evaluation of a field-based high-throughput phenotyping platform. *Functional Plant Biology*. 41:68-79.

2. Fahlgren et al., (2015). A versatile phenotyping system and analytics platform reveals diverse temporal responses to water availability in setaria. *Molecular Plant.* 8:1520-1535.

3. Pauli et al., (2016). Field-based high-throughput plant phenotyping reveals the temporal patterns of quantitative trait loci associated with stress-responsive traits in cotton. *G3 (Bethesda)*. 6:865-879.

*This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2016-67009-25639.* 

Title: Ecological Factors Affecting Carbon Flow from Surface Litter

Deanna Lopez<sup>1</sup>, Sanna Sevanto<sup>1</sup>, Thomas Yoshida<sup>1</sup>, and **John Dunbar**<sup>1\*</sup> (dunbar@lanl.gov)

<sup>1</sup>Los Alamos National Laboratory, Los Alamos, NM

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

**Project Goals:** The LANL Genomic Science SFA is focused on microbial communities in surface soil horizons and their functional processes that influence soil carbon storage and release. The SFA examines soil carbon cycling under conditions of environmental change to understand the metabolic and ecological roles of fungi and bacteria in surface soils in two important temperate biomes – forests and arid grass/shrub lands. In both biomes, fungal and bacterial biomass is concentrated in shallow surface soil strata where C and N cycling are major processes. Advancing fundamental knowledge of soil communities within the context of altered environmental regimes will improve our ability to predict and possibly manage ecosystem contributions to global climate. This involves discovery of fundamental principles at different scales that influence the organization, interactions, and response of soil communities.

For the work presented in this poster, our ongoing aim is to improve understanding of biological factors that influence soil carbon storage. The concentration of soil carbon declines with soil depth but the residence time increases dramatically, ranging from a few years at the surface to thousands of years in mineral horizons a meter below the surface. There are differing views on the source of the carbon stored long-term in deep soil horizons. One view is that dissolved organic carbon (DOC) from decomposition of surface litter percolates to deeper horizons and binds to protective mineral layers. A second view is that processes linked to plant roots are the primary source. A third view is that both phenomena above contribute to long-term carbon storage in deep horizons, but the relative contribution of each process is unknown. At present, we are focusing on fluxes from decomposing surface litter, examining factors such as initial microbial biomass abundance, succession, and priority effects that control carbon flow and fate.

We found complex relationships between initial biomass abundance and carbon flux. In this work, sterile plant litter in sand microcosms was inoculated with dilutions of soil microbial communities from two different soils. After 30 days of litter decomposition, bacteria diversity correlated with the ten-fold reductions in initial microbial abundance. In contrast, the fungal community diversity at day 30 was similar across treatments, consistent with the outgrowth and convergence of fungal-dominated decomposer guilds. A surprising outcome was that CO<sub>2</sub> efflux rates varied with microbial community abundance. Despite rapid initial responses in all communities, lower initial community abundance resulted in extended period of reduced cumulative CO<sub>2</sub> release. Every ten-fold drop in initial biomass abundance resulted in a 6 to 13% reduction in cumulative CO<sub>2</sub> output over the incubation period, with variation depending on the source of the parent microbial community. These results may account for the inconsistent and conflicting outcomes reported in the recent body of studies on the "home-field advantage" hypothesis (i.e. plant litter is decomposed more efficiently by native communities than by

foreign communities). Comparing outcomes from communities that differ in initial abundance provides insight into the role of less abundant species on carbon flow.

<u>Funding statement</u>: The information in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2017LANLF260).

Title: Microbial Community Traits Linked to Carbon Flux Patterns in Soil

Renee Johansen,<sup>1,\*</sup>(rjohansen@lanl.gov), Deanna Lopez<sup>1</sup>, Rebecca Mueller<sup>1,2</sup>, Thomas Yoshida<sup>1</sup>, and **John Dunbar**<sup>1</sup>

<sup>1</sup>Los Alamos National Laboratory, Los Alamos, NM <sup>2</sup> Present address, Montana State University, Bozeman, MT

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

**Project Goals:** The LANL Genomic Science SFA is focused on microbial communities in surface soil horizons and their functional processes that influence soil carbon storage and release. The SFA examines soil carbon cycling under conditions of environmental change to understand the metabolic and ecological roles of fungi and bacteria in surface soils in two important temperate biomes – forests and arid grass/shrub lands. In both biomes, fungal and bacterial biomass is concentrated in shallow surface soil strata where C and N cycling are major processes. Advancing fundamental knowledge of soil communities within the context of altered environmental regimes will improve our ability to predict and possibly manage ecosystem contributions to global climate. This involves discovery of fundamental principles at different scales that influence the organization, interactions, and response of soil communities.

For the work presented here, our ongoing aim is to identify microbial community traits linked to different carbon flux patterns. When new organic carbon enters soil ecosystems, a fraction is respired and a fraction is stored for some period of time. The balance between these two general fates determines if soil is a carbon source or sink. Globally, the strongest factors controlling flux patterns are physical-chemical variables (temperature, oxygen, and pH) and biophysical processes (aggregate formation and sorption onto protective mineral surfaces). However, microbial community composition also plays a role.

We found that carbon flux patterns can vary dramatically with subtle changes in the composition of complex microbial communities. To achieve this, we monitored carbon flux patterns from plant-litter-decomposing microbial communities derived from 200 soils from the dry grassland biome of the United States. We exploited the fact that soil microbial communities in a common biome change subtly with spatial distance. Liquid extraction of communities from each sample and transplantation into a common environment (sand microcosms) enabled evaluation of different community configurations without additional confounding factors such as geochemistry and plant cover. We found that communities representing the most divergent carbon cycling outcomes had common traits. This is the first demonstration of specific microbial community traits linked to divergent soil carbon cycling patterns in the same environment.

<u>Funding statement</u>: The information in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2017LANLF260). **Title:** Comparative Analyses of the Genomes and Secretomes of Ascomycota Fungi Reveal Diverse Functions in Plant Biomass Decomposition.

Jean F. Challacombe<sup>1</sup>\* (jchalla@lanl.gov), Cedar Hesse<sup>1,2</sup>, Gary Xie<sup>1</sup>, Lisa M. Bramer<sup>3</sup>, Lee Ann McCue<sup>3</sup>, John Dunbar<sup>1</sup>, **Cheryl R. Kuske<sup>1</sup>** 

<sup>1</sup> Los Alamos National Laboratory, Los Alamos, NM; <sup>2</sup> Present address, USDA Agricultural Research Service, Corvallis, OR; <sup>3</sup> Pacific Northwest National Laboratory, Richland, WA

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

**Project Goals:** The LANL Genomic Science SFA is focused on microbial communities in surface soil horizons and their functional processes that influence soil carbon storage and release. The SFA examines soil carbon cycling under conditions of environmental change to understand the metabolic and ecological roles of fungi and bacteria in surface soils in two important temperate ecosystems – forests and arid grass/shrub lands. In both biomes fungal and bacterial biomass is concentrated in shallow surface soil strata where C and N cycling are major processes. Advancing fundamental knowledge of soil communities within the context of altered environmental regimens will improve our ability to predict and possibly manage ecosystem contributions to global climate. This involves discovery of fundamental principles at different scales that influence the organization, interactions, and response of soil communities.

In arid grasslands and shrublands, the dominant fungi in surface soils are members of the Ascomycota phylum. Their functions in arid soils, where organic matter, nutrients and water are very low or only periodically available, are unknown. Potential roles include seasonal plant biomass decomposition, direct interactions with plants as endophytes or pathogens that induce selective disassembly of plant tissues, or as integral members of cyanobacteria-dominated biological soil crusts. We isolated and taxonomically typed several thousand fungal isolates from an arid grassland. The genomes of five Ascomycota genera that were abundant in multiple microhabitats were sequenced. Their secreted proteomes (secretomes) when grown on different carbon sources (chitin, native bunchgrass or pine wood) were determined in replicated cultures, through a collaborative project with the Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory. Comparisons of the genomes and secretomes of these five fungi have revealed similarities and differences in their secretomes, also with respect to particular carbon growth substrates, that will advance our fundamental knowledge of the roles that soil fungi play in microbial communities.

Ascomycota genomes were assembled using Velvet version 1.2.10 [1], and gene prediction was accomplished using Augustus version 3.0.3 [2]. Protein coding sequences were functionally annotated by BLASTP [3] against the nr database. For each gene, function was automatically assigned based on the top hit using an in-house script. Protein spectral counts (representing protein abundances) were obtained and mapped to predicted proteins in each fungal genome by the EMSL proteomics resource. Spectral count data were averaged across the technical replicates for each fungus and each treatment; the means, standard deviations, standard errors were calculated. Differences in protein expression were assessed by fold change analysis. For each

fungus grown on each carbon substrate, the fold change of the protein counts for each condition was calculated compared to each other condition and pairwise p-values were calculated. The volcano plots below in Figure 1 show some of the results.



Figure 1. Volcano plots showing results of the Fold Change analysis of protein spectral counts for two of the fungal species sequenced and analyzed for this study: Coniochaeta CK134 and Chaetomium CK152. The plots emphasize the magnitude of differences in protein expression for each fungus grown on grass as the main carbon source compared to pine. The values on the x-axis represent the  $log_2(FoldChange)$  and on the y-axis the Significance values represent  $-log_{10}(p-value)$ . These conversions give the data a nice spread for visualization. The dots represent actual proteins in each genome, colored by the Functional Categories in the legend. NOTE: Functional Categories related to plant biomass decomposition are colored in shades of green. The dots to the left of the 0 line represent decreased expression, while dots to the right of the 0 line represent increased expression, when the fungus was grown on grass compared to pine.

Each of the five fungi secreted a characteristic set of enzymes when grown on individual carbon substrates. While homologs of some secreted proteins were identified in more than one fungal genome, the overall secretome of each fungus showed differences. Results of this study will increase our understanding of the contributions of fungal carbon metabolism to arid land microbiomes and ecosystems, and will contribute to the identification of fundamental principles at different scales that influence the organization, interactions, and response of soil communities to variations in their environment.

## References

- [1] D. Zerbino, and E. Birney, Genome Research 18 (2008).
- [2] M. Stanke, A. Tzvetkova, and B. Morgenstern, Genome Biol. 7 Suppl 1 (2006).
- [3] S. F. Altschul *et al.*, J. Mol. Biol. **215** (1990).

Funding statement: The information in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2017LANLF260).

**Title:** Meta-omic, Enzymatic and Soil Chemical Measures Identify Surface Soil Decomposition Processes as Influenced by N Amendment in a Temperate Pine Forest

**Cheryl R. Kuske<sup>1</sup>**\* (<u>Kuske@lanl.gov</u>), Robert L. Sinsabaugh<sup>2</sup>, Rebecca Mueller<sup>1,3</sup>, La Verne Gallegos-Graves<sup>1</sup>, Lawrence O. Ticknor<sup>1</sup>, John Dunbar<sup>1</sup>

<sup>1</sup> Los Alamos National Laboratory, Los Alamos, NM

<sup>2</sup> University of New Mexico, Albuquerque, NM

<sup>3</sup> Present address, Montana State University, Bozeman, MT

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

**Project Goals:** The LANL Genomic Science SFA is focused on microbial communities in surface soil horizons and their functional processes that influence soil carbon storage and release. The SFA examines soil carbon cycling under conditions of environmental change to understand the metabolic and ecological roles of fungi and bacteria in surface soils in two important temperate ecosystems – forests and arid grass/shrub lands. In both biomes fungal and bacterial biomass is concentrated in shallow surface soil strata where C and N cycling are major processes. Advancing fundamental knowledge of soil communities within the context of altered environmental regimens will improve our ability to predict and possibly manage ecosystem contributions to global climate. This involves discovery of fundamental principles at different scales that influence the organization, interactions, and response of soil communities.

Anthropogenic N deposition is a chronic and increasing condition in temperate regions that may strongly influence C cycling dynamics. Our goal is to provide an understanding of the active and responsive components of temperate forest surface soils that contribute to C cycling and their collective responses to environmental change. This poster describes the current status of our field experiments in a temperate pine forest (Duke forest, NC) after eleven years of N fertilization (annual application of ammonium nitrate pellets, 11.2 g N m<sup>-2</sup>). To link microbial community structure with functions in response to N fertilization, we measured total, fungal and bacterial biomass (by rRNA gene qPCR); fungal and bacterial community richness, diversity and composition (by MiSeq rRNA gene sequencing); activities of seven exoenzymes in the total community measured in soil; and collective metatranscriptome gene expression of the bacterial, archaeal, and fungal communities. The field site was extensively sampled with at least 10 field replicates for each of three soil horizons (Oa, Oe, and A horizons to 10 cm depth), in each treatment plot. We present correlations among these datasets and with soil chemistry measures, and will present our inferences about the resident fungal and bacterial communities, their enzyme activities, and local geochemistry across shallow soil horizons in the pine forest.

All soil chemistry measures except pH and nitrate declined with depth over 10 cm. This included organic matter, P, K, metal ions, total N and total C. The long term N amendment depressed soil pH, but increased nitrate, P, K, Fe and Mn concentration in each horizon. Biomass, measured as extractable DNA, decreased significantly with depth and with N-amendment. Bacteria SSU rRNA gene copy number followed the same trend. Fungal SSU RNA gene copy number dropped

an order of magnitude across a few cm depth with no effect of N-amendment. Hydrolase activities per g soil declined sharply with soil depth, largely in response to declining SOC. In contrast, oxidase activities increased with soil depth on both a dry mass and SOC basis. A vector analysis of hydrolytic EEA responses to N treatment showed significant increase in vector length, indicating a relative increase in C limitation, and vector angle, indicating increased P limitation relative to N. For all horizons, N treatment suppressed phenol oxidase (POX) activity relative to B-galactosidase (BG) activity. As a result, the BG/POX ratio increased by 43%, 184%, and 55%, for Oa, Oe and A horizons, respectively. This trend was similar to that for the hydrolytic C:N activity ratio BG(NAG+LAP), suggesting that increased N availability also reduced the potential for N mining from recalcitrant SOM. Carbon use efficiency (CUE) increased significantly with soil depth and decreased with N treatment.

Bacterial community richness and diversity were negatively and significantly reduced, and bacterial community composition was altered, in N fertilized plots at all three depths. In contrast, the fungal community richness and diversity were unaffected by N fertilization in the top two depth intervals, and were only enriched in the A horizon. The fungal community composition was significantly different across soil depth and with N fertilization. Twenty-seven of the thirty most abundant OTUs were Basidiomycota. Eight genera were significantly, positively affected by the N fertilization and eight genera were negatively affected; all were Basidiomycota genera that function as ectomycorrhizal associates with pine or as wood rotters.

In summary, across a 10 cm soil depth column, (i) all of the soil chemistry measures were statistically different with depth. Correspondingly, (ii) the profile of enzyme activity, including enzymes responsible for C, N and P acquisition and oxidative decomposition, differed with soil depth. In response to N amendment, patterns of enzyme activity suggest a shift to greater C and P acquisition activity, and a reduced potential for mining C from recalcitrant organic matter and N acquisition activities. In the metatranscriptomes, (iii) transcripts representing these enzyme activities showed similar depth and N-response trends, and general patterns of CAZyme transcripts shifted across depth and with N fertilization. Both fungal and bacterial transcripts contributed to the observed transcriptome patterns. (iv) Fungal and bacterial community biomass was reduced with soil depth and with N fertilization, and the composition of both communities shifted with N fertilization at each depth interval.

The concerted changes in soil chemistry, microbial biomass, gene expression and enzyme activities across the shallow depth gradient, and in response to N fertilization, illustrate the dramatic impact that long-term N amendment may have on C cycling and storage processes in surface soils of temperate forests. The results presented here illustrate how important it is to select appropriate sampling sites and depths when trying to map local soil processes, and when defining soil traits that may be useful for establishing soil processes. Forests deposit considerable organic matter to the soil surface; the extent to which this organic matter is decomposed or stored in the soil as fixed carbon is an important parameter to understand when predicting responses of natural forest ecosystems to climate changes.

<u>Funding statement:</u> The information in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2017LANLF260).

# Plant-Microbe Interfaces: Populus genomics, genetics and molecular biology

**Jay Chen**<sup>1\*</sup> (<u>chenj@ornl.gov</u>), Wellington Muchero,<sup>1</sup> Jessy L. Labbé,<sup>1</sup> David J. Weston,<sup>1</sup> Xiaohan Yang,<sup>1</sup> Udaya C. Kalluri,<sup>1</sup> Sara S. Jawdy,<sup>1</sup> Lee E. Gunter,<sup>1</sup> and **Gerald A. Tuskan<sup>1</sup>** 

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

# http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) developing metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Previous and ongoing PMI SFA research has led to the development and application of a number of Populus genomic and genetic resource tools for realizing the project goals of the PMI SFA. These include ~1,000 Populus trichocarpa natural variants for Genome-Wide Association Studies (GWAS), ~400 P. trichocarpa × P. deltoides hybrids for Quantitative Trait Locus (QTL) mapping, Populus protoplast transient expression system and Populus hairy root transient expression system for molecular, cellular and biochemical validation, and a Populus transformation platform for generating stable transgenic lines over- or under-expressing the gene of interest. These resources and tools have enabled us to identify genetic loci regulating the interactions between Populus and microbes and to characterize these interactions at the molecular and biochemical levels. In particular, they have enabled us to effectively evaluate the host genotype influence on microbial community composition, diversity and function, and to pinpoint the action of these genetic regulators in the signaling events in *Populus*-microbial interactions. We have identified a lectin receptor-like kinase as a key regulator of Populus-Laccaria bicolor interactions. By using Populus transgenic lines and a bacterial lactonase, we will evaluate the role of quorum sensing in shaping the composition and structure of the *Populus* microbiome.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

# Plant Microbes Interfaces: Probing the molecular mechanisms of plant-bacterial interactions

**Caroline S. Harwood**<sup>1\*</sup> (<u>csh5@uw.edu</u>), Bruna G. Coutinho,<sup>1</sup> Yasuhiro Oda,<sup>1</sup> Monica R. Cesinger,<sup>1</sup> Amy L. Schaefer,<sup>1</sup> Dale Pelletier,<sup>2</sup> Jennifer Morrell-Falvey,<sup>2</sup> **E. Peter Greenberg**,<sup>1</sup> and Mitchel Doktycz<sup>2</sup>

<sup>1</sup>University of Washington, Seattle WA; <sup>2</sup>Biosciences Division, Oak Ridge National Laboratory, Knoxville TN

## http://pmi.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers of diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

An analysis of 154 genome sequences of Proteobacteria isolated from root tissues of the *Populus deltoides* revealed that many (n=32) of them encode a transcription factor that is closely related to a protein named OryR from the rice pathogen Xanthomonas oryzae. OryR detects an unknown plant compound and when bound by the plant ligand, activates transcription of genes involved in virulence. We have been investigating an OryR homolog, which we call PipR, in the *Populus* non-pathogenic endophyte *Pseudomonas* sp. GM79. Our recently published work showed that GM79 PipR activates expression of the downstream peptidase gene, *pipA*, in response to plant leaf macerates, peptide-rich hydrolysates (peptone), and high concentrations of a specific tripeptide (ser-his-ser). Recent work has focused on identifying compounds that can activate *pipA* in a PipRdependent manner at concentrations that might be present in plants. Mutant analyses of genes flanking pipR suggest that the PipR signal(s) enter the bacterial cells by active transport via an ABC-type transporter. The periplasmic binding protein component of the ABC-type transporter binds the plant and peptone signal(s) tightly, and we used this as a tool to purify and characterize the signal(s) from peptone. We identified several D-isomer peptides that can activate the PipR system, some at physiologically relevant (low mM) concentrations. The finding that PipR is most responsive to D-form peptides suggests similar compounds could be the active signal in *Populus* macerates. Plants are known to produce a variety of peptides that are involved in signaling for the regulation of growth and development and in defense against pathogen infection. Our findings raise the

possibility that plant-produced peptides may also be important for establishing beneficial plant-microbe interactions. We have also identified PipR homologs in other members of the *Populus* microbiome and have surveyed them for PipR activity. We believe that a better understanding of these PipR-type plant signal receptors and their plant signals is of general importance as they occur in dozens of bacterial species that are associated with economically important plants. As a complementary project, we are interested in understanding the role of secondary metabolites in the PMI microbiome and have identified a variety of potentially interesting gene clusters using antiSMASH analyses.

# Plant-Microbe Interfaces: Plant-based Genome-Wide Association Viriome/Microbiome Analysis (GWAVA)

Piet Jones,<sup>1,2</sup> Wellington Muchero,<sup>1</sup> Jay Chen,<sup>1</sup> Dale Pelletier,<sup>1</sup> Gerald Tuskan<sup>1\*</sup> (<u>tuskanga@ornl.gov</u>), Carissa Bleker,<sup>1,2</sup> Marilyn Roosinck,<sup>3</sup> and **Daniel Jacobson**<sup>1,2</sup>

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; <sup>2</sup>University of Tennessee, Knoxville, TN; <sup>3</sup>Pennsylvania State University, University Park, PA

## http://pmi.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Metagenomic or metatranscriptomic analyses can capture the diversity and functional potential of a microbiome, though viruses are often a neglected component. The microbiome (here defined as the combination of the bacteria, fungi, archaea and viruses found in plant tissue) not only interacts with the host, and thus may elicit or suppress a defense response, but there may be interactions among the constituents of the microbiome. This multi-directional influence can have a number of phenotypic responses, both pathogenic as well as mutualistic. To better understand these influences we have applied modified genome-wide association (GWAS) methods in order to associate host genotypic variation to the putative viral, bacterial, fungal and archaeal community composition. Metatranscriptome samples extracted from *Populus trichocarpa* xylem tissue were used to identify the different taxa present via several computational approaches. The resultant members of the microbiome then served as phenotypes in a GWAS analysis involving 444 genotypes of *Populus trichocarpa* (against over 8 million single-nucleotide polymorphisms). After correcting for multiple hypotheses testing, and viewing the results in a network context, we find shared associations between microbiome phenotypes and receptor kinases, signal transduction genes, transcription factors, xenobiotic stress, biotic stress, cell organization, biological control agents and other stress related genes among other functions. The resultant network provides a rich framework for biologically driven hypothesis generation. Thus the use of microbiome constituents as GWAS phenotypes is helping to elucidate the host mechanisms responsible for host-microbiome interactions (and the regulation thereof) and provides indications of possible interactions among different members of the microbiome itself.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

# Plant-Microbe Interfaces: Comparative genomics and functional characterization of *Populus*-associated endophytes

**Jennifer Morrell-Falvey**<sup>1\*</sup> (<u>morrellj11@ornl.gov</u>), Miriam Land,<sup>1</sup> Amber Bible,<sup>1</sup> Carmen Foster,<sup>1</sup> Dale Pelletier,<sup>1</sup> and **Mitchel Doktycz**<sup>1</sup>

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

# http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

*Populus* trees host a diverse community of microbes found closely associated with the plant roots (rhizosphere) and inside plant tissues (endosphere). These complex communities are shaped by the functional diversity expressed by their members as well as the environmental forces acting on them. Previous studies have shown that the microbial community found inside the plant is less diverse than the host-associated rhizosphere community, consistent with the idea that the endosphere represents a unique environmental niche. The mechanisms by which some microbes are able to bypass plant defenses and survive within plant tissues, however, are not well understood. To better understand what makes an endophyte an endophyte, we are using a combination of comparative genomics, microbial physiology, and molecular genetics to characterize and identify genetically tractable model endophytes to study the molecular mechanisms that promote an endophytic lifestyle. Starting with a collection of 30 genome sequenced endophytes isolated from *Populus*, we are characterizing plant-associated microbial behaviors including phytohormone production, siderophore production, motility, biofilm formation, and plant colonization. In addition, we are determining whether these isolates are genetically tractable by constructing fluorescent strains through plasmid transformation or chromosomal integration. This collection also includes thirteen Variovorax isolates, twelve of which were isolated from the endosphere and one from the rhizosphere of *Populus*. Genomic comparisons indicate that these isolates are diverse within the Variovorax genus, despite being isolated from a common host. We show that the rhizosphere-isolated strain Variovorax sp.

YR216 is the most distantly-related of the *Populus*-associated isolates. Unlike the endophyte isolates, *Variovorax* sp. YR216 lacks ACC deaminase activity and does not efficiently colonize root tissue. By combining comparative genomic analyses with molecular genetics, these studies will provide insights into the microbial pathways and signaling events involved in plant host recognition and survival.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research.
## Plant-Microbe Interfaces: Understanding the *Populus* microbiome structure in response to host stress

Collin M. Timm,<sup>1</sup> Gregory B. Hurst<sup>2\*</sup> (<u>hurstgb@ornl.gov</u>), Alyssa A. Carrell,<sup>1</sup> Sara Jawdy,<sup>1</sup> Lee E. Gunter,<sup>1</sup> Charles J. Doktycz,<sup>2</sup> Keiji G. Asano,<sup>2</sup> Christopher W. Schadt,<sup>1</sup> Nancy L. Engle,<sup>1</sup> Timothy J. Tschaplinski,<sup>1</sup> **David J. Weston**,<sup>1</sup> **Dale A. Pelletier**,<sup>1</sup> Gerald A. Tuskan,<sup>1</sup> and **Mitchel J. Doktycz**<sup>1</sup>

<sup>1</sup>Biosciences Division and <sup>2</sup>Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

#### http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Adverse conditions can affect plants, leading to decreased plant growth, productivity, and survival, and resulting in failure or poor yields of crops and biofeedstocks. In some cases, the microbial community associated with plants has been shown to alleviate plant stress and increase plant growth under suboptimal growing conditions. A systematic understanding of how the microbial community changes under these conditions is required to understand the function of the phytobiome system and the contribution of the microbiome to water utilization, nutrient uptake, and ultimately yield. Using a microbiome replacement strategy, we studied how the microbiome of Populus deltoides changes in response to diverse environmental conditions of water limitation, light limitation (shading), and metal toxicity. While plant responses to treatments in growth, photosynthesis, gene expression and metabolite profiles were varied, we identified a core set of bacterial genera that change in abundance in response to host stress. The results of this study indicate substantial structure in the plant microbiome community and identify potential drivers of the phytobiome response to stress. We further investigated the hypothesis that different plant stresses would affect the complements of secondary metabolites produced by microbial communities associated with the roots. We extracted methanol-soluble compounds from the soil samples at the conclusion of the stress experiment, analyzed the

extracts using liquid chromatography-tandem mass spectrometry (LC-MS-MS), and examined the data as a function of applied stress. Many similar MS-MS spectra were obtained across all treatments, but subsets characteristic of each abiotic stress could also be identified.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research.

### Plant-Microbe Interfaces: Dissecting the microbiome of *Populus* tree species from the soil to the canopy using amplicon sequencing and shotgun metagenomic analyses

Melissa A. Cregger<sup>1\*</sup>(<u>creggerma@ornl.gov</u>), Allison M. Veach,<sup>1</sup> Miranda Crouch,<sup>1</sup> Ian Hodges,<sup>1,6</sup> Zamin K. Yang,<sup>1</sup> Debbie Weighill,<sup>1,2</sup> Piet Jones,<sup>1,2</sup> Carissa Bleker,<sup>1,2</sup> Armin Geiger,<sup>1,2</sup> Mircea Podar,<sup>1,2</sup> Rytas Vilgalys,<sup>3</sup> Timothy Rials,<sup>4</sup> Susannah G. Tringe,<sup>5</sup> Dale A. Pelletier,<sup>1,2</sup> Daniel A. Jacobson,<sup>1,2</sup> and **Christopher W. Schadt**<sup>1,2</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, TN; <sup>2</sup>Univ. of Tennessee, Knoxville, TN; <sup>3</sup>Duke Univ., Durham, NC; <sup>4</sup>Univ. of Tennessee Institute of Agriculture, Knoxville TN; <sup>5</sup>DOE Joint Genome Institute, Walnut Creek, CA, Stanford University, Palo Alto, CA

*Populus* trees are broadly distributed in nature, widely used in pulp and paper production, and a potentially important bioenergy feedstock. In recent efforts, we examined 30 plant tissue/habitat types extracted from plantation grown *P. deltoides* and hybrid *P. trichocarpa x deltoides* (TxD) trees. The microbiome composition was assessed by 16S-rRNA gene (bacteria) and ITS2-rRNA gene (fungal) measurements of 300 samples across multiple tissue types and tissue ages. For select habitats (root endosphere, rhizosphere, and soil), we also applied a differential and density gradient centrifugation method to enrich the microbial cells from plant tissues prior to DNA extraction and shotgun metagenomic sequencing. For each tree type, bacterial and fungal community structure from rRNA gene amplicons varied significantly across leaf, stem, roots and soil/rhizosphere tissue/habitat types. Leaf habitat types had significantly lower OTU richness compared to root and stem habitats and featured decreasing abundance of Proteobacteria from leaf, to stem, to root, to soil. The oldest woody stem tissues were also distinguishable from the younger 1<sup>st</sup> and 2<sup>nd</sup> year tissues, and enriched in *Firmicutes*, consistent with potentially anaerobic environments. Within the leaf, stem, root, and soil habitats, bacterial community structure in P. *deltoides* samples could be differentiated from TxD hybrid samples. Interestingly, the leaf environment from the TxD hybrid samples had high levels of fungal pathogens relative to the P. *deltoides* leaf samples (primarily *Septoria musciva* and a *Marsonina* sp.). These data suggest the differential abundance of these pathogens between the species, may be driving changes in the overall microbiome structure, and follow on investigations are underway to address this. Comparative metagenomic analyses and amplicon-based datasets were completed for root endophyte, rhizosphere and soil samples and show a high degree of congruence and confirm our methods reduced host DNA contamination to < 2%. Metagenome functional gene profiles based on ordination and network analyses of PFAM enrichment patterns, show distinct clustering based on sample type (endosphere, rhizosphere, and soil) and host genotype (P. deltoides vs. TxD). Results from this and future work should greatly enhance our community-based and functional understanding of plant microbiomes in this important model species.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research.

## Plant-Microbe Interfaces: Linking host genotype fitness and soil conditions to microbiome community assembly in the *Populus* root – soil interface

Veach, A.,<sup>1</sup> M. Cregger,<sup>1</sup> Z. Yang,<sup>1</sup> D. Yip,<sup>1</sup> N. Engle,<sup>1</sup> T. Tschaplinski,<sup>1</sup> J. Labbé,<sup>1</sup> G. Tuskan,<sup>1</sup> and **C. Schadt**<sup>1\*</sup> (schadtcw@ornl.gov)

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

#### http://pmi.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

*Populus* is a commercially and ecologically important trees species as well as a potential biofuel feedstock source. Populus species commonly occur in riparian areas throughout most of North America where they are considered a keystone species. Populus have become an important model host system to study plant-microbe interactions due to their amenability to laboratory and greenhouse experimental manipulation, as well as molecular biology tools enabled from genome sequencing. We conducted a trap-plant experiment using 12 Populus genotype clones that vary in higher-order salicylate concentration and composition (e.g., salicortin, salicin, tremuloidin, populin) to better understand how genotype, chemotype and phenotype of the host plant might influence plant and soil microbiome assembly and composition. To assess the relative importance of host genotype vs. soil conditions and origin, we planted 5 replicate cuttings per genotype (N = 120) in 2 soil types (2:1 sterile sand:soil inoculum) collected from two Oregon locations (Corvallis and Clatskanie) and allowed plants to grow for 4 months. At the end of the experiment, leaf chlorophyll content, leaf growth (number of new leaves since transplant), and net photosynthetic rate differed across genotypes and soil type (P < 0.01). In addition, ectomycorrhizal colonization of root tips differs between genotypes measured thus far (ranges between 0 - 17%). Host chemotypes were confirmed for roots from experimental samples via GS-MS analysis, and showed host total salicylate concentrations varying from 1221 - 10610  $\mu$ g/g fresh weight (FW), tremuloidin varying from 17 – 225  $\mu$ g/g FW, and populin ranging from  $0 - 8.9 \,\mu\text{g/g}$  FW, depending on plant genotype. Microbiome data collection is ongoing and will be investigating belowground (root endosphere and rhizosphere soils) bacterial and fungal community composition effects of host chemotype, genotype and soils via Illumina MiSeq targeted amplicon sequencing, population sizes as detected by qPCR of 16S and ITS2 rRNA genes, and mycorrhizal colonization rates.

## Plant-Microbe Interfaces: Metabolic consequences of the introduction of a *Populus trichocarpa* lectin receptor-like kinase into *Arabidopsis thaliana*, a non-ectomycorrhizal host species

**Timothy Tschaplinski**<sup>1\*</sup>(tschaplinstj@ornl.gov), Zhihao Zhang,<sup>1</sup> Jessy Labbé,<sup>1</sup> Wellington Muchero,<sup>1</sup> Aurélie Deveau<sup>2</sup>, Claire Veneault-Fourrey<sup>2</sup>, Yongil Yang,<sup>1</sup> Priya Ranjan,<sup>2</sup> Sara Jawdy,<sup>1</sup> Daniel Jacobson,<sup>1</sup> Jin-Gui Chen,<sup>1</sup> Francis Martin<sup>2</sup>, Jerry Tuskan,<sup>1</sup> and Mitchel Doktycz<sup>1</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, TN; and <sup>2</sup>University of Tennessee-Knoxville, Knoxville, TN; <sup>2</sup>INRA, UMR 1136 INRA-University of Lorraine, Interactions Arbres/Microorganismes, Laboratory of Excellence ARBRE, INRA-Nancy, 54280, France

#### http://pmi.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

A lectin receptor-like kinase (RLK) was identified in black cottonwood (*Populus trichocarpa*) that facilitates the initiation of a symbiotic ectomycorrhizal association with *Laccaria bicolor*. This *Populus* RLK was introduced via genetic transformation into *Arabidopsis thaliana*, a non-ectomycorrhizal host species, generating transgenic lines; LK8, LK100. The metabolomic profiles of mature transgenic *Arabidopsis* plants were determined by gas chromatography-mass spectrometry (GC-MS) and contrasted with wild-type 'Col' plants. The effects of transgene insertion on the metabolomic profiles were determined + and – *L. bicolor*, a *Populus* fungal symbiont. Additionally, given that this *Populus* RLK is predicted to bind mannose, the metabolomic profiles of plants with and without the RLK gene were grown + and – a brief (4 h) exposure to mannose vs glucose.

*L. bicolor* generates an intense defense response in *Arabidopsis*. Metabolites up-regulated in response included defense metabolites histidine, sinapoyl malate, kaempferol, quercetin, sinapic acid-4-O-glucoside, other sinapoyl conjugates, but sterols, such as cholesterol and campesterol, and monolignol glucosides, coniferin and syringin, were also elevated. Many of the highest up-regulated metabolomic fold changes associated with exposure to *L. bicolor* were suppressed by

the RLK gene insertion. Exposure to *L. bicolor* alone also induces large declines nitrogenous compounds and organic acids, metabolites, previously assumed to be consumed by the host's symbiont. Also of note are reductions in defense signaling and related metabolites, including salicylic acid and the defense priming metabolite, azelaic acid. These declines are likely indicative of altered C partitioning to defense, with RLK, again, lessening the degree of these responses. Furthermore, mannose addition tended to reduce the fundamental transgene upregulated responses (especially in the weaker expressing LK100), whereas glucose had less of an effect.

In summary, *L. bicolor* induces an intense defense response in non-host, wildtype *Arabidopsis* 'Col', but these defense responses were suppressed by the presence of RLK gene. Interestingly, many of the highest metabolite fold changes associated with exposure to *L. bicolor* are the same as those driven solely by RLK gene, indicating that gene insertion, itself, triggers a defense response. The presence of mannose, a putative ligand, reduced the magnitude of the gene-induced defense response. The concomitant declines in a number of fatty acids and amino acids by both *L. bicolor* and RLK gene insertion that are typically attributed to utilization by the fungal symbiont, are likely the result of the altered C partitioning to defense. The RLK gene primary effect is to suppress the defense response induced by the potential fungal symbiont, thereby, facilitating the symbiosis.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

## Plant-Microbe Interfaces: Characterizing the diversity and function of the ectomycorrhizome of *Populus trichocarpa*

Alejandro Rojas-Flechas,<sup>1</sup> Brian Looney,<sup>2,3</sup> Sunny Liao,<sup>1</sup> Jay Chen,<sup>2</sup> Jessy Labbé,<sup>2</sup> and Rytas Vilgalys<sup>1</sup>\*(<u>fungi@duke.edu</u>)

<sup>1</sup>Duke University, Durham, NC; <sup>2</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; <sup>3</sup>University of Tennessee, Knoxville, TN;

#### http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The *Populus* root microbiome harbors a diverse community of ectomycorrhizal fungi (EMF) that significantly increases nutrient uptake and acquisition by the plant host while also providing protection against antagonistic parasites. Over 30 genera of EMF are known to associate with *Populus*, including many groups of mushroom-producing families including *Boletaceae*, *Russulaceae*, *Cortinariaceae*, *Tricholomataceae*, and *Amanitaceae*. A major aim of the PMI project is the collection, isolation, and characterization of the major EMF fungal associates of *P. trichocarpa* across its range in the Pacific Northwest. In Autumn 2015 and 2016, we surveyed macrofungal diversity under native *P. trichocarpa* forests from five core watersheds on the Squamish (BC), Snohomish (WA), Puyallup (WA), Columbia (OR and WA), and Willamette (OR) rivers. This resulted in over 150 collections of EMF fruit body collections and sampling of bulk soil used in bioassay studies. All sporocarp collections were plated on modified Melin-Norkrans medium, photographed, spore printed, and dried for identification and accession into fungal herbaria. The first fungal collections have been identified while the second samples from the second campaign are being identified by a consortium of taxonomic experts using the ITS barcode marker as well as morphological features.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

## Plant-Microbe Interfaces: Model communities of *Populus* bacterial isolates to study mechanisms of microbiome function

**Dave Weston**<sup>1\*</sup> (westondj@ornl.gov), Collin Timm,<sup>1</sup> Alyssa Carrell,<sup>1</sup> Sara Jawdy,<sup>1</sup> Lee Gunter,<sup>1</sup> Shen Lu,<sup>1</sup> Nancy Engle,<sup>1</sup> Timothy Tschaplinski,<sup>1</sup> **Dale Pelletier**,<sup>1</sup> Gerald A Tuskan,<sup>1</sup> and Mitchel Doktycz<sup>1</sup>

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

#### http://PMI.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

The function of the plant-microbiome system is a result of complex interactions among microbiome members, host plant and the environment. Direct experimentation on the microbiome is difficult due to the inability to culture the microbiome in the laboratory. Our approach is to test hypotheses generated from natural system studies within model communities constructed from cultivable and genome-sequenced members of *Populus trichocarpa* genotypes, fungi and bacteria isolates. Our bacterial isolate collection consists of >2700 bacterial strains, of which ~250 have been sequenced to date through a JGI CSP. Using genomic content, strain functional data, and microbiome community data as a guide, we select communities of organisms to colonize germ-free *Populus* cuttings in constructed community experiments.

In studies of microbiome response to adverse plant growth conditions, we observe a common set of bacteria that respond to plant stress. To study the contribution of individuals to this complex phenotype, we developed two separate 10-member bacterial communities representing >90% of the estimated functional diversity present within natural microbiome metagenomes (by KEGG annotations). The 10 strains for each community represent abundant and diverse orders identified in previous studies of natural *Populus* microbiomes, including  $\alpha$ -,  $\beta$ - and  $\gamma$ - *Proteobacteria*, *Bacilli*, and *Actinobacteria*. Both communities increased root growth when inoculated on axenic host plants. Both communities were dominated by specific genera within beta- and gammaproteobacteria, despite isolation from different host species. We observe positive and negative correlations between community members in replicate samples, an example of emergent behavior that would not be observed in one-on-one studies or genera level sequencing studies. When subjected to carbon limitation stress (shading), we observe a host transcriptional response indicative of pathogen defense, suggesting that host carbon dynamics may influence microbiome structuring. This system provides a framework by which complex natural host-microbe interactions can be studied within a more simplified constructed community.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S Department of Energy, Office of Science, Biological and Environmental Research.

## Plant-Microbe Interfaces: Small-RNA (sRNA) and open reading frame (sORF) response to endo- and ecto-mycorrhizal symbioses in *Populus*

Ritesh Mewalal,<sup>1</sup> Hengfu Yin,<sup>1</sup> Rongbin Hu,<sup>1</sup> Sara Jawdy,<sup>1</sup> Patrice Vion,<sup>2</sup> François Le Tacon,<sup>2</sup> Gerald A. Tuskan,<sup>1</sup> Jessy L. Labbé,<sup>1</sup> **Xiaohan Yang**<sup>1\*</sup> (<u>yangx@ornl.gov</u>)

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA; <sup>2</sup>INRA, UMR 1136 INRA-University of Lorraine, Interactions Arbres/Microorganismes, Laboratory of Excellence ARBRE, INRA-Nancy, 54280, France

#### http://pmi.ornl.gov

Project Goals: The goal of the PMI SFA is to understand the genome-dependent molecular and cellular events involved in establishing and maintaining beneficial interactions between plants and microbes. *Populus* and its associated microbial community serves as the experimental system for understanding how these molecular events manifest themselves within the spatially, structurally, and temporally complex scales of natural systems. To achieve this goal, we focus on 1) characterizing host and environmental drivers for diversity and function in the *Populus* microbiome, 2) utilizing microbial model system studies to elucidate *Populus*-microbial interactions at the molecular level and dissecting the signals and pathways responsible for initiating and maintaining microbial relationships and 3) develop metabolic and genomic modeling of these interactions to aid in interpreting the molecular mechanisms shaping the *Populus*-microbial interface.

Mycorrhizal fungi are a diverse group of beneficial symbionts that colonize the roots of more than 90% of higher plant species. In doing so, these fungi play an important role in the maintenance of the plant health by promoting water cycling, facilitating nutrient exchange, and protection from a variety of biotic and abiotic stresses; and in exchange, the fungus receives plant-fixed carbon (Bonfante and Genre 2010; Smith and Read 2008). Mycorrhizal fungi can be classified as fungi colonizing the intercellular spaces of the roots (ectomycorrhizas, ECM) e.g., *Laccaria bicolor* or developing within the root cells (endomycorrhizas/arbuscular mycorrhiza fungi, AMF) e.g., *Rhizophagus irregularis* (formerly *Glomus intraradices*) (Bonfante and Genre 2010). The genetic contribution from the plant and fungi for the establishment and maintenance of this mutualistic symbiosis is somewhat unclear. A number of studies support the hypothesis that fungi-derived protein signals, or effectors, largely facilitate the symbiotic interaction. In this regard, the genome of *L. bicolor* revealed a large number of small secreted proteins (SSPs), many of which are expressed and accumulated in the fungal hyphae colonizing root tissue, potentially facilitating symbiosis (Martin et al. 2008).

Plant colonization by fungi is highly specific and requires that the plant distinguishes between beneficial and pathogenic fungi, and ensures that the relationship remains advantageous.

Therefore, it may be naive to believe that the plant remain a silent partner in a beneficial symbiotic interaction and this notion was challenged by several studies (el Zahar Haichar et al. 2014). In the current study, we characterized the small RNA response to AMF and ECM fungi. We performed small RNA-sequencing of roots from P. deltoides infected with R. irregularis and L. bicolor, and P. trichocarpa infected with R. irregularis and L. bicolor. Roots of mock uninfected were used as a control. We found significantly differentially expressed transcripts between the treatments and control (either upregulated or downregulated), of which some are novel transcripts aligning to non-genic regions of *P. trichocarpa*. Since sRNA can potentially encode sORF (Li et al. 2014; Ruiz-Orera et al. 2014), we used Open Reading Frame (ORF) Finder (Wheeler et al. 2003) to scan the significantly differentially expressed sRNAs for potential sORF. In total, we could predict several sORFs in the differentially expressed transcripts from the treatments in *P. deltoides* and *P. trichocarpa* respectively. Also, we analyzed the small RNA sequencing datasets to identify microRNAs (miRNAs) genes that were responsive to mycorrhiza treatments. Our analysis identified 369 putative miRNAs, of which 239 were classified as 'plant novel' miRNAs and134 showed up-/down-regulated expression in response to fungal inoculation. Furthermore, 157 miRNA-target pairs were revealed through degradome analysis.

Our results suggest that the response from the plants are in fact far more complex than previously thought and that sRNA may be an important factor underlying plant-microbe symbiotic interaction.

#### References

- Bonfante P, Genre A (2010) Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. Nature Communications. 1:48.
- el Zahar Haichar F, Santaella C, Heulin T, Achouak W (2014) Root exudates mediated interactions belowground. Soil Biology and Biochemistry. 77:69-80.
- Martin F, Aerts A, Ahrén D, Brun A, Danchin E, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature. 452:88-92.

Smith SE, Read DJ (2008) Mycorrhizal symbiosis. Academic press.

The Plant Microbe Interfaces Scientific Focus Area is sponsored by the Genomic Science Program, U.S. Department of Energy, Office of Science, Biological and Environmental Research.

## Spatially Resolved Rhizosphere Function for Elucidating Key Controls on Below-ground Nutrient Interactions

**James Moran<sup>1\*</sup>** (james.moran@pnnl.gov), Elizabeth Denis<sup>1</sup>, Peter Ilhardt<sup>1</sup>, Mary Lipton<sup>1</sup>, John Cliff<sup>1</sup>, Eirik Krogstad<sup>1</sup>, Kimbrelle Thommasson<sup>1</sup>, Erin Fuller<sup>1</sup>, Cory Overman<sup>1</sup>, and Sarah Roley<sup>2</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, Washington

<sup>2</sup>Michigan State University, W.K. Kellogg Biological Station, Hickory Corners, Michigan

http://www.pnl.gov/biology/programs/fsfa/people/moran.stm

Project Goals: The main goals of this project are to elucidate key microbiological and geochemical controls on nutrient exchange through the rhizosphere and the role that spatial organization within the root-rhizosphere-soil continuum plays in nutrient transfer. Enhanced, spatially-resolved understanding of nutrient exchange within the rhizosphere can identify key variables amenable to manipulation as part of an effective rhizosphere management program targeting enhanced plant productivity. Our aims are directed towards 1) characterizing the spatial organization of nutrient exchange between soil and plant roots, 2) identifying key microbial functions within zones of high nutrient transfer, 3)

evaluating whether directed geochemical and/or microbiological modifications can be used to stimulate nutrient exchange to foster improved plant biomass productivity.

The central hypothesis we are testing is that spatially focused regions belowground funnel a disproportionate amount of nutrients to a plant root. Further, we hypothesize that the location of these resulting nutrient exchange hotspots are not stochastically distributed throughout the rhizosphere but, rather, that they are controlled by microenvironmental conditions resulting from a combination of local microbiological communities in conjunction with host soil geochemistry. Initial experiments are using labeled  $(^{13}C)$  carbon dioxide to track photosynthetic uptake and release of resulting rhizodeposits. A series of stable isotope analyses will be used to quantify rates of organic carbon release along transects both parallel and perpendicular to the plant root axes to assess short-term spatial heterogeneity. These measurements will be performed at increasingly resolved spatial scales (Figure 2) where each set of analyses informs the spatial zone of interest for the next, more





targeted analysis; bulk analysis (mm scale) will inform community level analyses (50 µm resolution, using laser ablation isotope ratio mass spectrometry) which will inform cellular-level analyses (10's nm, using nanoSIMS). Spatially-resolved elemental analysis (using laser induced breakdown spectroscopy) of these same regions helps reveal potential correlations between

mineral or element content and localization of plant-derived organics. Proteomic analysis can provide identification of specific taxonomic groups participating in nutrient exchange (through label uptake into specific proteins trackable to the soil metagenome) and also track changes in protein synthesis associated with specific microbial functions that will accompany future elemental supplementation experiments.

Initial growth experiments are underway using controlled switchgrass mesocosms. These mesocosms are constructed with soil harvested from test fields at the Kellogg Biological Station (Hickory Corners, MI) that have been under switchgrass production for nearly a decade. In the



Figure 2: Stable isotope measurements focus on increasing spatial resolution to quantify the amount of carbon and nitrogen assimilated into the system and to enable partitioning of this uptake between the roots, rhizosphere, and soil.

event that we identify specific components of the system that correlate with increased rates of carbon exchange, we will perform supplementation experiments in similar mesocosms to assess whether small-scale nutrient supplementation can stimulate expansion of nutrient exchange hotspots and the extent to which this increases plant productivity. A better understanding of relevant subsurface nutrient exchange processes will enable future increases in plant production on less fertile soil and with fewer artificial nutrient additions as needed to promote energy and food security.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER). This contribution originates from an Early Career Research Award granted at the Pacific Northwest National Laboratory (PNNL).

#### Molecular interactions of the plant-soil-microbe continuum of bioenergy ecosystems

**Kirsten Hofmockel**<sup>1</sup>\* (kirsten.hofmockel@pnnl.gov)

<sup>1</sup>Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

Project Goals: The overall goal of this project is to test if plant-microbe interactions are limited to influencing the rate of C accrual, while mineralogy regulates the sink capacity of biofuel cropping systems. To accomplish this goal, we are (1) identifying the microbial functions and biopolymers of microbial necromass that contribute to soil C accumulation under controlled conditions, (2) characterizing microbial necromass accumulation in response to crop selection and edaphic factors in situ and (3) generating long-term, crosssite data that can be used to model C cycling in bioenergy cropping systems under different soil conditions.

Abstract text. To be competitive in the biofuel energy market, cellulosic feedstocks need to be high yielding and carbon neutral or negative. Generating sustainable feedstock production systems in marginal lands depends on the plant-microbe-soil interactions controlling the formation and stabilization of soil organic C. While some systems seem to show substantial increases in soil organic C under perennial cropping systems, others have more moderate increases that cannot be explained by relationships with climate or soil texture alone (Lemus and Lal 2005). Instead, microbial residues have come to light as a substantial source of soil organic C. Still there is limited knowledge of the complement of molecules that comprise microbially derived soil organic C. While many studies invoke microbial necromass as a key contributor to soil organic matter, only a few studies have investigated the chemical composition soil fractions by sensitive mass spectrometric methods (Golchin et al. 1997, Sleutel et al. 2011). To manage biofuel cropping systems in a sustainable manner, it is essential that we understand how plantmicrobe-soil interactions regulate the accrual and stability of soil organic matter, and in particular the relative contribution of microbial biopolymers to soil organic C pools of different stability (Ludwig et al. 2015). This Early Career research program addresses soil organic matter biogeochemistry at the molecular scale, focusing on the production and stabilization of microbial metabolites and residues in soil. Through lab and field experiments I aim to reveal the microbial mechanisms that regulate carbon (C) stabilization in soils dedicated to biofuel crops, and test identified mechanisms in ecosystem-scale field experiments. After identifying microbial signatures in lab trials, I will evaluate necromass moieties under field conditions, to advance a systems-level understanding of the plant-microbiome-soil continuum for bioenergy feedstock production.

#### References

1. Lemus, R., and R. Lal. 2005. Bioenergy crops and carbon sequestration. Critical Reviews in Plant

Sciences 24:1-21.

2. Golchin, A., J. A. Baldock, P. Clarke, T. Higashi, and J. Oades. 1997. The effects of vegetation and burning on the chemical composition of soil organic matter of a volcanic ash soil as shown by 13 C NMR spectroscopy. II. Density fractions. Geoderma 76:175-192.

3. Sleutel, S., P. Leinweber, E. Van Ranst, M. Kader, and K. Jegajeevagan. 2011. Organic matter in clay density fractions from sandy cropland soils with differing land-use history. Soil Science Society of America Journal 75:521-532.

4. Ludwig, M., J. Achtenhagen, A. Miltner, K.-U. Eckhardt, P. Leinweber, C. Emmerling, and S. Thiele-Bruhn. 2015. Microbial contribution to SOM quantity and quality in density fractions of temperate arable soils. Soil Biology and Biochemistry 81:311-322.

This research was supported by an Early Career Research Program award to K Hofmockel, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP 68292.

#### Engineering Polyketide Synthases to Generate Lightly Branched Biofuels

Constance B. Bailey,<sup>1,2</sup>\* (ConstanceBailey@lbl.gov), Amin Zargar,<sup>1,2</sup> Samuel Curran,<sup>3</sup> Rasha Anayah,<sup>3</sup> Leonard Katz,<sup>1,2,3</sup> and **Jay D. Keasling**<sup>1,2,3</sup>

<sup>1</sup>Joint BioEnergy Institute, Emeryville CA; <sup>2</sup>Lawrence Berkeley National Laboratory Division of Biological and Systems Engineering, Berkeley, CA; <sup>3</sup>University of California, Berkeley, Berkeley, CA

#### http://energy.gov/eere/bioenergy/co-optimization-fuels-engines

Project Goals: This project seeks to harness the biosynthetic capacities of polyketide synthases to generate lightly branched biofuels that will serve as excellent replacements for petrochemically-derived gasoline and diesel fuels. Selective methyl branching has a strong impact on the melting point and other physical properties of biofuel candidates, but can be challenging to install through other biosynthetic platforms, such as terpene synthases and fatty acid synthases. Thus, this work seeks to capitalize on the unique biosynthetic logic of polyketide synthases towards this end.

The biosynthetic logic of polyketide synthases (PKSs) has traditionally been harnessed to generate a diverse range of bioactive natural products. However, due to the ability to precisely tailor molecular structure through this biosynthetic logic, PKSs have great potential for synthetic biology applications in other areas, which have been largely underexplored. PKSs are particularly attractive candidates to generate fuels with optimal physical properties (such as lowered melting points) due to their ability to form selectively branched chemical architectures. This selective branching presents a stark advantage when compared to other commonly used biofuel platforms, such as fatty acid synthases. Current efforts in this area seeks to apply engineered PKSs to generate polyketide metabolites as petrochemical replacements for both compression ignition (diesel) and spark ignition (gasoline) fuels.

Specific engineering targets include a *de novo* strategy to generate a loading didomain that will prime a module in the middle of the borrelidin synthase. This borrelidin synthase has iterative activity, which is unusual for bacterial type I PKSs. Through a domain swap of mixed selectivity for branched and unbranched extender units, a series of compounds can be generated that could be excellent diesel replacement fuels or fuel-blends. To generate gasoline replacement fuels, an engineered version of the lipomycin PKS is being developed, which accepts small, lipophilic branched chain acyl-CoA starter units. The significant protein engineering efforts that have been undertaken to design these pathways have led to new fundamental insights regarding structural features and domain boundaries within PKSs.

*This work is supported by Department of Energy* BM0102060-05450-1004173/Agreement 29894/DE-AC02-05CH11231.

## Potential Mechanisms of Anaerobic Methane Oxidation and Challenges to Slow-Growing Microbial Communities

Xiaojia He<sup>1</sup>, **Christopher Kempes**<sup>2</sup>\* (<u>ckempes@gmail.com</u>), Grayson Chadwick<sup>3</sup>, Shawn McGlynn<sup>3</sup>, Yimeng Shi<sup>1</sup>, **Victoria Orphan**<sup>3</sup> and **Christof Meile**<sup>1</sup>\* (<u>cmeile@uga.edu</u>)

<sup>1</sup>University of Georgia, Athens, GA; <sup>2</sup>Santa Fe Institute, Santa Fe, NM; and <sup>3</sup>California Institute of Technology, Pasadena, CA

#### **Project Goals:**

The overarching goal to expand our understanding of the key microorganisms, metabolic strategies, and interspecies relationships involved in the formation and oxidation of methane in the environment. To study methane-cycling archaea, their syntrophic partners, and their ecophysiological properties across a range of spatial scales we will use single cell/single consortia targeted approaches combined with characterizations of field samples, geochemically-characterized laboratory microcosms, environmental bioreactors, and defined syntrophic co-cultures, and utilize novel meta'omics strategies, state-of-the-art analytical imaging methodologies and stable isotope geochemistry. Ultimately, our goal is to develop and refine predictive models of the larger scale biogeochemical processes mediated by methane-metabolizing microorganisms.

#### Abstract:

Consortia of anaerobic methanotrophic archaea and deltaproteobacteria have been shown to oxidize methane and reduce sulfate in marine sediments, thereby limiting the rise of methane into the water column and ultimately the atmosphere. However, the mechanism of the interactions between the two groups of microorganisms remains uncertain, with recent studies (e.g., McGlynn et al. 2015, Wegener et al. 2015) pointing towards direct electron transport (DIET). Here, we use reactive transport modeling to investigate potential mechanisms of methane oxidation. To assess the potential, kinetic and thermodynamic constraints on anaerobic methane oxidation, we consider (1) the exchange of electron donors such as molecular hydrogen produced by the archaea and used by the bacteria in sulfate reduction, (2) a pathway in which the archaea directly couple methane oxidation to sulfate reduction with disulfide produced as the intermediate, which then gets disproportionated to sulfide and sulfate by the bacterial partner, and (3) direct electron transport from the archaea to bacteria through nanowires. We show that reaction kinetics, transport intensities, and energetic considerations all could decisively impact the overall rate of methane consumption. Simulation results indicate that the exchange of chemical intermediates lead to oxidation rates that fall short of measured ones. In contrast,

simulations with disulfide as intermediate and through extracellular electron transfer can achieve methane oxidation rates that match or even exceed lab measured reaction rates. To distinguish between the two mechanisms, we compare the computed spatial distribution of reaction rates to the spatial patterns within the microbial consortia as measured by <sup>15</sup>N-ammonium uptake measured with FISH-nanoSIMS. These observations show that the activities are independent of the spatial proximity of archaea and their bacterial partners in consortia. This is best reproduced in the simulations of DIET, supporting the recent evidence for such a potential mechanism.

An overall challenge in modeling microbial communities is constraining the possible range of physiology. We will address this challenge using results that define the cross-species tradeoffs in bacterial cell composition, physiological function, and metabolism, spanning five orders of magnitude in cell size (Kempes et al. 2016). Within this framework it has been shown that the largest and smallest bacteria are limited by physical space and energetic requirements (Kempes et al. 2012) and we will connect these previous efforts to define the challenges facing slow-growth microbial communities.

#### References

- McGlynn, S. E., et al. (2015). Single cell activity reveals direct electron transfer in methanotrophic consortia. Nature 526 (7574): 531-535.
- 2. Wegener, G., et al. (2015). Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. Nature **526** (7574): 587-590.
- 3. Kempes et al. (2016). Evolutionary tradeoffs in cellular composition across diverse bacteria. The ISME Journal 10: 2145-2157
- 4. Kempes et al. (2012). Growth, metabolic partitioning, and the size of microorganisms. PNAS 109: 495-500

*This research is supported by the Genomic Sciences Program in the DOE Office of Science, Biological and Environmental Research DE - SC0016469* 

#### Anabolic activity in Geobacter biofilms as a function of distance to insoluble electron acceptor

Grayson L Chadwick,<sup>1\*</sup> Fernanda Jimenez<sup>2</sup>, Daniel R Bond<sup>2</sup>, and Victoria J Orphan<sup>1</sup>

<sup>1</sup>California Institute of Technology, Pasadena, California; and <sup>2</sup>University of Minnesota, Minneapolis.

## Project Goals: To investigate the spatial patterns of cellular activity during extracellular electron transfer by a pure culture grown on an electrode and applying these results to further our understanding of electron transfer during syntrophic anaerobic oxidation of methane.

Microorganisms in the environment inhabit complex ecosystems. These ecosystems are often characterized by a high degree of spatial heterogeneity in terms of chemistry, nutrient concentrations and species distribution. This heterogeneity leads to a huge diversity of ecological niches, defined not only by abiotic factors, but also a spectrum of interspecies relationships from predation and parasitism to mutualism and symbiosis. For a complete understanding of any ecological system, we need to be able to understand the way in which ecological pr ocesses vary over different spatial scales.

A central question in microbial ecology is what is the importance of spatial structure of microbial communities, particularly when large aggregations of microorganisms create and respond to complex gradients of nutrients and waste products within the community. Questions along these lines are common in almost all subfields of microbiology including medical microbiology, wastewater treatment, and environmental microbiology. Unfortunately, it is often quite difficult to assay microbial activity at single-cell resolution within environmental communities.

The application of stable isotope probing coupled to nanometer scale secondary ion mass spectrometry (nanoSIMS) provides a tool with which to assess the assimilatory activity of a wide array of microorganisms in the environment at subcellular spatial resolution. In order to better understand the connection between stable isotope assimilation and microbial activity and cell-cell interactions, we are applying this technique to controlled *Geobacter* biofilms grown in a microbial fuel cell.

*Geobacter* forms thick (10's of cell layers thick) biofilms on anodes of microbial fuel cells. The electrons produced by their catabolic oxidation of acetate are transported through the conductive biofilm to the surface of the electrode. This system serves as a model for microbe-mineral interactions, as well as a growing list of natural microbial syntrophies in nature (e.g. anaerobic oxidation of methane) that are likely dependent on such direct electron transfer processes. The benefit of analyzing *Geobacter* grown on an electrode is that the metabolism is well understood, it is genetically tractable, and models aimed at capturing the important aspects of their metabolism can be simplified to 1 dimension. Although much research has been conducted on these model microorganisms, it is still not known how their cellular activity varies with distance to the electrode surface.

Our stable isotope probing experiments and nanoSIMS analysis have demonstrated the cells in the Geobacter biofilm have a negative correlation between their cellular activity and distance to the electrode, implying that a penalty exits with increasing distance to the terminal electron accepter even though the biofilms have been show to be electrically conductive. This data, coupled with future experiments varying parameters such as electrode potential, nutrient concentrations, and genetic EET mutants will allow us to make better models for this process in *Geobacter* biofilms, and hopefully begin to extend these models to more complex associations in the environment. Of particular interest are the syntrophic communities of anaerobic methane oxidizing archaea and sulfate reducing bacteria which have been suggested to be performing extracellular electron transport.

#### Multi-scale Modeling of Circadian Rhythms: From Metabolism to Regulation and Back

Bill Cannon<sup>1,\*</sup> (bill@pnnl.gov), Jeremy Zucker<sup>1</sup>, Jennifer Hurley<sup>2</sup>, Scott Baker & Jay Dunlap<sup>3</sup>

<sup>1</sup> Pacific Northwest National Laboratory, Richland, WA, <sup>2</sup> Rensselaer Polytechnic Institute, Troy, NY, <sup>3</sup> Dartmouth College, Hanover, NH

**Goals:** The goal of this research is to develop and implement a new computational and theoretical method for modeling biological systems that fills a gap in modeling mass action dynamics. Based on statistical thermodynamics, the method bridges data-poor scales (parameters for mass action kinetics) and data-rich scales (chemical potentials of metabolites, and metabolite, protein & transcript data) to enable predictive modeling from enzymatic reactions ( $10^{-3}$  to  $10^{0}$  s<sup>-1</sup>) to gene and protein regulation (~20 minutes) to circadian rhythms (24 hours). We are:

- Implementing an approach to the law of mass action that uses chemical potentials rather than rate constants. This approach involves a rescaling of the fast degrees of freedom, resulting in a compression of the time-dependence to fewer relative scales. Steady state processes can be 'telescopically' modeled to address the scale of interest while collapsing faster scales.
- Using the new method to understand the relationship between central metabolism and circadian rhythms in *Neurospora crassa* by using a multi-scale model of metabolism that will include regulation of the circadian clock.

**Abstract**. Predictive modeling relies on solving equations in which the necessary equation parameters are either based on first principles, such as Hamiltonian systems, or on empirical data such diffusion constants or rate parameters. This requirement has hampered the predictive modeling of biological systems in that the relevant scales (e.g. those in metabolism) are too large and complex to be modeled by first principles, and the necessary rate constants are not generally available. We will address this challenge by implementing a new approach to the law of mass action that does not require rate parameters but instead uses chemical potentials (1). This new approach is possible because of advances in statistical thermodynamic methods in the last 20 years. Due to the statistical formulation of the theory, the tools are capable of direct integration of metabolomics and proteomics data. We will use these tools to fundamentally understand the relationship between metabolism (2) and molecular circadian clocks with regard to the role of the circadian clock in increasing the metabolic efficiency of the cell (3).



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

Circadian clocks lie at the epicenter of cellular physiology for both fungal and mammalian cells, both of which share clocks with equivalent regulatory architecture (4). At the core of these clocks, a heterodimeric transcription factor (TF) drives expression of genes whose protein products feed back, physically interact with, and depress the activity of their heterodimeric activator. This negative feedback loop, yielding oscillatory TF activity, is the basis of fungal and animal circadian rhythms. Output from the clock occurs when these TFs regulate genes whose products do not impact the core feedback loop. Cellular clocks in mammalian cells regulate  $\sim$ 15% of genes. In the aggregate nearly all human genes are clock-regulated in some cell type, yielding the profoundly rhythmic metabolism that is characteristic of humans and that has a major impact on both normal and disease physiology including sleep/wake cycles. Neurospora crassa is the best studied cellular circadian system and is a well-established model for eukaryotic including mammalian clocks. In addition, Neurospora is engineered to overproduce cellulases, essential components in biomass deconstruction for biofuel precursors. Neurospora provides an extremely tractable system in which to pioneer modeling of these cellular clocks and their influence on metabolism. As shown below, stages in the circadian cycle are well marked by visual observation and transcriptomics. All Neurospora genes encoding enzymes have been mapped to their corresponding steps in metabolism. In a data set well beyond anything available in any other circadian system, the entire assemblage of clock-controlled genes has been described, and data are in hand to delineate all clock-controlled proteins, including enzymes, and the clock-controlled metabolites to which they give rise (5). Each step from gene to protein to metabolite is regulated and the entire assemblage can be modeled using this unparalleled data





Heat map showing rhythmic expression of the *Neurospora* genome over 48 hrs in constant darkness.

#### References

- 1. Cannon WR & Baker SE (2016) Non-Steady State Mass Action Dynamics Without Rate Constants: Dynamics of Coupled Reactions using Chemical Potentials. *submitted*.
- 2. Dreyfuss JM, *et al.* (2013) Reconstruction and validation of a genome-scale metabolic model for the filamentous fungus Neurospora crassa using FARM. *PLoS Comput Biol* 9(7):e1003126.
- 3. Hurley JM, Loros JJ, & Dunlap JC (2016) The circadian system as an organizer of metabolism. *Fungal Genet Biol* 90:39-43.
- 4. Baker CL, Loros JJ, & Dunlap JC (2012) The circadian clock of Neurospora crassa. *FEMS Microbiol Rev* 36(1):95-110.
- 5. Hurley JM, *et al.* (2014) Analysis of clock-regulated genes in Neurospora reveals widespread posttranscriptional control of metabolic potential. *Proc Natl Acad Sci U S A* 111(48):16995-17002.

Funding statement. This project is supported by the U.S. Department of Energy's Office of Biological and Environmental Research and the National Institute of Biomedical Imaging and Bioengineering. PNNL is operated by Battelle for the U.S. Department of Energy under Contract DE-AC06-76RLO.

#### The Algal Ferredoxin Interactome

Venkataramanan Subramanian<sup>1</sup>\* (venkat.subramanian@nrel.gov), Matt Wecker<sup>2</sup>, Wei Xiong<sup>1</sup>, Alexandra Dubini<sup>3</sup>, David Gonzalez-Ballester<sup>3</sup>, Patrice Hamel<sup>4</sup>, **Maria Ghirardi<sup>1</sup>** 

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

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

**Abstract:** Ferredoxins (FDX) are small, iron-sulfur cluster-containing proteins with strong negative redox potentials (-320 to -450 mV) that mediate redox transfer in various metabolic pathways. There are 13 FDX isoforms predicted in the *Chlamydomonas* genome that are differentially expressed in response to varying environmental conditions. Previous research in our lab has established a global FDX interaction network that suggests that they may have both specific and redundant functions (Peden et al. 2013). Among other results, that work revealed that (1) FDX2 may have an overlapping role with FDX1 in donating electrons for H<sub>2</sub> production and NADP<sup>+</sup> reduction, which we confirmed in vitro (Boehm et al., 2015); (2) FDX3, together with FDX1 and FDX2 is involved in nitrogen assimilation; (3) FDX4 interacts with glycolytic enzymes and enzymes involved in protection against ROS; and (4) FDX5 may be required for hydrogenase maturation, together with FDX2 and FDX4, and has been shown to be involved in fatty acid synthesis (Yang et al. 2015).

This year, we focused our efforts on the roles of FDX1, 2 and 5. The *FDX5* knock-out mutant strain (fdx5) was characterized under different growth conditions. When compared to its wild-type (WT), CC-124, we observed decreased H<sub>2</sub> production following incubation in the dark, low light and low light but with an open reactor (the two light conditions require both photosynthesis and fermentation to occur concomitantly). The fdx5 strain was also more sensitive to H<sub>2</sub>O<sub>2</sub> than the WT as detected by inhibition of growth on H<sub>2</sub>O<sub>2</sub>-containing plates exposed to HL. Finally, in addition to its lack of growth in the dark (Yang et al. 2015), no differences were observed between WT and mutant with respect to starch accumulation, lipid and organic acid production in any of the above studied conditions. On the other hand, fdx5 showed less starch and low fermentative metabolite accumulation (including H<sub>2</sub>) levels under sulfur (S) deprived conditions. Studies comparing the levels of the various FDXs in this mutant are being conducted.

FDX1 and FDX2 knock-down mutants were generated in the *Chlamydomonas* 704 (nit<sup>+</sup>) strain using micro-RNA-mediated silencing techniques. Analyses of the *fdx1* mutants revealed a strain with a 55% drop in *fdx1* transcript levels, which was accompanied by an *increase* in growth when cultivated photoautotrophically or photoheterotrophically in liquid medium at non-saturating light intensity. The photosynthetic capacity of the mutant (under saturating light) was

12% lower than that of WT, and its respiratory capacity 80% of WT. However, the mutant's initial rates of  $H_2$  photoproduction measured under saturating light were over 2X higher than its WT counterpart. Maximum production rates were achieved after 2 hours of anaerobic induction for both strains. Inhibitor experiments and starch level evaluations are underway to examine the source of reductant to the hydrogenase under various illumination conditions.

Under S-deprivation, various changes between WT and fdx l were observed: the O<sub>2</sub> concentration in the gas phase was still high after 48 hours of S-deprivation in both strains (although even higher in the fdx l mutant vs WT) and did not decrease substantially during the remainder of the treatment; H<sub>2</sub> detection was delayed by 24 h in the fdx l mutant, but the levels of H<sub>2</sub> accumulated were 3.5-fold higher after 96 hours of –S treatment when compared to the WT. Among the fermentative products, fdx l showed large amounts of lactate accumulation in the extracellular medium, while formate secretion was lowed compared to the WT. Clearly, there is a shift in the nature of the fermentation products as well as on the time-point at which each one is detected in the mutant vs. WT strain. We are currently determining starch levels and changes in the expression of all ferredoxins to verify if there is a complementary up-regulation of any of the minor ferredoxins (particularly FDX2) in fdx l.

The FDX2 knock-down mutant, which expresses only about 20% of the *FDX2* gene, shows lower nitrite assimilation activity, as expected, with nitrite secretion observed when the culture is grown in the presence of nitrate. The mutant produces less  $H_2$  than WT when grown under TAP-NH<sub>4</sub>, S-deprivation, and in either low light or dark conditions. These results confirm the involvement of FDX2 in nitrogen assimilation and suggest its role in  $H_2$  production (fermentative and photoproduction) under stress conditions. As in the case of the *fdx1* mutant, transcript and protein levels of all the FDX isoforms will be determined in this mutant.

Finally, double fdx 2/fdx 5 mutants, as well as overexpressing strains of FDX2, FDX5 and both (in a WT background and subsequently in in the fdx 1 mutant) are being constructed.

#### References and publications generated by this project and reference above:

- 1. Boehm M, Alahuhta M, Mulder DW, Peden EA, Long H, Brunecky R, Lunin VV, King PW, Ghirardi ML, Dubini A. (2015). Photosynth. Res. DOI 10.1007/s11120-015-0198-6.
- Peden EA, Boehm M, Mulder DW, Davis R, Old WM, King PW, Ghirardi ML, Dubini A. (2013). J. Biol. Chem. 288(49): 35192-209.
- Yang W, Wittkopp TM, Li X, Warakanont J, Dubini A, Catalanotti C, Kim RG, Nowack EC, Mackinder LC, Aksoy M, Page MD, D'Adamo S, Saroussi S, Heinnickel M, Johnson X, Richaud P, Alric J, Boehm M, Jonikas MC, Benning C, Merchant SS, Posewitz MC, Grossman AR. (2015). Proc. Natl. Acad. Sci. U S A. 112(48):14978-83.
- 4. **Subramanian, V, Dubini, A and Ghirardi, ML** (submitted review). In "Microbial Hydrogen Production: Achievements and Perspectives" (Eds. Seibert M. and Torzillo, G.) Springer; *submitted*.

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

#### Design, Synthesis, and Testing Toward a 57-Codon Genome

Nili Ostrov,<sup>1</sup> Matthieu Landon,<sup>1,2,3</sup> Marc Guell,<sup>1,4</sup> Gleb Kuznetsov,<sup>1,5</sup> Jun Teramoto,<sup>1,6</sup> Natalie Cervantes,<sup>1</sup> Minerva Zhou,<sup>7</sup> Kerry Singh,<sup>7</sup> Michael G. Napolitano,<sup>1,8</sup> Mark Moosburner,<sup>1</sup> Ellen Shrock,<sup>1</sup> Benjamin W. Pruitt,<sup>4</sup> Nicholas Conway,<sup>4</sup> Daniel B. Goodman,<sup>1,4</sup> Cameron L. Gardner,<sup>1</sup> Gary Tyree,<sup>1</sup> Alexandra Gonzales,<sup>1</sup> Barry L. Wanner,<sup>1,9</sup> Julie E. Norville,<sup>1</sup> Marc J. Lajoie,<sup>10</sup> **George M. Church**<sup>1,4</sup>\* (church@genetics.med.harvard.edu)

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, MA; <sup>2</sup>Program in Systems Biology, Harvard University, Cambridge, MA; <sup>3</sup>Ecole des Mines de Paris, Mines Paristech, Paris 75272, France; <sup>4</sup>Wyss Institute for Biologically Inspired Engineering, Boston; <sup>5</sup>Program in Biophysics, Harvard University, Boston; <sup>6</sup>Department of Biological Sciences, Purdue University, West Lafayette, IN; <sup>7</sup>Department of Biological Engineering, MIT, Cambridge, MA; <sup>8</sup>Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA; <sup>9</sup>Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA.; <sup>10</sup>University of Washington, Seattle, WA

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

#### **Project Goals:**

We report progress towards assembly of a 3.97 MB, 57-codon *Escherichia coli* genome in which seven codons were replaced with synonymous alternatives across all protein coding genes. Target codons AGA (Arg), AGG (Arg), AGC (Ser), AGU (Ser), UUG (Leu), UUA (Leu) and UAG (Stop) were chosen by in-house designed software, guided by experimentally validated synonymous recoding rules<sup>1,2</sup>. Thus far, we have completed computational design and de novo synthesis of the recoded genome in 87 ~50kb segments, as well as functional testing of 85% of all recoded genes. We developed a robust pipeline for design, assembly, testing and troubleshooting of recoded segments, with only 25 of 3031 tested genes showing lethal effects. Here, we present our pipeline for the final genome construction step, combining multiple recoded segments into a single strain in a hierarchical fashion. This work underscores the feasibility of rewriting genomes and establishes a framework for assembly of synthetic organisms.

#### **Publications**

- 1. Ostrov et al. (2016) Science 353(6301). doi: 10.1126/science.aaf3639.
- 2. Napolitano et al. (2016) PNAS 113(38). doi: 10.1073/pnas.1605856113

Funding for this work was provided by U.S Department of Energy grant DE-FG02-02ER63445 and Defense Advanced Research Projects Agency grant BAA-12-64. B.L.W. was supported by NSF grant 106394; G.K. was supported by DOD NDSEG Fellowship; D.B.G. was supported by NSF Graduate Research Fellowship; E.S. was supported by the Origins of Life Initiative at Harvard University.

#### **Exploring Species Specificity of Lambda Red Recombination**

Gabriel Filsinger<sup>1\*</sup> (<u>filsinger@g.harvard.edu</u>), Tim Wannier,<sup>1</sup> Xavier Rios,<sup>1</sup> Chris Gregg,<sup>1</sup> Marc Lajoie,<sup>2</sup> George M. Church<sup>1</sup>

<sup>1</sup>Harvard Medical School, Department of Genetics, Boston, MA; <sup>2</sup>University of Washington, Seattle, WA

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

## Project Goals: To extend recombineering methods to organisms other than *E. coli* by identifying and overcoming the host-specificity of Beta recombinases.

Recombination using the lambda red protein Beta has been extensively used in *E. coli* to incorporate insertions, deletions, and point mutations into the genome at chosen loci. Although Beta increases the rate of ssDNA recombination in *E. coli* 10,000 fold, this catalytic recombination activity has not been observed in other bacteria such as *Lactobacillus* or *Corynebacterium*. Interestingly, the family of single-stranded annealing proteins that includes Beta has been found to have species-specific activities and unpredictable efficiencies, making it difficult to design a generalized method for porting Beta recombination across organisms. In order to explore mechanisms of generalizing the recombinate activity, we use *Lactobacillus* and *E. coli* as two model organisms with orthogonal recombination systems to probe the source of species specificity.

*This project has been funded by DOE grant DE-FG02-02ER63445 and a NSF Graduate Research Fellowship.* 

#### Non-Standard Amino Acid Incorporation in the New Era of Recoded Genomes

Aditya M. Kunjapur<sup>1\*</sup> (<u>aditya\_kunjapur@hms.harvard.edu</u>), Erkin Kuru,<sup>1</sup> Devon Stork,<sup>1</sup> Kamesh Narasimhan,<sup>1</sup> Po-yi Huang,<sup>1</sup> Daniel J. Mandell,<sup>1</sup> John Aach,<sup>1</sup>, and **George M. Church<sup>1</sup>** 

<sup>1</sup>Harvard Medical School, Department of Genetics, Boston, Massachusetts, USA

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

Project Goals: The advent of scalable genome engineering and assembly technologies has inspired efforts aimed at removing all instances of particular codons by synonymous substitution throughout a genome. In the resulting recoded organisms, codons are free for reassignment to amino acids beyond the canonical twenty without competition from endogenous tRNAs or release factors. This increases the ease of producing proteins containing non-standard amino acids (nsAAs), which have previously been shown to have numerous desirable properties. Here, we review our recent efforts to use nsAAs for new functions, such as for biocontainment. Moving forward, we are interested in improving the activity and selectivity of amino-acyl tRNA synthetases (AARSs) and other components of translational machinery for incorporation of D-amino acids and for novel nsAAs.

During the last decade, the Church Lab has been at the forefront of developing genome engineering technologies for recoding. The development of multiplex automatable genome engineering (MAGE) provided a foundational approach to recoding at the scale of hundreds of codons (1). Conjugative assembly genome engineering (CAGE) enabled the parallelization of recoding efforts within portions of an individual genome that could later be hierarchically reassembled (2). A major product of these efforts was the creation of an *Escherichia coli* strain devoid of amber (UAG) codons and release factor 1, known as C321. $\Delta$ A (3). Since then, MAGE and CAGE have also been used to construct an *E. coli* strain without the rare arginine codons AGA and AGG in any essential genes (4). Moving forward, genome synthesis and assembly are being used to construct a 57-codon strain, which is nicely reviewed in another poster (5).

Non-standard amino acids (nsAAs) can site-specifically be incorporated into proteins *in vivo* across prokaryotes and eukaryotes using orthogonal pairs of amino acyl tRNA synthetases (AARSs) and orthogonal tRNA (6–9). Some of the many functions that nsAAs have been shown to provide include photocrosslinking (10, 11), functionalization (12), structure determination (13), fluorescence (14, 15), and metal binding (16). Despite numerous demonstrations of nsAA potential, nsAAs use has been limited by the absence of a dedicated codon that is free of competition from endogenous translation machinery. In addition, we recently used the engineered C321. $\Delta$ A strain to demonstrate a new use for nsAAs: biocontainment dependent on the nsAA biphenylalanine (bipA) (17). Our ability to biocontain is especially important given that recoding can bestow an organism with increased virus resistance (3).

As we work toward a 57-codon genome that will include a dedicated codon for biocontainment,

we intend to first improve the activity and selectivity of BipARS, which is the AARS for bipA. Selectivity at the AARS level will be vital to ensuring the fidelity of translation during the envisioned simultaneous use of multiple types of nsAAs, and we are interested in ways to distinguish nsAAs at the molecular level *in vivo*. In addition, we are looking towards other nsAAs with interests in fluorogenic and D-amino acids (D-AAs). D-AA-containing proteins have potential to greatly expand protein conformation space (*18*) and thereby enable novel protein functions, and can also enhance protein thermostability (*19*). D-AAs are discriminated from L-AAs at 3 steps in core translation machinery: aminoacylation of tRNAs by aminoacyl-tRNA synthetases (aaRSes), formation of ternary complexes with EF-Tu-GTP and their delivery to the ribosome, and the ribosome's own catalysis of peptide bond formation.

#### References

- H. H. Wang *et al.*, Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. 460, 894–898 (2009).
- 2. F. J. Isaacs *et al.*, Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science (80-. ).* **333**, 348–353 (2011).
- 3. M. J. Lajoie *et al.*, Genomically Recoded Organisms Expand Biological Functions. *Science (80-. ).* **342**, 357–360 (2013).
- 4. M. G. Napolitano *et al.*, Emergent rules for codon choice elucidated by editing rare arginine codons in Escherichia coli. *Proc. Natl. Acad. Sci.* **113**, E5588–E5597 (2016).
- 5. N. Ostrov *et al.*, Design, synthesis, and testing toward a 57-codon genome. *Science (80-. ).* **353**, 819–822 (2016).
- 6. L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the genetic code of Escherichia coli. *Science* (80-.). **292**, 498–500 (2001).
- 7. J. W. Chin et al., An expanded eukaryotic genetic code. Science (80-. ). 301, 964–967 (2003).
- L. Wang, J. Xie, P. G. Schultz, EXPANDING THE GENETIC CODE. Annu. Rev. Biophys. Biomol. Struct. 35, 225–249 (2006).
- J. W. Chin, Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* 83, 379–408 (2014).
- 10. J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci.* **99**, 11020–11024 (2002).
- 11. J. W. Chin *et al.*, Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia coli. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).
- 12. L. Wang, Z. Zhang, A. Brock, P. G. Schultz, Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 56–61 (2003).
- 13. J. Xie *et al.*, The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat Biotechnol.* **22**, 1297–1301 (2004).
- 14. D. Summerer *et al.*, A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9785–9 (2006).
- 15. J. Wang, J. Xie, P. G. Schultz, A Genetically Encoded Fluorescent Amino Acid. J. Am. Chem. Soc. 128, 8738–8739 (2006).
- 16. J. Xie, W. Liu, P. G. Schultz, A Genetically Encoded Bidentate, Metal-Binding Amino Acid. *Angew. Chemie.* **119**, 9399–9402 (2007).
- D. J. Mandell *et al.*, Biocontainment of genetically modified organisms by synthetic protein design. *Nature*. 518, 55–60 (2015).
- 18. S. Rana *et al.*, A double catgrip mixed L and D mini protein only 20 residues long. *Bioorg. Med. Chem.* 15, 3874-82 (2007).
- 19. A. Rodriguez-Granillo *et al.*, Computational design of thermostabilizing D-amino acid substitutions. *J. Am. Chem. Soc.* **133**, 18750-9 (2011).

This project has been graciously funded by DOE grant DE-FG02-02ER63445.

## **Retron Library Recombineering: Construction and Functional Interrogation of Trackable Genomic Edits**

Max G. Schubert<sup>1,2</sup>\* (<u>mschubert@g.harvard.edu</u>), Daniel Goodman,<sup>2,3</sup> George M. Church<sup>1,2,3</sup>

<sup>1</sup>Harvard University, Cambridge, MA; <sup>2</sup>Wyss Institute of Harvard, Boston, MA; <sup>3</sup>Massachussetts Institute of Technology, Cambridge, MA

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

Project Goals: Develop methods for synthesis of specific DNAs *in vivo* using Retron constructs. Use the DNAs produced to conduct efficient, multiplexed genome-engineering, and develop a tracking methodology wherein mutants are uniquely barcoded for phenotype measurement.

Advances in DNA synthesis and DNA sequencing enable powerful new systematic surveys of genomes, but precise genome editing has yet to be applied to genome-wide studies. Transposoninsertion sequencing and CRISPR knockout screens both efficiently alter genomes, and rapidly evaluate thousands of mutants as a pooled library using Next-Generation Sequencing (NGS)<sup>1</sup>. However these methods non-specifically ablate a locus using marker or scar sequences, and cannot create specific edits/alleles of interest. Here we present Retron Library Recombineering, a method that efficiently makes precise edits across the E. coli genome, and enables NGS tracking of the resulting edited cells. Single-stranded DNA (ssDNA) are generated in vivo using natural retro-elements called retrons, and these ssDNA serve as donors for lambda-RED recombineering<sup>2</sup>. This technique is more efficient at producing desired edits than existing methods, but more importantly, it enables the abundance of all mutant cells in a library to be measured by NGS of the cassette. We demonstrate efficient genome editing and accurate fitness measurement using this technique, and explore this technique for "de-bugging" the construction of genetically recoded organisms, and "de-convoluting" large sets of mutations observed in evolution experiments. Thus, the fitness contribution of thousands of individual mutations, and their epistatic relationships, can be defined all in one growth vessel. Unlike current highthroughput genome-wide screens, this method can be used to explore natural sequence variation or new genotypes of interest. Retrons are a simple, potent mechanism for producing ssDNA in vivo and will have diverse applications for precise genome engineering.

#### References

1. A. N. Gray *et al.*, High-throughput bacterial functional genomics in the sequencing era. *Curr. Opin. Microbiol.* **27**, 86–95 (2015).

2. F. Farzadfard, T. K. Lu, Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science*. **346**, 1256272 (2014).

This project has been funded by DOE grant DE-FG02-02ER63445.

#### <sup>13</sup>C-assisted flux elucidation using genome-scale carbon mapping models

#### Saratram Gopalakrishnan<sup>1\*</sup> and Costas D. Maranas<sup>1</sup>

<sup>1</sup>Department of Chemical Engineering, The Pennsylvania State University, University Park, PA

Project Goals: This project aims to expand existing carbon mapping models from core models to genome-scale models for flux elucidation using 13C-MFA using tools, algorithms and reaction information contained within the MetRxn database. The constructed mapping models will be deployed for genome-scale flux elucidation to obtain insights into the impact of model scale-up and loss of information, sensitivity of flux distributions to biomass composition, and novel carbon backbone scrambling patterns and pathway usage in cyanobacteria.

Metabolic models used in <sup>13</sup>C metabolic flux analysis generally include a limited number of reactions primarily from central metabolism. They typically omit degradation pathways, complete cofactor balances, and atom transition contributions for reactions outside central metabolism. This study addresses the impact on prediction fidelity of scaling-up bacterial and cyanobacterial mapping models to a genome-scale. The core mapping model for E. coli employed in this study accounts for 75 reactions and 65 metabolites primarily from central metabolism. The genomescale metabolic mapping model (GSMM) (imEco726, 668 reaction and 566 metabolites) is constructed using as a basis the *i*AF1260 model upon eliminating reactions guaranteed not to carry flux based on growth and fermentation data for a minimal glucose growth medium. This GSMM model identifies all peripheral metabolic pathways contributing to small metabolite recycling, alternate routes to lower glycolysis and the TCA cycle, multiple transhydrogenase mechanisms. The cyanobacterial mapping models imSyn711 and imSyf608 detail reaction atom mapping for all carbon-balanced reactions in cyanobacteria Synechocystis PCC 6803 and Synechococcus PCC 7942. 221 unique cyanobacterial reactions contribute to 67 novel carbon paths identified using an EMU-based depth-first search algorithm spanning Calvin cycle, photorespiration, an expanded glyoxylate metabolism, and corrinoid biosynthetic pathways. Differences in metabolism around the 3-phosphoglycerate (3PG) metabolic node and glyceraldehyde-3-phosphate (G3P) are flagged as the primary source of carbon scrambling patterns unique to the cyanobacteria with acetate recycling from porphyrin biosynthesis as a secondary source. The interplay between enzymes of the non-oxidative pentose phosphate pathway and carbon fixation via the Calvin cycle, results in rearrangements of C1 and C2 carbons of the pentose phosphates allowing for the transfer of fixed <sup>13</sup>C-CO<sub>2</sub> to all atoms of G3P and 3PG metabolites. In contrast, in *E. coli* carbon scrambling occurs only in upper glycolysis, pentose phosphate pathway, and the Entner-Doudoroff pathways. Furthermore, cyanobacterial mapping models reveal the presence of an additional <sup>13</sup>C incorporation path via glyoxylate metabolism and photorespiration contributing to three unique carbon arrangement patterns on triose phosphates, absent in E. coli. Labeling data for 17 amino acid fragments obtained from cells fed with glucose labeled at the second carbon was used to obtain fluxes and ranges for E. coli, whereas labeling distributions of 15 central metabolites obtained via <sup>13</sup>C-labeled bicarbonate are used for flux and range elucidation in *Svnechocvstis*. Metabolic fluxes and confidence intervals are estimated, for both core and genome-scale mapping models, by minimizing the sum of square of differences between predicted and experimentally measured isotopic steady-state labeling distributions for E. coli, and isotope labeling dynamics for Synechocystis.

Overall, we find that both the topology and estimated values of the metabolic fluxes remain largely consistent between the core and GSMM models for *E. coli*. Stepping up to a genome-scale mapping model leads to wider flux inference ranges for 20 key reactions present in the core model. The glycolysis flux range doubles due to the possibility of active gluconeogenesis, the TCA flux range expanded by 80% due to the availability of a bypass through arginine consistent with labeling data, and the transhydrogenase reaction flux was essentially unresolved due to the presence of as many as five routes for the inter-conversion of NADPH to NADH afforded by the genome-scale model. By globally accounting for ATP demands in the GSMM model the unused ATP decreased drastically with the lower bound matching the maintenance ATP requirement. A non-zero flux for the arginine degradation pathway was identified to meet biomass precursor demands as detailed in the iAF1260 model. Inferred ranges for 81% of the reactions in the genomescale metabolic (GSM) model varied less than one-tenth of the basis glucose uptake rate (95% confidence test). This is because as many as 411 reactions in the GSM are growth coupled meaning that the single measurement of biomass formation rate locks the reaction flux values. This implies that accurate biomass formation rate and composition are critical for resolving metabolic fluxes away from central metabolism and suggests the importance of biomass composition (re)assessment under different genetic and environmental backgrounds. In addition to better recapitulation of experimentally observed labeling distributions of all measured central metabolites, flux elucidation using the cyanobacterial mapping model predicts the existence of a serine biosynthesis route from 3PG and trace flux through the GABA shunt. The loss of information associated with mapping fluxes from MFA on a core model to a GSM model is quantified and its implications on inferences drawn on the metabolic capabilities of E. coli and cyanobacteria are discussed.

Supported by funding from the U.S. Department of Energy by grants DE-FG02-05ER25684 and DE-SC0012722.

Application of the MetRxn database to highlight multi-tissue/organisms and expansion to include algorithms for predicting novel reactions and pathways

Chiam Yu Ng<sup>1</sup>\* (cun121@psu.edu), Lin Wang<sup>1</sup>, Akhil Kumar<sup>2</sup>, Siu Hung Joshua Chan<sup>1</sup>, Margaret Simons<sup>1</sup>, Anupam Chowdhury<sup>1</sup> and **Costas D. Maranas<sup>1</sup>** 

<sup>1</sup>Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; <sup>2</sup>The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA

http://www.maranasgroup.com/

Project Goals: The project aims to apply and extend the MetRxn database. MetRxn was used to create expanded genome-scale models that include complex interactions through multiple tissues and organisms by examining the metabolism in maize and microbial communities. Additionally, MetRxn is being expanded to include algorithms for predicting novel reactions and pathways by examining promiscuous enzymes and designing novel pathways accounting for thermodynamic feasibility and protein costs.

The developed MetRxn knowledgebase (<u>www.metrxn.che.psu.edu</u>) is a unified repository of 8 databases and 112 metabolic models, resulting in 44,784 unique reactions and a million plus unique metabolites. In addition, the database includes 6,211 reaction rules developed using Canonical Labelling for Clique Approximation (CLCA), which leverages prime factorization. We focus on applying the MetRxn database to expand models to investigate the metabolic behaviors of complex organisms and microbial communities. By reconstructing multi-tissue and multiorganism models, we can use flux balance analysis to determine the interactions between different cell/tissue-types or organisms, examine metabolic changes among growth stages, and predict the community abundance of microorganisms. Additionally, algorithms are being developed to utilize the standardized database and reaction rules in MetRxn predicting novel reactions from enzyme promiscuity and novel pathway combinations accounting for thermodynamic feasibility and protein costs.

The MetRxn database was applied to highlight the interactions between multiple tissues and organisms. First, a whole-plant metabolic model of maize was developed by reconstructing the root, stalk, leaf, kernel and tassel tissues using the phloem to transport metabolites. The whole-plant model was simulated at three growth conditions: vegetative leaf growth, tassel development, and kernel filling. Growth rate and dry weight proportions were used to define the plant biomass congruity and normalize the transport reaction flux between tissues. Using parsimonious analysis, the total number of reactions required during each growth stage was identified with approximately 80% of the reactions in the vegetative leaf growth stage common among all growth stages. The model predicts the expected transport of sucrose from photosynthetic tissues into the root, where sucrose is consumed for energy and produces organic acids that are recycled to the photosynthetic tissues. Additionally, reactions from MetRxn were combined for multiple microbial organisms to create a community model. Steady-state community metabolism was predicted for long-term stability using a novel framework, SteadyCom. In the absence of the steady-state constraint, faster growing organisms will displace other microbes and disrupt the interactions predicted by non-steady-state methods. The SteadyCom framework is scalable to large communities and compatible with existing constraintbased modeling techniques. The algorithm's capability to predict relative abundances and interorganism relationships was demonstrated in a community of four E. coli auxotrophic mutants.

The standardized MetRxn reaction database and reaction rules within the newly developed quasiPath framework were leveraged to design novel reaction pathways in a mass balanced fashion. While designing novel biotransformations from the substrate to product, we

consider design elements such as network size, non-linear pathway topology, mass-conservation, cofactor balance, thermodynamic feasibility and chassis selection. We demonstrate the capability of quasiPath in imputing the biochemical gaps in gut metabolism. In addition, biotransformation models for the aerobic conversion of polycylic aromatic hydrocarbons (PAH), found in industrial effluents to commercially valuable compounds are also presented. We also contrast the philosophies of quasiPath with recent retrosynthesis efforts.

Factors such as stoichiometric and cofactor balancing, thermodynamic feasibility and pathway metabolic burden have to be carefully considered while designing a novel metabolic pathway from the collection of reactions in MetRxn. We have previously developed the MILPbased optStoic/minFlux pathway designing procedure which identifies a minimal network of flux carrying reactions that satisfy the overall design equation while obeying mass, energy and overall thermodynamic balances. In this project, we extended the procedure to explore the design principle of the natural glycolytic pathways. Although there are a plethora of routes through which glucose can be converted into pyruvate, the canonical ED and EMP glycolytic pathways are widely adopted by different species despite their difference in energy efficiency. We thus employed the modified optStoic/minFlux method to prospect for over 37,301 possible routes between glucose and pyruvate at different pre-specified stoichiometric yields of ATP. Subsequently, we filtered all candidate pathways based on their thermodynamic feasibility and quantified the minimal protein cost of the feasible pathways. A trade-off plot between energy efficiency and protein cost for each of the feasible routes revealed that the naturally evolved ED and EMP pathways are among the most protein cost-efficient pathways in their respective ATP yield categories. The pathway design and analysis procedure developed here can be applied to other important bioconversion pathways.

Supported by funding from the U.S. Department of Energy by grant DE-FG02-05ER25684 and DE-SC0008091.

#### **RDP: Data and Tools for Microbial Community Analysis**

Benli Chai<sup>1</sup>\* (chaibenl@msu.edu), Santosh Gunturu,<sup>1</sup> Leo Tift,<sup>1</sup> Yanni Sun,<sup>1</sup> C. Titus Brown,<sup>2</sup> James Tiedje,<sup>1</sup> **James Cole**<sup>1</sup>

<sup>1</sup>Michigan State University, East Lansing, Michigan 48824; <sup>2</sup>University of California-Davis, California 95616

#### **Project Goals:**

RDP offers aligned and annotated rRNA and important ecofunctional gene sequences with related analysis services to the research community. These services help researchers with the discovery and characterization of microbes important to bioenergy production, biogeochemical cycles, climate change, greenhouse gas production, and environmental bioremediation.

http://rdp.cme.msu.edu http://fungene.cme.msu.edu https://github.com/rdpstaff

RDP's data collections include 3,356,809 16S rRNA and 125,525 fungal 28S rRNA sequences as of December 2016. Over the past year, RDP websites (Cole et al., 2014) were visited, on average, by 8412 researchers (unique IPs) in 16,639 analysis sessions each month.

During 2016, we updated the RDP Classifier (Wang et al., 2007) and the underlying RDP Taxonomy two times to reflect recently discovered bacterial, archaeal, and fungal lineages and latest taxonomic emendations. The RDP Taxonomy now models over 2500 bacteria and archaea genera (including about 100 unofficial genera), with over 13,000 training sequences. RDP Classifier has updated its "Warcup" training set (Deshpande et al., 2016) with the latest, much improved version for rapid fungal classification using Internal Transcribed Spacer (ITS) sequences. Besides the "Warcup set", RDP Classifier also offers UNITE set trained fungal classification. Most RDP tools are now available as open source command-line versions through RDP's GitHub repository (https://github.com/rdpstaff). This includes our recently published Xander software package (Wang et al., 2015). Xander incorporates our novel method for assembling protein-coding sequences for genes of interest from a metagenomic dataset. In addition to the software packages, the repository includes additional resources including examples, documentation and tutorials. These command- line tools provide researchers with an independent method to analyze their own data, including high-throughput data and many of these tools are already used in third-party pipelines. These stand-alone versions of our tools have been created for easy porting to KBase in the future.

RDP's FunGene Pipeline & Repository (Fish et al., 2013) provides databases for 270 protein coding genes useful as phylogenetic markers and for following important ecological functions. In addition to the aligned and annotated gene and protein sequences, FunGene provides online analysis functions and tools for selecting subsets of sequences for download and further analysis. Use of the FunGene web, on average, was 778 researchers per month in 1402

analysis sessions. During the past year, we updated FunGene data releases five times from searches of the primary sequence databases.

We have continued improving the website for better performance in accessing reference sequences and analyzing amplicon data. In addition to optimizing existing gene models in N and C cycles, we have added more genes of environmental importance, such as an additional N cycle gene model for improved homolog detection and characterization, five new models for major classes of bacterial phytase genes (HAPs, PTPs, PAPs, and BPPs) and an alkaline phosphatase gene (ALP), which play an important role in general phosphorus cycle in different environments in catalyzing the cleavage of phosphate groups from the indigestible organic forms and make it bioavailable.

#### **References:**

Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske, and J. M. Tiedje. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42 (Database issue): D633-D642. doi: 10.1093/nar/gkt1244

Fish, J. A., B. Chai, Q. Wang, Y. Sun, C. T. Brown, J. M. Tiedje, and J. R. Cole. (2013). FunGene: the functional gene pipeline and repository. *Front Microbiol.* 4: 291. doi: 10.3389/fmicb.2013.00291

Wang, Q., J. A. Fish, M. Gilman, Y. Sun, C. T. Brown, J. M. Tiedje, and J. R. Cole. (2015). Xander: employing a novel method for efficient gene-targeted metagenomic assembly. *Microbiome* 3: 32 doi: 10.1186/s40168-015-0093-6

Wang, Q, G. M. Garrity, J. M. Tiedje, and J. R. Cole. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol.* 73: 5261-5267. doi: 10.1128/AEM.00062-07

Deshpande, V., Q. Wang, P. Greenfield, M. Charleston, A. Porras-Alfaro, C. R. Kuske, J. R. Cole, D. J. Midgley, and N. Tran-Dinh. (2016). Fungal identification using a Bayesian Classifier and the 'Warcup' training set of Internal Transcribed Spacer sequences. *Mycologia* 108(1):1-5. doi: 10.3852/14-293

This research was supported by the Office of Science (BER), U.S. Department of Energy Grant No. DE-FG02-99ER62848, with contributions from Office of Science (BER), U.S. Department of Energy Grant Nos. DE-SC0014108, DE-SC0010715, DE-FC02- 07ER64494, NIEHS Superfund Research Program Award 5P42ES004911-23 and National Science Foundation Award No. DBI-1356380.

## Is there a direct link between free-living nitrogen fixation rates and nitrogen mineralization rates?

Darian Smercina,<sup>1</sup> Lisa K. Tiemann<sup>1</sup>\* (ltiemann@msu.edu), Sarah Evans<sup>1,2</sup>, Maren Friesen<sup>1</sup>

<sup>1</sup>Michigan State University, East Lansing, MI; <sup>2</sup>Kellogg Biological Station, Hickory Corners, MI

http://rhizosphere.msu.edu

**Project Goals:** The primary goals of this project are to increase our understanding of how rhizosphere microbiomes in perennial biofuel cropping systems (PBCS) use C resources, especially root exudates, to fuel N-transformations. We are particularly interested in trade-offs between free-living N-fixation and N-mineralization and how root exudates support both processes.

Representing the main pathways to plant accessible nitrogen (N), the relationship between freeliving N fixation (N-fix) and N mineralization (N-min) is relatively unexplored. In particular, free-living N-fix, thought to be supported mainly by plant root exudates, has often been overlooked. Currently, there is no information on the link between N-fix and N-min rates available in the literature. In order to address this knowledge gap, we are using a three-pronged approach, including a meta-analysis, a greenhouse study and field experiments.

Following an extensive literature search, we found 12 papers that simultaneously reported N-fix and N-min rates. Surprisingly, these data indicated a positive relationship between N-fix and Nmin rates; however, the scarcity of data limits our ability to draw any strong conclusions. We have explored the relationship between N-fix and N-min in a controlled greenhouse experiment using switchgrass (Panicum virgatum) because recent evidence suggests switchgrass may support free-living N-fix when N limited. Indeed, in our study switchgrass and soils exposed to N limiting conditions experienced no adverse effects, namely no differences in plant growth or tissue chemistry (C:N) or soil enzyme activities compared to non-N limiting conditions. Soils used in this study are from marginal lands, low in soil organic matter and N, so it is likely N deficits are compensated for via N-fix. Analysis of 15N2 -Fix and gross N-min rates, determined via 15N pool dilution, will elucidate this source of N. Finally, our field experiment encompasses six marginal land sites across MI and WI, part of the Great Lakes Bioenergy Research Center. In 2016, we measured N-fix and N-min rates in switchgrass monoculture plots at all six sites once, at the peak of growing season, and bi-weekly, from April to September, at two MI field sites. Data collected to date from two MI sites show no difference in N-min rates in N fertilized versus unfertilized switchgrass plots. We anticipate greater N-fix at these sites in the unfertilized relative to the N fertilized plots. Overall, evidence suggests free-living N-fix is an important source of N, which is often overlooked despite its potentially large impacts on plant productivity and soil microbial activities such as enzyme production and N-min.

This project, entitled "Connecting nitrogen transformations mediated by the rhizosphere microbiome to perennial cropping system productivity in marginal lands" is supported by the Office of Biological and Environmental Research in the DOE Office of Science under Award Number DE-FOA-0001207.
# The Effects of Growth Promoting Rhizobacteria and Endophytes on Switchgrass Growth and Root Architecture

Bana Abolibdeh,<sup>1</sup>\* (libdehba@msu.edu), Laney Hult,<sup>1</sup>Katheryne Johnston,<sup>1</sup>Maren Friesen<sup>1</sup>, Sarah Evans<sup>1,2</sup> and Lisa Tiemann<sup>1</sup>

<sup>1</sup>Michigan State University, East Lansing; <sup>2</sup>Kellogg Biological Station, Hickory Corners, MI

http://rhizosphere.msu.edu

**Project Goals:** The primary goals of this projectare to increase are understanding of how rhizosphere microbiomes in perennial biofuelcropping systems (PBCS) gain C resources to fuel N-transformations and interact with plant root exudation patterns and physiological pathways.

In this study, we aim to understand interactions between plant growth promoting bacteria in the switchgrass rhizosphere and endosphere. Various populations of microorganisms inhabiting the same environment often compete for resources in that environment through the secretion of enzymes or chemicals that are inhibitory to their competitors. In this study we will focus on known plant growth promoting bacteria including two N<sub>2</sub>-fixing bacteria, Azospirillumbrasilense(a Gram negative rhizobacterium) and Paenibacilluspolymyxa(a Gram positive endophyte), and two non-N<sub>2</sub>-fixing bacteria, *Pseudomonas penetrans*(a Gram negative rhizobacterium) and *Bacillus cereus*(a Gram positive endophyte). After planting and through growth, switch grass seedlings will be inoculated with all individual, all paired combinations and the four-way combination of bacteria by adding 5 ml of 10<sup>7</sup> CFU ml<sup>-1</sup> culture broth to an otherwise sterile sand growth media. A modified Hoagland solution will be added to support switchgrass growth, but N levels in the growth media will be halved. Switchgrass seedling growth will be observed and monitored over the course of four weeks and 12 weeks, until harvest. Switchgrass seedling growth will be assessed through tissue analyses of root, shoot and leaf C and N and characterization of root architecture. Microbial growth and abundance will be assessedvia MPN counts on selective media after isolating organisms in the rhizosphere (root surfaces) and endosphere (root, stem and leaf tissue). These data will help us understand how N<sub>2</sub>fixing bacteria and other growth promoting bacteria interact with each other and influence switchgrass growth.

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

**Jack Brown**<sup>\*1</sup>, Jim B. Davis<sup>1</sup>, Aaron Esser<sup>2</sup>, Kurt Schroeder<sup>1</sup>, Fangming Xiao<sup>1</sup> and Zhang Zhiwu<sup>2</sup>. <sup>1</sup>University of Idaho. <sup>2</sup> Washington State University. \* Presenting PI (<u>jbrown@uidaho.edu</u>)

Non-food Brassica crops have potential to produce non-food oil feedstocks suitable for biofuel production. Developing oilseed Brassica cultivars with higher seed and oil yield, high oil quality, and blackleg resistance and grown with low input costs will reduce importation, help break our fossil fuels dependence, and increase national security. In this project we will (1) Develop superior non-food grade oilseed cultivars suitable for a range of PNW and other US environments with blackleg and abiotic stress resistance suitable for high-quality biofuel feedstocks; (2) Dissect genetic architecture of industrial Brassica oilseed germplasm for yield, quality, and resistance to blackleg fungus through a genome-wide association study with highdensity genetic markers; (3) Develop and apply marker-assisted genomic selection techniques for blackleg resistance that will accelerate development of disease resistant oilseed cultivars; and (4) Use bioinformatics to identify putative PRR resistance genes that confer durable resistance to blackleg and use transgenic approaches to introgress PRR genes into adapted cultivar background. This project will utilize germplasm and the long-term industrial oilseed breeding program at the UI and the agronomy and molecular biology expertise from the UI and WSU. The genetically superior Brassica non-food grade oilseed cultivars developed will drastically increase domestic industrial oilseed production. In addition, we will identify new and novel plant resistance genes for blackleg resistance and molecular marker assisted selection tools to accelerate plant breeding procedures. Increased domestic production of biofuel feedstock oil will reduce dependence and importation of fossil fuels, increase fuel security, and ensure US agricultural competitiveness with foreign countries.

Analysis of an Abundant Bacterial Genus in a Leaf Litter Community

Alexander B. Chase (abchase@uci.edu),<sup>1</sup>\* Adam C. Martiny,<sup>1</sup> Ulas Karaoz,<sup>2</sup> Eoin L. Brodie,<sup>2</sup> and Jennifer B.H. Martiny<sup>1</sup>

<sup>1</sup>University of California, Irvine, CA; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, CA

http://jmartiny.bio.uci.edu/styled-2/index.html http://allison.bio.uci.edu/projects/a-trait-based-framework-for/

Project Goals: With increasing availability of cultured and uncultured (metagenomereconstructed and single-cell) genomes, it is now feasible to study the ecology and genetic diversity of specific microbial populations in complex communities. Previously, using 16S rRNA sequencing, we identified *Curtobacterium* (phylum: Actinobacteria, family: Microbacteriaceae) to be the most abundant taxon (~18% relative abundance) in a leaf litter grassland community in Southern California (Matulich *et al.* 2015). Here we analyze shotgun metagenomic libraries (Berlemont *et al.* 2014) at this site and ask: 1) How the abundance of *Curtobacterium*, and more specifically clades within *Curtobacterium*, responded over a two-year period (2010-2012) and across three treatments: drought, increased nitrogen, and control. 2) What is the genetic variation within populations of *Curtobacterium* over time?

We found that traditional characterization of microbial abundance in metagenomic libraries (i.e. MG-RAST) lack the genomic representation in their database to accurately quantify relevant taxa of interest, particularly within soil communities. To address these concerns, we isolated and sequenced 14 Curtobacterium genomes (Chase et al. 2016) and constructed a robust reference database and phylogenetic analyses into the Microbacteriaceae family. To quantify microbial abundance, we established an approach to extract and analyze single-copy marker genes within the metagenomic data. Normalizing the metagenomic data using 30 phylogenetically conserved marker genes, we estimated that the relative abundance of all Microbacteriaceae ranges from 22.76% to 29.12% across the samples. By placing short metagenomic reads onto this family reference tree, we showed significant strain-specific responses within clades of Curtobacterium over time. To validate these results, we then recruited and mapped reads to 4 genomes representing multiple clades. The mapping corroborated the abundance of specific clades within *Curtobacterium*, suggesting clade-specific responses to changing environmental conditions. Mapping of the reads further revealed genomic variation across clades, specifically with regards to the presence and/or absence of particular glycoside hydrolase genes. Together these results show that targeted analysis at fine taxonomic resolution can reveal population-specific characterizations that would otherwise be masked by broad taxonomic designations.

### References

Berlemont, R., Allison, S. D., Weihe, C., Lu, Y., Brodie, E. L., Martiny, J. B. H., & Martiny, A. C. (2014). Cellulolytic potential under environmental changes in microbial communities from grassland litter. Frontiers in Microbiology, 5, 639. http://doi.org/10.3389/fmicb.2014.00639

Chase, A. B., Arevalo, P., Polz, M. F., Berlemont, R., & Martiny, J. B. H. (2016). Evidence for Ecological Flexibility in the Cosmopolitan Genus Curtobacterium. Frontiers in Microbiology, 7, 1874. http://doi.org/10.3389/fmicb.2016.01874

Matulich, K. L., Weihe, C., Allison, S. D., Amend, A. S., Berlemont, R., Goulden, M. L., ... Martiny, J. B. H. (2015). Temporal variation overshadows the response of leaf litter microbial communities to simulated global change. The ISME Journal, 9(11), 2477–2489. http://doi.org/10.1038/ismej.2015.58

This research was funded by the Genomic Science Program, the (now retired) Program in Ecosystem Research (PER), and direct support to Lawrence Berkeley National Laboratory. The Genomic Science Program is part of the Biological System Science Division (BSSD) and PER was part of the Climate and Environmental Sciences Division (CESD) of Biological and Environmental Research (BER) in the US Department of Energy Office of Science. Long-term priming-induced changes in permafrost soil organic matter decomposition E Pegoraro<sup>1</sup>, R Bracho<sup>2</sup>, EAG Schuur<sup>1</sup>\* (<u>Ted.Schuur@nau.edu</u>), Junyi Liang<sup>3</sup>, Wenting Feng<sup>3</sup>, Lauren E. Hale<sup>3</sup>, **Liyou Wu<sup>3</sup>**, **Zhili He<sup>3</sup>**, **Jim Cole<sup>4</sup>**, **James M. Tiedje<sup>4</sup>**, **Konstantinos Konstantinidis<sup>5</sup>**, **Yiqi Luo<sup>4</sup> and Jizhong Zhou<sup>3</sup>** 

<sup>1</sup>Department of Biological Sciences, Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ.

<sup>2</sup>School of Forest Resources & Conservation, University of Florida, Gainesville, FL.
<sup>3</sup>Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK;

<sup>4</sup>Center for Microbial Ecology, Michigan State University, East Lansing, MI;

<sup>5</sup>Center for Bioinformatics and Computational Genomics and School of Biology, Georgia Institute of Technology, Atlanta, GA.

Project goal: The overall goal of this project is to advance system-level predictive understanding of the feedbacks of belowground microbial communities to multiple climate change factors and their impacts on soil carbon (C) cycling processes. The main objectives of this integrative project are to (i) determine the responses of microbial community structure, functions and activities to an increased input of easily decomposable C substrates to soil (priming effects); (ii) determine the extent to which priming enhances mineralization of native soil C; (iii) determine what proportion of the increased mineralization of native soil C is old C; (iv) determine if substrate input with different C quality distinctively affects microbial activity and soil organic matter decomposition; and (v) develop integrated bioinformatics and modeling approaches to scale information across different organizational levels. This study focuses on using laboratory incubations of soil as an isolated system to understand the influence of microbial processes on the release of C, and their response to changes in easily decomposable C substrate inputs.

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

Additional input of easily decomposable carbon can stimulate microbial activity, consequently increasing soil organic matter decomposition rates. This phenomenon, known as the priming effect, can exacerbate the effects of climate change by releasing more  $CO_2$  from soils however, the extent to which it could decrease soil carbon is unknown. We incubated permafrost soil from a moist acidic tundra site in Healy, Alaska for 456 days at 15° C. Soil from surface and deep layers were amended with three pulses of uniformly <sup>13</sup>C labeled glucose or cellulose, every 152 days. Substrate addition resulted in higher respiration rates in glucose amended soils; however, positive priming was only observed in deep layers. This suggests that microbes in deep layers are limited in energy, and the addition of easily decomposable carbon increases native soil organic matter decomposition. Here we also show, through data-model integration and synthesis, that approximately 58% of newly added C is transferred into SOC via replenishment, whereas the additional loss of old soil organic C (SOC) due to priming effect only accounts for 8.4% of the added new C in the first year after a one-time new C input. As a result, the new C input leads to a net increase in SOC, ranging from 40% to 49% of the added new

C. The magnitude of the net increase in SOC is positively correlated with the nitrogen-to-C ratio of the added substrates. Furthermore, a 100-year modeling experiment confirms that an increase in new C input leads to significant SOC accumulation over time. Our findings suggest that increasing plant productivity and the consequent increase in C input to soils likely promote SOC storage despite the enhanced decomposition of old C, which impacts C turnover time, but not necessarily C stocks.

This work is supported by the US Department of Energy, Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program (DE-SC00010715).

# Multiomic Insights into the Activity and Dynamics of Soil Nitrifier Communities in Midwestern Agricultural Soils

Luis Orellana<sup>\*1</sup>(lhorellana@gatech.edu), Joanne Chee-Sanford<sup>2</sup>, Robert Sanford<sup>3</sup>, Frank Löffler<sup>4,5</sup>, and Konstantinos Konstantinidis<sup>1</sup>.

<sup>1</sup>Georgia Institute of Technology, Atlanta; <sup>2</sup>US Department of Agriculture—Agricultural Research Service, Urbana; <sup>3</sup>University of Illinois, Urbana; <sup>4</sup>University of Tennessee, Knoxville, Tennessee; <sup>5</sup>Bioscience Division, Oak Ridge National Laboratory, Oak Ridge.

### **Project Goals:**

The goals of this project are to fill existing knowledge gaps in our understanding of N-flux and associated C-turnover in soils and sediments. Novel information about the diversity, distribution, abundance and gene expression contributing to N-transformation is required to link desirable (i.e., N-retention) and undesirable (i.e., N-loss, such as  $N_2O$  emissions) activities with measurable microbial parameters. Linking molecular- and organismal-level information with environmental factors can be used to better predict the impact of land management practices on N- and C-turnover. Such integrated approaches generate novel information on multiple scales of resolution and contribute to system-level understanding of key nutrient cycles in soils. In the present work, we analyzed the response of microbial communities to agricultural practices (e.g., addition of N-fertilizer) in two distinct agricultural soils, an important soil ecosystem for bioenergy crop production in Midwest US.

### Abstract:

Assessing the impact of fertilizer overuse on microbial soil communities is important for a better understanding of the cycling of C and N in soils. However, integrating field metadata (e.g., temperature, moisture, and oxygen) with the activity and dynamics of microbial communities in order to provide a systems-level understanding of nutrient cycling remains challenging, especially for soils. To this end, we analyzed short-read metagenomes obtained from two agricultural sites with contrasting soil textures (sandy versus silt loam) during four seasons in 2012 at two depths: surface (0-5cm) and deep (20-30 cm). Distinct archaeal populations and N metabolism genes were characteristic of the deep samples. To overcome the limitations of fixed e-values cut-offs for annotation of short-read metagenomes and to reduce false positive matches, we developed a novel computational approach, called ROCker, that employs the receiver operating characteristic (ROC) curve to minimize the false discovery rate (FDR) based on how simulated shotgun metagenomic reads of known composition map onto well-curated reference protein sequences. ROCker typically showed 60-fold lower false positive rates compared to the common practice of using fixed e-values and hidden Markov models. Application of the ROCker approach to the time series metagenomes showed that most N cycling genes (e.g., nosZ, amoA and *nirK*, among others) varied in abundance over the course of the year. For instance, we found a remarkably high abundance of metagenomic reads related to the under-studied Clade II nosZ (reduction of N<sub>2</sub>O to N<sub>2</sub>) sequences, accounting for approximately 90% of the total nosZ reads found in both soil layers. Approximately 12% of the nosZ reads were taxonomically assigned to the Anaeromyxobacter genus, indicating their potential relevance in N<sub>2</sub>O reduction. Population binning allowed the recovery of 69 draft genomes, including novel nitrifier archaea and bacteria. Six bins encoding *amoA* represented five new members of the *Thaumarchaeota* phylum and

three nitrifier populations represented a new bacterial genus, most related to the commamox *Nitrospirae*. These nitrifier populations, especially the archaeal ones, were observed to sharply increase in abundance upon N fertilizer application in the 20-30 cm soil layer in sandy soils, suggesting that they were responsible for a large part of (the fertilized) ammonia oxidation. In addition, time-series metatranscriptomic analyses of N-amended soil mesocosms (simulating a fertilization event) showed a high correspondence between *amoA* transcript abundance and ammonium disappearance ( $r^2 ~ 0.96$ ), and that bacterial nitrifiers were comparatively more responsive than their archaeal counterparts. Collectively, our study identified measurable biomarkers and novel microbial populations controlling the fate of fertilizer N and N<sub>2</sub>O emissions in agricultural soils, revealed key differences in the responding populations from field and incubated soils, and advanced the molecular toolbox for studying *in-situ* processes.

### **References:**

- Robert A. Sanford, Darlene D. Wagner, Qingzhong Wu, Joanne C. Chee-Sanford, Sara H. Thomas, Claribel Cruz-García, Gina Rodríguez, Arturo Massol-Deyá, Kishore K. Krishnani, Kirsti M. Ritalahti, Silke Nissen, Konstantinos T. Konstantinidis, and Frank E. Löffler. 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. Proc. Natl. Acad. Sci. U. S. A. 109: 19709–14.
- 2. Orellana, L. H., L. M. Rodriguez-R, S. Higgins, J. C. Chee-Sanford, R. A. Sanford, K. M. Ritalahti, F. E. Löffler, and K. T. Konstantinidis. 2014. Detecting nitrous oxide reductase (nosZ) genes in soil metagenomes: method development and implications for the nitrogen cycle. mBio 5: e01193-14.
- 3. Orellana, L. H., L. M. Rodriguez-R and K. T. Konstantinidis. ROCker: a pipeline for accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores. Nucleic Acids Research. doi: 10.1093/nar/gkw900

**Funding Statement**: This research was supported by the US Department of Energy, Office of Biological and Environmental Research, Genomic Science Program, Award DE-SC0006662.

# Factors Affecting Nitrous Oxide Production from Ammonia Oxidizers and Possible Mitigation Options

Kelley A. Meinhardt,<sup>1</sup> Stuart E. Strand,<sup>1</sup> Manmeet W. Pannu,<sup>1</sup> Anthony Bertagnolli,<sup>1</sup> Dylan McCalmont,<sup>1</sup> Steven C. Fransen,<sup>2</sup> Sally Brown,<sup>1</sup> Karen L. Casciotti,<sup>3</sup> and **David A. Stahl**<sup>1\*</sup> (dastahl@uw.edu)

<sup>1</sup>University of Washington, Seattle, WA; <sup>2</sup>Washington State University, Prosser, WA; <sup>3</sup>Stanford University, Stanford, CA

Project Goals: Biofuels produced from perennial crops, like switchgrass (*Panicum virgatum*), are a potential alternative to fossil fuels. Traditional cultivation practices often require the addition of synthetic nitrogen. However, much of the applied nitrogen is lost to the leaching of nitrate produced by nitrifying microorganisms. In addition, those organisms contribute directly and indirectly to the production of the greenhouse gas nitrous oxide (N<sub>2</sub>O). Over 60% of anthropogenic N<sub>2</sub>O production originates from fertilized agricultural soils. The goals of this project were: 1) to evaluate the environmental and land management factors that affect N<sub>2</sub>O production from cultivated switchgrass, 2) to resolve the relative contributions of ammonia-oxidizing microorganisms to nitrification and the production of N<sub>2</sub>O, and 3) to identify methods to reduce or suppress N<sub>2</sub>O production in such systems.

Cellulosic ethanol produced from perennial feedstock crops, such as switchgrass (*Panicum virgatum*), is a potential replacement for fossil fuels. However, high biomass yields require the addition of nitrogen (N) fertilizers that often has negative environmental impacts, including nitrate leaching and N<sub>2</sub>O production. N<sub>2</sub>O has a global warming potential ~300X greater than carbon dioxide, and emissions are expected to continually increase due to higher demand for crops and livestock from a growing world population. The processes responsible for most N<sub>2</sub>O production are microbially controlled, and under oxic conditions the contributing populations are ammonia-oxidizing archaea (AOA) and bacteria (AOB). However, the extent to which each group contributes to ammonia oxidation and N<sub>2</sub>O production, and the environmental and land management factors that affect their populations, are not known, with published findings differing.

The contributions of these two groups to ammonia oxidation and  $N_2O$  production were investigated in the field and laboratory. Soil from fields of switchgrass located at two sites, Prosser and Paterson, WA, differing in soil texture (silt loam vs. loamy sand, respectively), were examined for ammonia oxidizer community changes with N source and agricultural practices, as well as for correlation with  $N_2O$  production. Complementary lab studies using field soils constrained the relative contribution of each group using transcription, isotope fractionation, and selective inhibitor analyses.

During two field seasons, N<sub>2</sub>O fluxes were measured monthly from April through October from fields of switchgrass receiving no N (control), 224 kg/y of inorganic N (agronomic rate), and N derived from intercropped leguminous alfalfa (*Medicago sativa*). Simultaneously, soil samples were collected and assayed for AOA and AOB abundance via quantification of the gene encoding one subunit of the ammonia monooxygenase (*amoA*). N<sub>2</sub>O fluxes from fertilized plots were up to 16X higher immediately after N application and irrigation than from unfertilized control plots. Importantly, intercropping reduced N<sub>2</sub>O flux to one-third of that from fertilized treatments. However, the 2-year average biomass yield from intercropped plots (16.8  $\pm$  1.1

Mg/ha/yr) was intermediate between fertilized  $(24.5 \pm 1.2 \text{ Mg/ha/yr})$  and control  $(10.5 \pm 1.1 \text{ Mg/ha/yr})$  plots. AOA were the dominant in all treatments and were more abundant in the intercropped treatment than the control and fertilized treatments. Only AOB abundance was positively correlated with N<sub>2</sub>O emissions.

Terminal restriction fragment (TRF) analysis of *amoA* revealed a significant impact of management on AOA populations. Communities in native soils were similar, despite sampling sites being some distance apart. In contrast, populations in managed fields were comprised of different genotypes of both AOA and AOB. Those changes in population structure were correlated with soil pH and texture (i.e., particle size).

The same field soils were used to establish a series of microcosms receiving, or not receiving, N fertilizer. Temporal fluctuations in N<sub>2</sub>O emissions and isotopic composition were associated with N amendment and changes of AOA and AOB genes and transcripts. AOB *amoA* gene counts increased 45-fold after 10 days, whereas AOA increased only 1.5-fold following N addition. Although *amoA* transcripts of both AOA and AOB increased during the experiment relative to no-N controls, AOB transcripts increased over 85-fold from day 0 to 10, indicating a much greater response of AOB to N addition. Control soils retained a  $\delta^{15}N^{\text{bulk}}$ -N<sub>2</sub>O signature between -7 to +3 per mil, indicative of AOA fractionation, while the  $\delta^{15}N^{\text{bulk}}$ -N<sub>2</sub>O from fertilized microcosms reached values of -49 per mil, which are within the fractionation range of AOB.

Selective inhibitors of nitrification were used to further document the relative contributions of AOA and AOB to ammonia oxidation and associated N<sub>2</sub>O emissions in response to N addition. Four inhibitor treatments were applied to control and fertilized soils: 1) acetylene to inhibit both AOA and AOB, 2) PTIO to selectively inhibit AOA, 3) 1-octyne to selectively inhibit AOB, and 4) no inhibitor (positive control). After 10 days, the acetylene-treated soil showed no decrease in ammonia or increase in nitrate or N<sub>2</sub>O, indicating both ammonia oxidizers had been inhibited. The 1-octyne-treated soils showed a lack of nitrate and N<sub>2</sub>O accumulation similar to the acetylene treatment, suggesting AOA produced very little of these compounds. The no inhibitor control showed a balanced consumption of ammonia and production of nitrate, while N<sub>2</sub>O production was high. In the PTIO treatment, nitrate and N<sub>2</sub>O producers. A titration of N additions with selective inhibitors indicated that, even when small amounts of synthetic N were supplied, AOB were stimulated and produced N<sub>2</sub>O in this soil.

Our results provide evidence that agricultural practices, as well as soil properties, can affect ammonia-oxidizing communities and  $N_2O$  emissions. Synthetic N fertilizer, even in small amounts, led to  $N_2O$  production by AOB. Intercropping switchgrass with alfalfa does enhance biomass yields (relative to control plots) and reduces  $N_2O$  emissions from AOB (relative to fertilized plots). Selective inhibitors that promote AOA and suppress AOB could be a useful strategy to reduce microbial production of  $N_2O$  in bioenergy feedstock croplands.

This research was supported by the Department of Energy's Genomic Science and Technology for Energy and the Environment grant DE-SC0006869.

### A population of copy number variants for poplar functional genomics.

Isabelle M. Henry<sup>1\*</sup> (imhenry@ucdavis.edu), Héloïse Bastiaanse<sup>2</sup>, Courtney Canning<sup>2</sup>, Matthew S. Zinkgraf<sup>2</sup>, Andrew T. Groover<sup>1,2</sup> and Luca Comai<sup>1</sup>

<sup>1</sup> University of California, Davis, <sup>2</sup> Pacific Southwest Research Station, U.S. Forest Service, Davis

http://comailab.genomecenter.ucdavis.edu/index.php/Poplar http://www.fs.fed.us/psw/programs/cb/staff/agroover/

Changes in gene dosage can affect gene function in multiple ways, and inducing dosage mutations (insertions and deletions) is a powerful approach to rapidly create wide phenotypic variation. We have produced and characterized a population of interspecific poplar hybrids carrying insertions and deletions tiling the entire genome multiple times. We are using this resource to identify genes that contribute to specific traits, initially focusing on poplar bioenergy traits. For this funding cycle, our specific objectives were to i) finish characterize and maintain this resource ii) investigate the phenotypic effects of dosage change and iii) exploit the indel germplasm for functional genomics. Outputs include a population of interspecific poplar hybrids carrying defined dosage variation and extensively characterized trait measurements, and possibly cultivars directly usable for bioenergy applications. The approach and tools developed here can be easily applied to other vegetatively-propagated species.

Poplar breeding is predominantly based on interspecific hybridization, harnessing the advantage of hybrid vigor. The resulting F1 hybrids frequently exhibit dosage variation, either in the form of whole genome duplication (triploidy) or in the form of copy number variation of pieces of, or entire chromosomes (aneuploidy). These variants can have transgressive phenotypes, sometimes desirable for biomass production or other characteristics of interest. Inducing large-scale copy number variation is therefore a rapid method for creating new variants. This approach is rarely used in sexual species because the resulting variation is often meiotically unstable but in clonally propagated species, such as the fast growing tree genus *Populus*, it holds many advantages. Using gamma irradiation of pollen grains, we have created a population of ~800 interspecific F1 sibling that vary in their chromosomal composition. We have found that approximately 50% of them carry at least one large-scale insertions or deletions (Henry et al, 2015). Phenotypic analysis confirmed that dosage variation is overall well tolerated and samples with distinct morphological alterations are present within the population.

We are completing the development of this unique resource for the research community and have started to use this germplasm as a functional genomics tool. Using precise phenotypic data, gene dosage information, and soon, associated gene expression information, we are investigating the role of gene dosage in poplar hybrid performance. By looking at the association between trait value and copy number across the genome, we are able to identify dosage qtls that contribute to a variety of quantitative traits (see Figure). We have so far focused primarily on phenotypic traits related to biomass production, leaf morphology and wood properties but have also started to acquire information about other traits such as drought or biotic stress tolerance and the population is available for other measurements as desired. Our population provides a powerful platform for both understanding gene function and the effect of gene dosage on phenotypes and

on poplar hybrid performance. This resource is publicly available for others to investigate specific traits of interest.



**Identification of dosage qtls. A.** For each F1 interspecific hybrid individual, the presence of an additional copy (insertion) or the deletion of a copy of a particular chromosome fragment was identified by detecting changes up or down in sequence coverage (arrows). **B.** Subset of lesions identified on chromosome 1. Lesions are tiled across the length of the chromosome. Dosage lesions vary in length (up to a whole chromosome) and a single individual carried up to 10 lesions. Deletions were most common but insertions were observed as well. For dosage qtl analysis (C), dosage markers were defined based on lesion boundaries. **C.** Dosage qtl analysis identified regions of the genomes for which copy number is associated with trait value. In this particular case, the timing of bud burst was recorded and found to be strongly associated with loci on 7 different chromosomes.

### References

1. Henry IM, Zinkgraf MS, Groover AT, and Comai L. A System for Dosage-Based Functional Genomics in Poplar. Plant Cell **2015**, 27:2370-2383.

*This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grants no. DE-SC0007183.* 

### Repurposing the Yeast Peroxisome for Compartmentalizing Multi-enzyme Pathways

Jennifer A. Samson,<sup>1,\*</sup> (j.samson@berkeley.edu), William C. DeLoache,<sup>1</sup> Zachary N. Russ,<sup>1</sup> and **John E. Dueber**<sup>1,2</sup>

<sup>1</sup>University of California, Berkeley; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, California

#### **Project Goals**

Our long-term project goal is to establish the peroxisome as a synthetic organelle that can compartmentalize multiple enzymes and be made selectively permeable to substrates and cofactors to achieve high flux for this engineered metabolic pathway. Success would enable pathways to function with minimal crosstalk with cytosolic factors as well as engineer environments distinct from the cytosol (e.g., redox state or pH could potentially be altered in this subcompartment). Towards this overall goal, we are striving to repurpose the peroxisome in *S. cerevisiae* and are addressing multiple challenges that will be discussed herein: 1. efficiently targeting heterologous cargo into the peroxisome lumen, 2. determining the rules for natural metabolite permeability through this organelle's lipid bilayer, 3. testing the efficacy of compartmentalizing a model pathway, and 4. genetically engineering the peroxisome membrane to have reduced permeability.

### Abstract

Eukaryotic organisms have evolved organelles for spatially separating certain biochemical reactions from others as well as the creation of distinct chemical environments. Similarly, it would be valuable to have synthetic control over an organelle such that the identity of the imported enzymes can be determined to catalyze desired reactions and avoid undesired reactions that may occur in the cytosol. We have decided to proceed towards this grand aim by repurposing an existing organelle - the peroxisome. The choice of this organelle were several-fold. First, the peroxisome is not required for *S. cerevisiae* viability provided long chain fatty acids are not used as a sole carbon source. Second, this organelle is induced to impressive sizes in some methanol utilizing yeasts such as *Pichia pastoris* and *Hansenula polymorpha*, sometimes greater than 70 percent of total cell volume. Third, the peroxisome has been naturally specialized by several organisms to alter the composition of cargo protein for new functions.

Our recent work has determined a size-dependent (sieve-like) leakiness to small molecules smaller than approximately 400-730 Daltons (Deloache, Russ, & Dueber, 2016). These measurements were made using an enzyme sequestration assay we developed for *in vivo* determination of permeability, removing the need for organelle purification that has complicated previous studies in the field; as such, our work provides the strongest evidence yet on this long-standing debate. However, the permeability of the peroxisome membrane in *S. cerevisiae* is a substantial limitation for repurposing this organelle as a generalizable compartment for multi-enzyme pathways. As such, we are faced with the challenge of engineering the peroxisome for reduced permeability to small molecules. We hypothesize that the protein

importomer (the complex enabling translocation of protein from the cytosol into the peroxisome lumen) is the major source of this leakiness. However, knocking out the importomer sets up a "chicken and egg" conundrum: how can proteins be trafficked into the peroxisome lumen in the absence of the protein importomer?

To determine whether the importomer is the main determinant for peroxisome permeability we have developed a strategy to target heterologous membrane proteins to the peroxisome, that have been rerouted from the ER. The ER provides the hub for sorting and trafficking membrane proteins to many organelles, including the plasma membrane, Golgi apparatus, vacuole, and peroxisomes. We have shown that the Neurospora crassa cellodextrin transporter 1 (CDT1) can be targeted to the peroxisome membrane and that CDT1 is functional at the peroxisome membrane, observed by transport of Xglucoside substrates into the peroxisome lumen. Going forward, we aim to use this strategy to 1) traffic soluble protein to the peroxisome that has firstly been directed to the ER lumen using a heterologous membrane protein scaffold and 2) direct orthogonal protein transporters from systems with tight regulation of small molecule transport across the membrane, such as the TAT translocon from the plasma membrane of E. coli. These systems can be screened in cells with the native protein import machinery knocked-out in order to determine the genes/factors responsible for peroxisome permeability using our in vivo assay (Deloache, Russ, & Dueber, 2016), in a systematic way. Ultimately, engineering robust methods to target soluble protein to the peroxisome lumen, in the absence of the native machinery, requires sophisticated understanding of both cell biology and engineering disciplines to address these challenges.

#### References

1. Deloache, W. C., Russ, Z. N., & Dueber, J. (2016). Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways. Nature Communications, 7, 11152. http://doi.org/10.1038/ncomms11152

#### Funding statement.

The funding for this project is from DOE BER Award DE-SC0008084.

#### Notes on abstract:

- Note the placement of superscripts in the authors and affiliations.
- URL above should be specific to the project. More than one URL is permitted.
- **References** can be **Publications** instead, if needed. Use any common style for these citations.

### Application of machine learning and active learning to enhance chemical yields in microbes

Prashant Kumar<sup>1</sup>\* (pkumar29@wisc.edu), Paul A. Adamczyk<sup>1,2</sup>, Xiaolin Zhang<sup>1</sup>, Parameswaran Ramanathan<sup>3</sup>, and **Jennifer L. Reed**<sup>1,2</sup>

<sup>1</sup>Department of Chemical and Biological Engineering, UW-Madison, Madison, WI 53706

<sup>2</sup>DOE Great Lakes Bioenergy Research Center, Madison, WI 53706

<sup>3</sup>Department of Electrical and Computer Engineering, UW-Madison, Madison, WI 53706

Project Goals: Cyanobacteria offer a promising route for directly converting solar energy and CO<sub>2</sub> into biofuels. The objectives of this research are to integrate modeling and experimental approaches to guide development of a butanol producing cyanobacterium, *Synechococcus* sp. PCC 7002. New computational approaches will be developed to facilitate these efforts which will (1) design experiments and analyze their results, and (2) identify genetic engineering strategies for improving butanol production in *Synechococcus* sp. PCC 7002. Experiments will subsequently be performed to construct and analyze *Synechococcus* sp. PCC 7002 strains engineered for butanol production. The developed approaches will be systematically applied to suggest genetic engineering strategies for improving production of a variety of biofuels in five other microorganisms. This research will support the U.S. Department of Energy's mission for developing renewable ways of producing advanced biofuels.

Renewable sources of transportation fuels are needed to reduce the amount of oil used to satisfy transportation energy needs in the U.S. and to alleviate our dependence on foreign sources of oil. Microbes can be used to produce a wide variety of liquid biofuels including: ethanol, butanol, isobutanol, isoprene, hydrogen, and alkanes. Cyanobacteria offer an alternative route for converting solar energy and  $CO_2$  into biofuels, without the need for using lignocellulosic biomass as an intermediate. The biofuel production capabilities of microbes can be improved through metabolic engineering, where metabolic and regulatory processes are adjusted using targeted genetic manipulations. Traditionally, metabolic engineering strategies are found through manual inspection of metabolic pathways, where enzymes involved in biosynthesis are overexpressed or added, competing pathways are eliminated, and the performance of resulting strains are evaluated. However, such manual approaches cannot predict the effects that these changes will have on metabolism and the enzyme levels needed to optimize flux through a metabolic pathway.

Efforts have been made to develop computational methods to study metabolic and regulatory networks of microbes and identify the genetic interventions needed to produce desired high-value metabolite(s) from low-cost substrates. Genome-scale constraint-based metabolic models rely on information on reaction stoichiometry to identify reaction or gene deletion and addition strategies to enhance product yield. However, these models cannot predict how changes to enzyme levels will impact metabolic fluxes. In contrast, smaller kinetic models can be used to suggest additional strategies based on changes in enzyme levels and/or kinetic properties, but these kinetic models need a lot of experimental data (e.g., proteomic, metabolomic, and fluxomic data) to parameterize them. Hence, there is a need for computational methods that can predict expression levels needed to achieve metabolic engineering goals with limited amounts of experimental data and no kinetic details about the system.

We developed an active learning framework called **ActiveOpt** to design expression constructs for a metabolic pathway of interest. ActiveOpt does not need a detailed kinetic model and instead uses a linear Support Vector Machine (SVM) classifier to predict product yields or productivities (either high or low) from ribosome binding site strengths estimated by the RBS Calculator [1]. ActiveOpt initially trains a SVM classifier from a few experiments, where RBSs in gene expression constructs are varied and product yields are measured and labeled (as high/low yield), and then proposes subsequent experiments to be conducted. ActiveOpt, with relatively little experimental data and no mechanistic or kinetic details of the pathway, can be used to design experiments to achieve high biochemical yields in a small number of experiments.

ActiveOpt was tested on two separate datasets: (1) a newly generated valine yield dataset and (2) a published neurosporene productivity dataset [2]. The valine dataset included 91 experiments, in which two plasmids, that express nine valine biosynthesis and exporter genes (*ilvBNIHCDE* and *ygaZH*) with varying RBS strengths, were transformed into *Escherichia coli* and valine yields were measured in glucose+acetate minimal medium. A leave-one-out cross validation showed that SVM classifiers built from this dataset have high precision (75%) and recall (87%). Starting with just a few of the 91 possible experiments, ActiveOpt could identify expression constructs resulting in at least 95% of the highest measured valine yield (across all conducted 91 experiments) in a small number of experiments (typically <7) and identify the genes whose RBS strengths significantly affect valine yield. Further, ActiveOpt was used to propose four new experiments (beyond the original 91 valine experiments) that were predicted to have high yields. Valine yields in those four new experiments were found to be high, with one strain having 53.38% of the maximum theoretical yield, which is close to the best yield found in the 91 previously conducted experiments (54.70%). ActiveOpt was also tested on a previously published neurosporene dataset, and the algorithm could again identify the expression constructs with high productivity in less than 10 experiments as compared to the 101 experiments conducted in the original study.

These results show that ActiveOpt can efficiently design gene expression constructs that lead to high chemical yield in organisms in very small numbers of experiments. It can also identify the genes whose expression (as predicted by RBS strengths) significantly influence biochemical production. Our next step is to use ActiveOpt to design experiments to identify strains with high butanol yield in the cyanobacterium *Synechococcus* sp. PCC 7002.

### References

- 1. Salis, H.M., E.A. Mirsky, and C.A. Voigt, *Automated design of synthetic ribosome binding sites to control protein expression.* Nat Biotechnol, 2009. **27**(10): p. 946-50.
- 2. Farasat, I., et al., *Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria.* Mol Syst Biol, 2014. **10**: p. 731.

*Funding statement.* This work was supported by the Office of Science (BER), U.S. Department of Energy (DE-SC0008103), the W.M. Keck Foundation, and the College of Engineering at the University of Wisconsin-Madison.

**Factors Governing Mutualism Dynamics in a Hydrogen-Producing Coculture** Alexandra L. McCully,<sup>1</sup> Breah LaSarre<sup>1</sup>, Ryan K. Fritts<sup>1</sup>, Maureen C. Onyeziri<sup>1</sup>, Jay T. Lennon<sup>1</sup>, and **James B. McKinlay<sup>1\*</sup>** (jmckinla@indiana.edu) <sup>1</sup>Indiana University, Bloomington

# Project Goals: The goals of this project are to (i) develop a stable hydrogen gas-producing coculture between *Rhodopseudomonas palustris* and fermentative microbes, such as *Escherichia coli*, (ii) use genetic, biochemical, evolutionary, and systems biology approaches to characterize and manipulate microbial interactions and H<sub>2</sub> production.

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

We developed an anaerobic coculture between fermentative *Escherichia coli* and an engineered strain of phototrophic *Rhodopseudomonas palustris* (Nx) that together convert carbohydrates into H<sub>2</sub> gas, a potential biofuel. The two bacteria form a syntrophic relationship wherein *E. coli* ferments glucose and excretes essential carbon (organic acids) for *R. palustris* while *R. palustris* fixes N<sub>2</sub> and excretes essential nitrogen (NH<sub>4</sub><sup>+</sup>) for *E. coli*. The bidirectional exchange ensures stable coexistence as two populations converge on a common equilibrium from starting ratios spanning over six orders of magnitude favoring either species. Growth and metabolic trends are highly reproducible over serial transfers. We have also developed kinetic models that accurately describe coculture dynamics and can be used to predict the effects of various perturbations<sup>1,2</sup>. We are now using computational and experimental approaches to address the importance of core metabolic traits in deciding the phenotypic behavior of this community.

To generate ATP, *E. coli* must excrete organic acids under fermentative conditions. Thus, even under complete nitrogen-starvation, *E. coli* will continue to ferment glucose to organic acids to maintain itself. This trait has important implications for coexistence<sup>2</sup>. Our model suggested that growth-independent fermentation is critical for sustaining coexistence at low levels of  $NH_4^+$  transfer. When growth-independent fermentation was excluded from the model, the coculture was predicted to go extinct below a threashold level of  $NH_4^+$  transfer. We decreased  $NH_4^+$  transfer in our coculture by lowering the supply of N<sub>2</sub>. Coexistence was observed at even the lowest levels of  $NH_4^+$  transfer that we tested, suggesting an essential role for growth independent fermentation. The continuous excretion of organic acids stimulates *R. palustris* growth and reciprocal  $NH_4^+$  excretion, eventually lifting both species out of starvation. Organic acids are an obligate waste product of fermentative organisms and an important carbon and electron shuttle in anaerobic communities. Thus the importance of growth-independent fermentation in establishing and maintaining cross-feeding relationships could be widespread in nutrient-limited environments. The highest coculture H<sub>2</sub> yields were also observed under severely nitrogen-limiting conditions, exceeding the theoretical maximum fermentative H<sub>2</sub> yield.

We also found that growth-independent fermentation can be detrimental to coexistence when the *E. coli* population is large. Such large populations amplify the rate of growth independent fermentation and result in rapid organic acid accumulation that acidifies the medium before growth can occur. This result highlights a dual role for organic acids, moving along a continuum between being beneficial versus detrimental.

We also determined that the relative benefit *R. palustris* receives from organic acids can be influenced by its own level of cooperative  $NH_4^+$  excretion<sup>1</sup>. Our model predicted that more  $NH_4^+$  excretion would result in a less efficient utilization of feedstock and a lower *R. palustris* cell density. The higher levels of  $NH_4^+$  would stimulate rapid *E. coli* growth and organic acids would be produced faster than *R. palustris* could consume them, resulting in a growth-inhibiting acidic pH. Thus, by cooperating more, *R. palustris* would change the nature of organic acids from a carbon source to a growth inhibitor. To test these predictions, we engineered a 'hyper-cooperator' strain of *R. palustris* that excretes 3-fold more  $NH_4^+$  than the Nx parent. Cocultures with the hyper-cooperator confirmed the predictions, as there were fewer *R. palustris* cells, a higher residual organic acid concentration, and a more acidic pH. Nonetheless, the hyper-cooperator and *E. coli* stably coexisted over serial transfers, albeit at a new equilibrium.

Our results inform on the potential for both positive and negative roles of organic acids in anaerobic food webs and the influence that microbial interactions can have on deciding these roles. Our results also have implications for the use of engineered  $NH_4^+$ -excreting, N<sub>2</sub>-fixing bacteria in both industrial and agricultural settings. Varied  $NH_4^+$  excretion rates can indirectly modify system behavior in potentially negative ways, such as altering the species ratio or misdirecting resources towards undesired products. In other words, optimizing  $NH_4^+$  cross-feeding does necessarily mean increasing  $NH_4^+$  cross-feeding.

### References

- 1. LaSarre, B, AL McCully, JT Lennon, and JB McKinlay. 2017. Microbial mutualism dynamics governed by dose-dependent toxicity of cross-fed nutrients. ISME J. doi: 10.1038/ismej.2016.141
- 2. McCully, AL\*, B LaSarre\*, and JB McKinlay. Growth-independent cross-feeding modifies boundaries for coexistence in a bacterial mutualism. biorxiv.org/content/early/2016/10/25/083386 \*equal contribution.

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

### **Evaluating the Cellulolytic Properties of Novel Fibrobacteres isolates from Mammalian Herbivores**

Anthony P. Neumann<sup>1</sup>, Caroline A. McCormick<sup>1</sup>, Paul J. Weimer<sup>2</sup>, and **Garret Suen**<sup>1\*</sup> (gsuen@wisc.edu)

<sup>1</sup>University of Wisconsin-Madison, Madison; <sup>2</sup>USDA-ARS Dairy Forage Research Center, Madison

Project Goals: Characterizing novel cellulose-degrading microbes is of great interest for the development of biofuels and bioproducts from cellulosic feedstocks. One group of bacteria includes the phylum Fibrobacteres, which are well known for the prolific cellulose-degrading properties. Surprisingly, our understanding of the entire phylum is limited to only two described species within the genus *Fibrobacter*. Here, we sought to expand our understanding of this important cellulolytic phylum by culturing novel isolates from a multitude of different herbivorous hosts. We tested each isolates ability to degrade cellulose and compared them to the type strain *Fibrobacter succinogenes*. This work expands on our existing knowledge about the Fibrobacteres and further provides access to an untapped diversity of novel cellulolytic enzymes.

Bacteria in the phylum Fibrobacteres are well known for the highly cellulolytic properties. For example, the type strain Fibrobacter succinogenes S85 is the most prolific cellulose degrader known. Non-culture based approaches have identified Fibrobacteres bacteria from the gastrointestinal tract of a wide range of mammalian herbivorous hosts, yet these species have thus far remained resistant to culturing efforts. Given the highly cellulolytic nature of the existing *Fibrobacter* species, we posited that species associated with mammalian herbivores would also exhibit similar properties. This idea is supported by the fact that F. succinogenes exists within the cow rumen, one of nature's most highly optimized plant biomass degrading ecosystems. To test this, we developed a novel culturing method and employed to obtain 45 new isolates from 11 different herbivorous hosts. Extensive phylogenetic analysis of these isolates expands our understanding of this phylum by revealing 6 new clades, with many representing new species and genera. Isolates tended to group by host, and between hosts there were distinct groupings into hindgut and foregut fermenters. We then tested each isolates ability to degrade both cellulose and hemicelluloses, and found that, like the type strain, all isolates were only able to degrade and utilize cellulose. These findings suggest that members of this phylum are cellulose specialists, as has been observed for the type strain. Further analysis of these isolate's fermentation products revealed that, like F. succinogenes, all cultures produced succinate as the major fermentation product, with lesser amounts of acetate and formate. Our results expand our understanding of the Fibrobacteres and provides insights into the ecology and evolution of a prolific cellulose-degrading specialist.

*This work was supported by a DOE Early Career Research Program Award DE–SC0008104 to GS.* 

# Differential Expression, Regulatory Divergence, and Sex Dimorphism Pervade the Shrub Willow (*Salix* spp.) Transcriptome

Craig H. Carlson<sup>1\*</sup>, Yongwook Choi<sup>2</sup>, Agnes Chan<sup>2</sup>, Michelle J. Serapiglia<sup>1</sup>, Fred E. Gouker<sup>1</sup>, Stephen P. DiFazio<sup>3</sup>, Christopher D. Town<sup>2</sup>, and Lawrence B. Smart<sup>1</sup>

<sup>1</sup>Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY

<sup>2</sup> Plant Genomics Group, J. Craig Venter Institute, Rockville, MD

<sup>3</sup> Department of Biology, West Virginia University, Morgantown, WV

http://willow.cals.cornell.edu

**Project Goals:** Many studies have highlighted the complex, multigenic basis for heterosis (hybrid vigor) in inbred crops. Despite the lack of a consensus model, it is vital that we turn our attention to understanding heterosis in undomesticated, outcrossing and polyploid species, including shrub willow (*Salix*). A consistent trend in willow breeding is the success of triploid progeny produced from crosses between diploid and tetraploid species. We will quantify heterosis for yield and biomass traits across 8 families representing intraspecific and interspecific diploid and interspecific triploid progeny. We will quantify allele-specific gene expression and inheritance patterns in hybrid progeny using RNA-Seq. We will correlate these expression data with phenotypic characterization of heterosis for yield and biomass composition determined in replicated greenhouse and field trials. We will look for networks of coordinated gene regulation controlling yield and lignocellulosic deposition.

Recent genomic advances have provided the biomass feedstock community with new tools to improve traits related to biomass yield and wood chemical composition. In *Salix*, hybrid vigor is apparent in interspecific crosses and tends to be more pronounced in triploid progeny derived from the hybridization of tetraploid and diploid parents. Progeny and parents of full-sib intraand interspecific F<sub>1</sub> and F<sub>2</sub> shrub willow families have been examined in order to define the basic architectures and inheritance patterns of transcriptome-wide gene expression. Our main objectives of this study were (1) to test for differential expression among the transcriptomes of segregating diploid and triploid families with regards to contrasting sex, tissue type, and midparent expression, (2) to categorize transgressive gene expression into modes of inheritance, (3) to assess the magnitude of regulatory divergent expression, and (4) to correlate modules of co-expressed genes in parents and progeny with traits important for biomass yield. We show allele-specific expression is largely conserved in intraspecific families and highly divergent in interspecific families. Akin to the heterosis observed in the field, as the complexity of the cross increases, regulatory divergent gene expression increases. In addition, we utilize a full-sib F<sub>2</sub> S. *purpurea* mapping population, planted in a replicated field trial, to supplement expression data. Candidate genes that have been or have yet to be identified will be confirmed via allele-specific assays. These data will be used to develop predictive models of heterosis and complement the growing genomic resources available for the improvement of shrub willow bioenergy crops.

#### Figure 1. Magnitude of sexually dimorphic inheritance in Salix purpurea.

Boxplots (a) summarize the log<sub>2</sub> normalized expression differences for genes with sexually dimorphic inheritance patterns (teal) and those with same-sex inheritance (grey), by sex. Scatterplots compare log<sub>2</sub> normalized expression of F<sub>1</sub> (b) males and (c) females to the maternal (P1, x-axis) and paternal (P2, y-axis) expression. Points represent only genes with dimorphic inheritance patterns. Pie charts within the scatterplots summarize patterns of gene expression inheritance for genes with dimorphic gene expression for each sex. The scatterplot (d) illustrates overlain coordinates of gene expression inheritance for males and females, where each gene is represented by male ( $m_{xy}$ , points) and female ( $f_{xy}$ , arrows) vectors, connected by a single line segment. Each segment is equally divided by two colors which correspond to the male and female inheritance class for each gene as the absolute magnitude of dimorphic gene expression inheritance was calculated for each gene as the absolute Euclidean distance (L<sup>2</sup>) between  $m_{xy}$  and  $f_{xy}$  on the same Cartesian plane. For those genes with dimorphic inheritance, boxplot distributions (e) of conserved (grey) and nonadditive (teal) inheritance patterns for males and females and females depict differences in their absolute magnitude.



This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0008375.

#### Model-Guided Metabolic Engineering of Increased 2-Phenylethanol Production in Plants

Joseph Lynch<sup>1</sup>, Shaunak Ray\*<sup>2</sup>, Clint Chapple<sup>1</sup>, Natalia Dudareva<sup>1</sup>, John A. Morgan<sup>1,2</sup>
 <sup>1</sup>Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907
 <sup>2</sup>School of Chemical Engineering, Purdue University, West Lafayette, Indiana 47907

**Project Goals:** We are testing a metabolic engineering strategy for the overproduction of 2phenylethanol, a potential biofuel. Our approach is to first overexpress the enzymes catalyzing the multiple biosynthetic steps from phenylalanine to 2-phenylethanol in *Arabidopsis thaliana*. The strategy utilizes a single gene cassette to simultaneously express all proteins in near stoichiometric amounts under the control of a single promoter. The second part of the strategy is to increase the amount of the precursor, phenyalanine. Results from transgenic plants will be incorporated into kinetic models which will be used for identifying targets of future metabolic engineering strategies for optimized biofuel production.

2-Phenylethanol (2-PE) is a naturally occurring volatile organic compound with properties that make it a candidate oxygenate for petroleum-derived gasoline. However, its use for this purpose is limited by a lack of economically viable large scale production. 2-PE is produced naturally in some plant tissues via the sequential deamination/decarboxylation and reduction of phenylalnine. This pathway competes with the phenylpropanoid pathway for the common precursor phenylalanine. The phenylpropanoid pathway directs approximately 30% of carbon flux towards the biosynthesis of lignin, a major constituent of plant cell walls that impedes the process of cellulosic biofuel production. Therefore, we propose a genetic engineering strategy at the phenylalanine branch point, whereby a portion of the carbon flux towards lignin biosynthesis is diverted towards the production of an economically valuable product, 2-PE. Transgenic Arabidopsis thaliana were generated to introduce the pathway for production of 2-PE through overexpression of an aromatic aldehyde synthase (AAS) in tandem with phenylacetaldehyde reductase (PAR), which successfully increased production of 2-phenylethanol. However, as was previously reported for similar strategies, the *in planta* accumulation observed remains far lower than desired. To assess metabolic bottlenecks to further accumulation, excised 5-week-old stems and rosette leaves were exogenously fed different concentrations of <sup>13</sup>C<sub>6</sub>-ring labeled Phe. Both the amount and isotopic enrichment of downstream intermediates was quantified using LC-MS/MS at multiple time points after feeding. A kinetic model of the phenylpropanoid network was constructed, and the parameters were identified through non-linear optimization with training datasets, and validated with data from an independent experiment. In silico analysis of the results from our model predicted that the endogenous cytosolic Phe pools limit the 2-PE production in these transgenic plants. This prediction was tested by combining overexpression of PAR and PAAS with overexpression of a feedback-insensitive 3-deoxy-D-arabino-heptulosonate 7phosphate (DAHP) synthase, the latter of which has been previously shown to have hyper-induced phenylalanine biosynthesis in Arabidopsis. Additionally, increased substrate availability was also tested in the *pal1/pal2* double mutant background combined with the overexpression of PAR and PAAS. These transformations led to a significantly increased accumulation of 2-PE in transgenic

*Arabidopsis*. The use of kinetic modeling combined with time-course *in vivo* metabolite profiling is shown to be a promising approach to rationally engineer plants that accumulate high-value commodity chemicals.

This research is supported by the Office of Biological and Environmental Research in the US Department of Energy.

### Measurement and modeling of phenylpropanoid metabolic flux in Arabidopsis

Peng Wang<sup>1\*</sup> (wang1155@purdue.edu), Longyun Guo,<sup>1</sup> Rohit Jaini,<sup>2</sup> Antje Klempien,<sup>3</sup> Rachel M. McCoy,<sup>4</sup> John Morgan,<sup>1,2</sup> Natalia Dudareva,<sup>1</sup> **Clint Chapple**<sup>1</sup>

<sup>1</sup> Department of Biochemistry, Purdue University, West Lafayette, IN; <sup>2</sup> School of Chemical Engineering, Purdue University, West Lafayette, IN; <sup>3</sup> Department of Agronomy, Purdue University, West Lafayette, IN; <sup>4</sup>Department of Horticulture, Purdue University, West Lafayette, IN

Project Goals: We propose to develop a kinetic model for the shikimate and phenylpropanoid pathways. Kinetic models provide insights into the distribution of flux control, thus permitting more intelligent, predictive and effective design of experiments to modulate fluxes towards pathway end products. For this work, we will compare flux measurements in wild-type Arabidopsis plants to plants that are mutant or down-regulated for genes of the lignin biosynthetic pathway, and, those that have been metabolically engineered to bypass the shikimate dependent branch or direct carbon away from lignin biosynthesis to the production of 2-phenylethanol. The outcomes of our proposed kinetic modeling are to identify what remains unknown about the regulation and control of metabolic fluxes to lignin, and to allow development of strategies and predictions of what targets are the most promising candidates for alteration of metabolic flux to lignin.

Lignin is a complex aromatic polymer that is deposited along with polysaccharides in the plant secondary cell wall. Lignin provides strength and hydrophobicity to plant tissues but impedes the utilization of lignocellulosic biomass for biofuel production. Lignin is derived from the phenylpropanoid pathway, the architecture of which is well understood based upon the biochemical and genetic studies conducted to date. In contrast, we lack a systematic and guantitative view of the factors that control carbon flux into and within this branched metabolic pathway in plants. To explore the control of carbon allocation for phenylalanine and lignin biosynthesis, we have developed a kinetic model of the pathway in Arabidopsis to test the regulatory role of several key enzymes. We first established an experimental system for flux analysis using excised wild-type Arabidopsis stems. We found that excised stems continue to grow and lignify during feeding and showed that distribution of PAL and 4CL activities is consistent with the pattern of lignin deposition. When ring  ${}^{13}C_6$ -labeled phenylalanine ([ ${}^{13}C_6$ ]-Phe) was supplied to excised stems, corresponding isotopologues of a number of intermediates was quantified by LC/MS-MS, and incorporation of [<sup>13</sup>C<sub>6</sub>]-ring labeled monolignols into lignin was demonstrated by DFRC/GC/MS. Using this approach, we analyzed metabolite pool sizes and isotope abundances of the pathway intermediates in a time course from stems fed with [<sup>13</sup>C<sub>6</sub>]-Phe of different concentrations. The maximal activities of PAL and 4CL stayed constant during the feeding processes, while labeled lignin deposition rate increased as more labeled Phe was available. These results suggested that the availability of substrate Phe is one limiting factor for lignin flux in developing stems. In addition, we extended the feeding system to mutants that are defective in *PAL1 PAL2* and *4CL1* to investigate the control of these enzymes on lignin flux. These measurements were all used to develop our kinetic model of the lignin biosynthetic pathway.

This research is supported by the Office of Biological and Environmental Research in the US

Department of Energy.

### Kinetic modeling of the phenylpropanoid pathway in Arabidopsis

Longyun Guo<sup>1</sup>\* (guo165@purdue.edu), Peng Wang<sup>1</sup>, Rohit Jaini<sup>2</sup>, Natalia Dudareva<sup>1</sup>, Clint Chapple<sup>1</sup> and **John Morgan**<sup>1, 2</sup>

<sup>1</sup>Department of Biochemistry, Purdue University, West Lafayette, IN; <sup>2</sup>School of Chemical Engineering, Purdue University, West Lafayette, IN-47907

Project Goals: The project aims to generate a kinetic model of lignin biosynthesis in Arabidopsis to guide rational design for biofuel production. Lignin biosynthesis requires 11 enzyme families functioning together to produce 3 major monolignols from phenylalanine. The metabolic complexity makes it difficult to easily anticipate outcomes of metabolic engineering. A kinetic model is thus proposed to help both understanding and manipulation of this pathway. We are iteratively developing the model with data from *in vivo* substrate feeding to Arabidopsis stems. In parallel, we are also working on introducing an alternative pathway to 2-phenylethanol into Arabidopsis. The final goal is to rewire the fluxes around and downstream of phenylalanine to design an efficient biofuel-producing plant.

Lignocellulosic biomass in plant is an important source for bio-ethanol production, however, its efficient breakdown is limited by the cross-linking property of the lignin polymer in the secondary cell wall. To solve the problem, it is promising to develop transgenic plants with reduced amount of lignin or altered lignin composition. Lignin is mainly derived from pcoumaryl alcohol, coniferyl alcohol and sinapyl alcohol. These alcohols are synthesized in the cytosol from the common precursor phenylalanine. During these metabolic conversions, there are cases where multiple substrates compete for a common enzyme, and where single intermediates are funneled through multiple branch points. This complex metabolic network makes it challenging to rationally propose a genetic manipulation strategy for desired product yield, due to the lack of a global understanding of the pathway. An integrative experimentalmathematical framework was thus proposed to generate an *in vivo* kinetic model for both basic understanding and to serve as an *in silico* platform for manipulation outcome prediction. The Arabidopsis primary stem was selected as the experimental system for modeling lignin formation. Different concentrations of <sup>13</sup>C<sub>6</sub>-ring labeled phenylalanine were fed to excised 5-week-old stems to obtain both the amount and isotopic enrichment of downstream intermediates at multiple time points with LC-MS/MS. Meanwhile averaged lignin deposition rate was estimated from lignin content over development. Since a complete kinetic model is computationally intensive to parameterize, we decided to divide the whole pathway into three modules to train smaller models respectively. A base kinetic model for the first module was initially constructed using Michaelis-Menten kinetics. For the following model refining step, 36 possible metaboliteenzyme regulatory interactions were then systematically explored. Evidence of their existence(s) in vivo was examined with training datasets (0.1, 1 & 3 mM treatments) using Akaike's Information Criteria. The best performance model was then validated with an independent

dataset (0.3 mM treatment). With this workflow, we identified several previously unknown putative metabolite-enzyme interactions, and the current model can capture pathway dynamics over a wide range of feeding treatments. Base models for the other two modules have also been constructed, and similar workflow is planned to apply for model refining as well. The final combined kinetic model can be used to explore *in vivo* metabolic behaviors under different conditions.

This research is supported by the award DE-SC0008628 from the Office of Biological and Environmental Research in the US Department of Energy.

# Identifying the most efficient algal degrading communities: diversity, composition and mechanisms

Xiaoqian Yu<sup>1\*</sup>(<u>xy43@mit.edu</u>), Christopher H. Corzett, <sup>1</sup> Martin F. Polz<sup>1</sup> and Eric J. Alm<sup>1</sup>

<sup>1</sup>Massachusetts Institute of Technology, Cambridge, MA, USA

Project Goals: Marine algae hold great promise for biofuel production and have advantages over terrestrial biomass and freshwater algae. Despite this potential, little effort has been made to date to harness the enzymatic machinery that bacteria use to convert marine algal carbohydrates into bioenergy substrates. Our project capitalizes on this unexplored opportunity via three distinct activities: bioprospecting for novel algal polysaccharide-degrading genes, functional screening for enzymes with desired biochemical properties, and repackaging pathways in reusable genetic modules.

In the marine ecosystem, macroalgae are fast growing primary producers representing half of the carbon production and three-quarters of the biomass in coastal oceans. They are thus a rapidly replenishing resource for bioenergy applications. We are currently studying the degradation of macroalgae by natural microbial communities addressing three principal questions: (i) How does the diversity of a microbial community affect its productivity? (ii) Are there specific microbial communities that are especially productive when growing on algal material? (iii) What are the community compositions of these microbial communities and what are the relevant mechanisms?

We performed a microcosm experiment where marine microbial communities with different diversity and composition were generated from a natural seawater community using a removalof-species-by-dilution method. All communities were supplemented with one of the three following substrates: macroalgal (Fucus) extract, alginate (major structural polysaccharide of Fucus), or mannose (monosaccharide precursor for many sugars in Fucus). For all substrates, we found that the most productive communities were not those with the highest diversity, but those that were dominated by 3-4 species of bacteria after a period of growth during which the more productive species have been selected for. Specifically, for communities supplemented with alginate, we were able to identify a few genera of bacteria that were important predictors of community productivity using a random forest model. The most important predictor genera in the model was Shewanella, known for having versatile electron-accepting capacities that allow them to couple organic matter decomposition to various terminal electron receptors in different environments. All other top predictors belonged to the order Alteromonadales or Oceanospirillales, both known for their abilities to degrade dissolved organic material. We are currently in the process of scaling up the microcosm experiment so we can apply ecological models to identify the relative roles of selection, resource partitioning, and interactions in the most productive communities. We will also perform comparative transcriptomics across communities with high, medium and low productivity to reveal how different organisms could cooperate to degrade algae and form a highly productive community. This will allow us to no longer be restricted to one single organism when packaging metabolic genes into useful genetic modules for algae degradation and biofuel production; the suite of genes can now be picked from and optimized across multiple organisms.

This work is supported by the Office of Biological and Environmental Research in the Department of Energy Office of Science (DE-SC0008743).

### The Ecology of Macroalgal Polysaccharide Utilization: *Verrucomicrobia* Isolates Initiate Fucoidan and Carrageenan Degradation Cascades

Christopher Corzett<sup>1\*</sup> (corzett@mit.edu), Jan-Hendrik Hehemann,<sup>2</sup> Eric Alm,<sup>1</sup> and Martin  $Polz^1$ 

<sup>1</sup> Massachusetts Institute of Technology, Cambridge, MA, USA <sup>2</sup> Max Planck Institute for Marine Microbiology, Bremen, Germany

Project Goals: Marine algae hold great promise for biofuel production and have advantages over terrestrial biomass and freshwater algae. Despite this potential, little effort has been made to date to harness the enzymatic machinery that bacteria use to convert marine algal carbohydrates into bioenergy substrates. Our project capitalizes on this unexplored opportunity via three distinct activities: bioprospecting for novel algal polysaccharide-degrading genes, functional screening for enzymes with desired biochemical properties, and repackaging pathways in reusable genetic modules.

Marine macroalgae are vital players in the global carbon cycle, and polysaccharides represent a significant output of their primary production. Identifying the microbes and metabolic pathways responsible for degrading these sugars is not only crucial to understanding marine carbon flow, but also offers potential for biofuel production using seaweed feedstocks. Enrichment cultures from coastal waters yielded novel Verrucomicrobia isolates capable of degrading the sulfated polysaccharides fucoidan and carrageenan. These fucose- and galactose-based polysaccharides commonly found in brown or red seaweeds, respectively, are often recalcitrant to microbial degradation and require specialized enzymes to degrade. Strains capable of initiating complex breakdown cascades of these sulfated polysaccharides were sequenced, revealing Polysaccharide Utilization Loci (PULs) enriched with numerous and diverse Carbohydrate-Active Enzymes (CAZymes) and sulfatases. Transcriptional analyses revealed specific PULs induced by each macroalgal substrate: one carrageenan-specific PUL (encoding GH16, GH39, GH43 GH82, sulfatases and a TonB-dependant receptor homologue) and three fucoidan-specific PULs (encoding GH29, GH107, sulfatases, hypothetical CAZymes and putative transporters) were identified. CAZymes with potentially novel substrate specificity, particularly among the poorly characterized fucosidases, are the subject of ongoing enzymatic characterization. Culture- and plate-based assays also indicate specific combinations of isolates appear to complement one another and yield greater overall biomass accumulation, suggesting engineered organisms or communities with a full repertoire of enzymatic capabilities may facilitate the efficient conversion of algal biomass.

This work is supported by the Office of Biological and Environmental Research in the Department of Energy Office of Science (DE-SC0008743).

### Characterization of the Alginate Lyases and Laminarinases from Vibrio sp.

Ahmet Badur<sup>1</sup>, Ehar Ammar<sup>1</sup>, Matthew Plutz<sup>1</sup>, Geethika Yalamanchili<sup>1</sup>, Jan-Hendrik Hehemann<sup>3</sup>, Martin Polz<sup>2</sup>, and **Christopher V. Rao**<sup>1\*</sup> (cvrao@illinois.edu)

<sup>1</sup>Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign

<sup>2</sup>Department of Civil and Environmental Engineering, Massachusetts Institute of Technology <sup>3</sup>Center for Marine Environmental Sciences, University of Bremen

Project Goals: This project will harvest 'biomass to biofuel' pathways from algaeassociated bacteria, and develop these as reusable genetic parts. Marine algae hold great promise for biofuel production and have advantages over terrestrial biomass and freshwater algae. Despite this potential, little effort has been made to date to harness the enzymatic machinery that bacteria use to convert marine algal carbohydrates into bioenergy substrates. Our project capitalizes on this unexplored opportunity via three distinct activities: bioprospecting for novel algal polysaccharide-degrading genes, functional screening for enzymes with desired biochemical properties, and repackaging pathways in reusable genetic modules.

Brown seaweeds are an attractive source of feedstocks for biofuel production, since they have advantages over terrestrial feedstocks. Brown seaweeds have higher growth rates than terrestrial plants, and they lack crystalline cellulose and lignin. Additionally, brown seaweeds do not impinge on arable land, thus negating the conflict between food and fuel. Two of the primary components of brown seaweeds are alginate and laminarin. Alginate is a copolymer consisting of 1,4 linked epimers  $\alpha$ - L-guluronate (G) and  $\beta$ -D-mannuronate (M). The local structure of alginate can take one of three forms: short stretches of polyguluronate (polyG), short stretches of polymannuronate (polyM), or alternating sequences of guluronate and mannuronate. The enzymes that can degrade the linkages within alginate are called alginate lyases. Alginate lyases are classified based on their specific dyad G-G (EC 4.2.2.11), M-M (EC 4.2.2.3), and M-G/G-M bonds that they cleave. Additionally, alginate lyases are classified based on whether they have exolytic or endolytic cleavage. Laminarin is a polysaccharide consisting of  $\beta$ -1,3 and  $\beta$ -1,6 linked glucose. The enzymes that can degrade these linkages are called glycoside hydrolases (GHs). More specifically, the  $\beta$ -1,3 linkage is degraded by enzymes belonging to seven GH families: GH3, GH5, GH16, GH17, GH55, GH64, and GH81. β-1,6 degrading GHs are remain unknown.

We are investigating the mechanism of alginate metabolism within marine *Vibrio sp.* We previously characterized the alginate lyases in *V. spendidus* 12B01 [1]. A recent study investigated the ability of different marine *Vibrionaceae* bacteria to degrade alginate [2]. They identified significantly variability within two closely related *Vibrio splendidus* strains, 12B01 and 13B01. In particular, they found that *V. splendidus* 13B01 has significantly higher secreted alginate lyase activity than *V. splendidus* 12B01. To determine the source of this variability, we characterized the six alginate lyases in *V. splendidus* 13B01 using a combination of genomics, proteomics biochemical, and functional screening. These experiments revealed that a single alginate lyase PL7G, unique to *V. splendidus* 13B01, is critical for rapid extracellular alginate degradation.

We are also investigating the mechanism of laminarin degradation in *Vibrio breoganii* 1C10. This bacterium contains four putative laminarinases: LamA, LamB, LamC, and LamD. We cloned, purified, and enzymatically characterized these laminarinase. We also determined the specificity and endolytic/exolytic activity using NMR and MALDI-TOF spectrometry analysis. The transglycosylation ability of these lamarinases and the extent of the hydrolysis were also examined. These results now allow for metabolic engineering of microorganisms that degrade laminarin as their sole carbon source.

- 1. Badur, A.H., S.S. Jagtap, G. Yalamanchili, J.K. Lee, H. Zhao, and C.V. Rao, *Alginate Lyases from Alginate-Degrading Vibrio splendidus 12B01 Are Endolytic.* Appl Environ Microbiol, 2015.
- 2. Hehemann, J.H., P. Arevalo, M.S. Datta, X. Yu, C.H. Corzett, A. Henschel, S.P. Preheim, S. Timberlake, E.J. Alm, and M.F. Polz, *Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes.* Nat Commun, 2016. **7**: p. 12860.

This project is a part of the Biosystems Design Program supported by the Office of Biological and Environmental Research in the DOE Office of Science

### **Optimization of Alginate Utilization in Engineered Bacteria for Biofuels Production**

Jonathan Ning<sup>1,2\*</sup> (jning2@illinois.edu), Eva Garcia-Ruiz<sup>1,2</sup>, Sujit S. Jagtap<sup>1</sup> and Huimin Zhao<sup>1,2</sup>

<sup>1</sup>Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Illinois; <sup>2</sup>Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Illinois

### http://scs.illinois.edu/~zhaogrp/

Project Goals: This project is focused on unveiling pathways from algae-associated bacteria and refactoring them to use algal polysaccharide as a feedstock for biofuel production. Algal polysaccharides are considered a promising carbon/energy source and are emerging as an important feedstock for the production of biofuels. Despite this potential, little effort has been made to date to harness the enzymatic machinery that bacteria use to convert marine algal carbohydrates into bioenergy substrates. This project harnesses the unexplored bacterial polysaccharide-degrading pathways to 1) bioprospect novel algal polysaccharide-degrading genes, 2) characterize enzymes with desired biochemical properties, and 3) repackage pathways in reusable genetic modules. This project will yield a set of functional modules for the producing biofuels from marine macroalgae.

Marine macroalgae is emerging as an attractive feedstock for biofuels production. A number of marine microbes are able to degrade and catabolize efficiently macroalgal polysaccharides by specialized enzymatic pathways which convert these carbohydrates into bioenergy substrates. *Vibrio splendidus* is a marine bacterium capable of degrading and catabolizing alginate (a linear copolymer of two uronic acids:  $\beta$ -D-mannuronate and  $\alpha$ -L-guluronate) by specialized enzymatic pathways. In this work, we harnessed the alginate-degrading machinery from *V. splendidus* via heterologous expression in the highly genetically amenable host *Escherichia coli*. The alginate-degrading pathway in *V. splendidus* is clustered in two separated fragments of DNA that contain a set of genes for alginate transport and metabolism. This cluster of ~49 kb was assembled using the DNA assembler method and expressed on a fosmid in *E. coli* ATCC 8739. The resulting strain was able to grow on minimal media with alginate and oligoalginates as the sole carbon sources, albeit at a slow rate. To increase utilization rate of alginate, selected key *V. splendidus* pathway genes and homologs were overexpressed to identify degradation pathway bottlenecks and increase utilization rate. The resulting strain can be used as a basis for further optimization efforts using pathway-scale and genome-scale methods.

This study was supported by the U.S. Department of Energy DE-SC0008743.

## Nitrogen limitation and lipid production in Yarrowia lipolytica

Erin L. Bredeweg, Kyle R. Pomraning, Ziyu Dai, Scott E. Baker\* (scott.baker@pnnl.gov)

Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

# Project Goal: Our goal is to enable rapid engineering of the oleaginous yeast, *Yarrowia lipolytica*, for increased rates of biofuel production.

*Yarrowia lipolytica* is an oleaginous yeast with strong potential as a production host for biofuels and bioproducts. *Yarrowia* produces significant amounts of lipids, including triacylglycerols when nitrogen is a limiting nutrient. We have characterized system level metabolic and regulatory changes that occur upon nitrogen limitation using systems and cell biological approaches. These studies include the development of a green fluorescent protein (GFP) based "cell organelle atlas" to characterize the cellular changes that occur upon nitrogen limitation in *Y. lipolytica*. Genome-wide promoter analysis predicts that the ortholog of the carbon catabolite repression regulator, mig1/cre1/creA, as well as nitrogen responsive GATA transcription factors are key controllers of the cellular response to carbon and nitrogen levels. The research we will present includes a genetic and transcriptome analysis of the GATA transcription factors and their regulatory targets.

### **Publications**

- 1. Pomraning, K.R., Bredeweg, E.L., Baker, S.E. Regulation of nitrogen metabolism by GATA zinc-finger transcription factors in *Yarrowia lipolytica*. *Submitted*.
- 2. Bredeweg, E.L., Pomraning, K.R., Dai, Z., Nielsen, J. Kerkhoven, E.J., Baker, S.E. 2016. A molecular genetic toolbox for *Yarrowia lipolytica*. Biotech. Biofuels. *In press*.
- Pomraning, K.R., Kim, Y.-M., Nicora, C.D., Chu, R.K., Bredeweg, E.L., Purvine, S.O., Hu, D., Metz, T.O., Baker, S.E. 2016. Multi-omics analyses reveals regulators of the response to nitrogen limitation in *Yarrowia lipolytica*. BMC Genomics 17:138.
- 4. Kerkhoven, E. Pomraning, K.R., Baker, S.E., Nielsen, J. 2016. Regulation of amino acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. npj Systems Biology and Applications. 2:16005.
- 5. Pomraning, K., Baker, S.E. 2015. Draft genome sequence of the dimorphic yeast *Yarrowia lipolytica* strain W29. Genome Announc. 3(6). pii: e01211-15.
- Pomraning, K.R., Wei, S., Karagiosis, S.A., Kim, Y.M., Dohnalkova, A.C., Arey, B.W., Bredeweg, E.L., Orr, G., Metz, T.O., Baker, S.E. 2015. Comprehensive metabolomic, lipidomic and microscopic profiling of *Yarrowia lipolytica* during lipid accumulation identifies targets for increased lipogenesis. PLoS One. 10(4):e0123188.

*Funding for this research provided by the DOE BER Genome Science Program grant DE-SC0008744.* 

# Leucine Biosynthesis is Involved in Regulating High Lipid Accumulation in *Yarrowia lipolytica*

Eduard J. Kerkhoven<sup>\*1</sup> (eduardk@chalmers.se), Young-Mo Kim<sup>2</sup>, Siwei Wei<sup>2</sup>, Carrie D. Nicora<sup>2</sup>, Thomas L. Fillmore<sup>2</sup>, Samuel O. Purvine<sup>2</sup>, Bobbie-Jo Webb-Robertson<sup>3</sup>, Richard D. Smith<sup>2</sup>, Scott E. Baker<sup>2</sup>, Thomas O. Metz<sup>2</sup>, **Jens Nielsen**<sup>1</sup>

<sup>1</sup> Chalmers University of Technology, Gothenburg, Sweden; <sup>2</sup> Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA; <sup>3</sup> National Security Directorate, Pacific Northwest National Laboratory, Richland, WA

# Project Goals: Our goal is to elucidate the regulation of lipid metabolism in *Yarrowia lipolytica* to identify new targets to improve the TAG yield.

The yeast *Yarrowia lipolytica* is a potent accumulator of lipids and lipogenesis in this organism can be influenced by a variety of factors, such as genetics and environmental conditions. We have previously identified the involvement of the regulation of amino-acid metabolism to redirect flux to lipid accumulation [1]. We have followed up on this using a multifactorial study, thereby elucidating the effects of both genetic and environmental factors on regulation of lipogenesis in *Y. lipolytica* and identifying how two opposite regulatory states both result in lipid accumulation.

The study involved the comparison of a strain overexpressing diacylglycerol acyltransferase (DGA1) with a control strain grown under either nitrogen or carbon limited conditions. A strong correlation was observed between the responses on the levels of transcript and protein. Combination of DGA1 overexpression with nitrogen limitation resulted in high-level of lipid accumulation accompanied by downregulation of various amino acid biosynthetic pathways, with leucine in particular, and these changes were further correlated with a decrease in metabolic fluxes.

The downregulation in leucine biosynthetic pathway genes was supported by a measured decrease in the level of 2-isopropylmalate, an intermediate of leucine biosynthesis. Combining the multi-omics data with putative transcription factor binding motifs uncovered a contradictory role for TORC1 in controlling lipid accumulation, likely mediated through 2-isopropylmalate and a Leu3-like transcription factor.

### References

1. Kerkhoven EJ, Pomraning KR, Baker SE and Nielsen J. (2016) Regulation of amino-acid metabolism controls flux to lipid accumulation in Yarrowia lipolytica. npj Syst Biol App. doi:10.1038/npjsba.2016.5

Funded by U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, Genomic Science program (GSP), under Award Number DE-SC0008744. PNNL is operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. Proteomics analysis was supported by the GSP-funded Pan-omics Program at PNNL.

### Lipid Accumulation and its Impact on Amino Acid Metabolism in Saccharomyces cerevisiae

Michael Gossing\*1 (gossing@chalmers.se) and Jens Nielsen1

<sup>1</sup>Chalmers University of Technology, Gothenburg, Sweden

Project Goals: Lipids are a group of highly diverse molecules with a multitude of biological functions such as formation of biological membranes, storage of energy, cell signaling, and apoptosis. Triacylglycerides (TAG) function as energy storage and source of membrane building blocks. *Saccharomyces cerevisiae* was metabolically engineered to accumulate increased levels of TAG. We observed that redirecting metabolic flux towards formation of TAG affected amino acid metabolism on the level of transcript and metabolite.

We have engineered *Saccharomyces cerevisiae* to accumulate increased levels of TAG by introducing a push and pull on TAG biosynthesis. A push was introduced by overexpression of acetyl-CoA carboxylase double mutant *ACC1<sup>S659A S1157A</sup>* (*ACC1\*\**), while a pull was introduced by overexpression of phosphatidate phosphatase *PAH1* and diacylglycerol acyltransferase *DGA1*. The resulting strain was analyzed for changes in level of transcript and amino acids compared to a reference strain in batch cultivation during respiratory growth.

We observed a reduced abundance of total free amino acids in the engineered strain. This reduction was consistent with an observed downregulation of several amino acid biosynthetic genes. However, while all amino acids showed a lower absolute abundance (with the exception of aspartic acid), some amino acids showed a higher relative abundance in the total amino acid pool compared to the reference strain. Interestingly, a higher relative abundance was found for the branched-chain amino acids leucine, isoleucine and valine. Branched-chain amino acid metabolism in general, and leucine metabolism in particular, has been linked to lipid metabolism by several studies. A higher relative abundance was also found for the aromatic amino acids phenylalanine, tryptophan and tyrosine.

The interaction of lipid metabolism and amino acid metabolism highlights the complexity of the metabolic network in *Saccharomyces cerevisiae*. Studying this interaction in more detail will lead to more information about how these pathways relay information, and will lead to identification of metabolites and proteins involved in signaling.

Funded by U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, Genomic Science program (GSP), under Award Number DE-SC0008744.
#### **Poster Abstract**

## Optimal feed control for maximization of lipid production from dilute acetic acid by an engineered oleaginous yeast *Yarrowia lipolytica*

Jingyang Xu (xujy@mit.edu) and Gregory Stephanopoulos Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Room 56-422, Cambridge, MA

Acetic acid can be generated in syngas fermentation, lignocellulosic biomass degradation, and anaerobic digestion of organic wastes. Compared to sugar-based lipid production, microbial conversion of acetate into triacylglycerols for biofuel production can be economically more affordable due to low- or negative-cost feedstocks. The main issue is the dilute nature of acetate produced in such systems, a problem we tackled with cell recycle. To this end, we established an efficient process for converting dilute acetate into lipids by the oleaginous yeast, *Yarrowia lipolytica* engineered for lipid overproduction. The process design utilized low-strength acetic acid as substrate with dual feed of acetic acid and acetate, as well as nitrogen for cell growth with a cross filtration module for cell recycle. Various types of feed control were deployed including one based on a fermentation model and on-line measurement of the Respiratory Quotient. The optimized process was able to utilize 3% acetic acid for high density cell culture and achieved a lipid titer of 115 g/L, lipid conversion yield of 0.16g/g and an overall productivity of 0.8 g L<sup>-1</sup> h<sup>-1</sup>.

### Rewiring metabolism for maximum lipid production in oleaginous yeast Yarrowia lipolytica

Kangjian Qiao<sup>1</sup> (kjqiao@mit.edu) and Gregory Stephanopoulos<sup>1</sup>

<sup>1</sup>Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA

Project Goals: We focused on achieving a fundamental understanding of the metabolism of the oleaginous yeast *Yarrowia lipolytica* and developing tools to characterize and engineer it towards cost-effective lipids production. More specifically, we aimed to improve its fermentation characteristics towards the development of a cost-effective process which converts renewable resources to lipids for biodiesel production. The conversion yield and volumetric productivity on various carbon sources are the key metrics for optimization.

Production of lipids by microbial fermentation of carbohydrate feedstocks outcompetes oil crops in terms of productivity, however, presently achievable carbohydrates-to-lipids process yields are not yet at a point that can support cost-effective production of lipids and biodiesel. To maximize process yields, one needs to maximize lipid content as well as capture as many of the electrons generated from the catabolism of the available substrate as possible. We show that overall lipid process yield, Y, is significantly improved via introduction of synthetic pathways that effectively recycle glycolytic NADHs into cytosolic NADPH and acetyl-CoA to be used for lipid synthesis. Strain construction is guided by a quantitative model that predicts Y from the non-lipids biomass yield, lipid content, and yield of lipid synthesis pathway, Y<sub>L</sub>, with Y<sub>L</sub> becoming the key driver of process yield maximization at high lipid content. In total, thirteen rationally designed strain constructs were evaluated in shake flask and bioreactor experiments to identify the best strain that achieved a lipid titer of close to 100 g/L with a productivity of 1.2 g/L/h and an overall process yield 0.27 g FAME/g glucose. These figures of merit advance the commercialization opportunities of carbohydrate-based biodiesel production.

### Seteria viridis as a Model System to Accelerate Gene Discovery in Panicoid Grasses

Pu Huang<sup>1</sup>, Christine Shyu<sup>1</sup>, Hui Jiang<sup>1</sup>, Kerrie Barry<sup>2</sup>, Jerry Jenkins<sup>3</sup>, Jeremy Schmutz<sup>2,3</sup>, Mathew S. Box<sup>1</sup>, Chuanmei Zhu<sup>1</sup>, Allison Huskey<sup>1</sup>, Dustin Mayfield-Jones<sup>1</sup>, Xiaoping Li<sup>1</sup>, Elizabeth A. Kellogg<sup>1</sup>, and **Thomas P. Brutnell**<sup>1</sup>

<sup>1</sup>Donald Danforth Plant Science Center, Saint Louis, MO 63132

<sup>2</sup>Department of Energy Joint Genome Institute, Walnut Creek, CA 94598

<sup>3</sup>HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806

# Project Goals: This project aims to develop genetic resources and tools in *Setaria viridis* for gene identification and functional characterization. Outcomes from this research include the identification of genes that can be manipulated in target feedstocks for the improvement of agronomic traits including plant architecture, disease/pest resistance and flowering time. We are also developing new genome editing technologies and methods of genetic analysis to accelerate gene discovery and genome engineering efforts.

Setaria viridis is a C<sub>4</sub> panicoid grass with a diploid genome, rapid life cycle, small stature and high seed production that together facilitate genetic analyses. It serves as an ideal model system to accelerate gene discovery in bioenergy feedstocks including maize, sorghum, Miscanthus, switchgrass and sugarcane. Multiple forward genetics, reverse genetics, as well as high throughput phenotyping resources have been generated in S. viridis, including a chemically mutagenized population of over 15,000 families and a diverse panel of ~600 wild accessions. To demonstrate the utility of the system, we have conducted several reverse and forward genetic screens. In a forward genetic screen, we identified SvAUX1 as a regulator of inflorescence branching and gravitropism in S. viridis. We identified four single gene recessive sparse panicle phenotypes characterized by reduced and uneven branching of the inflorescence from approximately 2700 M2 families. A bulked segregant analysis was performed to identify the gene underlying the sparse panicle1 (spp1) phenotype, and spp1 was fine mapped to a ~1 Mb interval. Through complementation tests and deep sequencing of another mutant, spp3, a causal gene SvAUX1 was identified. This gene is one of the two genes in the ~1Mb interval and the only gene disruption shared between *spp1* and *spp3*. We further show that the maize ortholog of SvAUX1, ZmAUX1 plays similar role in inflorescence and root development, highlighting the utility of S. viridis in accelerating functional genomic studies in maize. In a second forward genetic screen, we identified a candidate gene (SvCO-like 1) that may control flowering time, a key trait for bioenergy feedstocks. This gene was found through bulked segregant analysis in a prolonged flowering mutant family. Further functional characterization of SvCO-like1 is currently ongoing. Through reverse genetics, we are characterizing the jasmonate (JA) signaling pathway in S. viridis that contributes to growth and defense responses that must be optimized for high yielding feedstocks that are resistant to pest pressures. In grasses, there are three genes encoding the jasmonate receptor CORONATINE INSENSITIVE (COI). Using CRISPR-Cas9

technology, we generated mutants in *COIIb*, and identified three independent alleles that exhibit an early flowering phenotype, similar to what has been described in *Arabidopsis* where only a single *COI* gene is present. Additional screens to identify *COIIa* and *COI2* mutants are underway to dissect *COI* gene function and gain deeper insight into the role of JA in regulating growth and defense responses in *S. viridis*. Taken together, multiple toolsets have been developed in *S. viridis* for rapid gene candidate identification and functional characterization, which will accelerate functional genomics studies in panicoid grasses.

### **Grasses Suppress Shoot-Borne Roots to Conserve Water During Drought**

Jose Sebastian<sup>1,#</sup>, Muh-Ching Yee<sup>1,#</sup>, Willian Goudinho Viana<sup>1,2</sup>, Rubén Rellán-Álvarez<sup>1,3</sup>, Max Feldman<sup>4</sup>, Henry Priest<sup>4</sup>, Charlotte Trontin<sup>1</sup>, Tak Lee<sup>5</sup>, Hui Jiang<sup>4</sup>, Ivan Baxter<sup>4,6</sup>, Todd C. Mockler<sup>4</sup>, Frank Hochholdinger<sup>7</sup>, José R. Dinneny<sup>1,\*</sup> (jdinneny@carnegiescience.edu) and **Thomas P. Brutnell<sup>4</sup>** 

<sup>1</sup>Carnegie Institution for Science, Department of Plant Biology, Stanford, CA, 94305, <sup>2</sup>CAPES Foundation, Ministry of Education of Brazil, Brasilia – DF, Zip Code 70.040-020, <sup>3</sup>Unidad de Genómica Avanzada, Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Mexico, <sup>4</sup>Danforth Plant Science Center, St. Louis, MO, <sup>5</sup>Department of Biotechnology, College of Life Sciences and Biotechnology, Yonsei University, Seoul, Korea, <sup>6</sup>USDA-ARS Plant Genetics Research Unit, Donald Danforth Plant Sciences Center, <sup>7</sup>University of Bonn, INRES - Crop Functional Genomics, Friedrich-Ebert-Allee 14453113 Bonn, Germany, <sup>#</sup> These authors contributed equally

### http://foxmillet.org/ and https://dinnenylab.me/

### **Project Goals:**

- Define the developmental consequences of drought on the root system of Setaria and related grasses.
- Determine the physiological significance of stress-mediated changes in root architecture.
- Develop a method for quantifying root system architecture under water deficit stress in soil-like conditions.
- Identify the signaling pathways that are regulated during drought that likely affect root development.

Abstract: Many important crops are members of the Poaceae family, which develop root systems characterized by a high-degree of root initiation from the belowground basal nodes of the shoot, termed the crown (1, 2). While this post-embryonic shoot-borne root system represents the major conduit for water uptake, little is known regarding what effect water availability has on its development. Here we demonstrate that in the model C<sub>4</sub> grass *Setaria viridis*, the crown locally senses water availability and suppresses post-emergence crown root growth under water deficit (1). This response was observed in field and growth room environments and in all grass species tested. Luminescence-based imaging of root systems grown in soil-like media revealed a shift in root growth from crown to primary-root derived branches, suggesting that primary-root-dominated architecture can be induced in *S. viridis* under certain stress conditions (1, 3). Crown roots of *Zea mays* (maize) and *Setaria italica*, domesticated relatives of teosinte and *S. viridis*, respectively, show reduced sensitivity to water deficit, suggesting that this response may have been influenced by human selection. Enhanced water status of maize mutants lacking crown roots suggests that, under water deficit, stronger suppression of crown roots may actually benefit crop productivity.

Significance Statement: Grasses, whose members constitute key food and bioenergy crops worldwide, utilize unique developmental programs to establish the root system from the shoot. Shoot-borne crown roots, originate near the soil surface and provide the main conduits by which the plant takes up water and nutrients. We show that crown root development is the major target of drought stress signaling. Water-deficit triggered crown-root arrest provides an important mechanism to conserve water under drought and this response is widely conserved across grass species. Substantial phenotypic variation exists in maize for this trait, which may be useful target in breeding efforts to improve drought tolerance.

### References

1) Sebastian J, Yee MC, Viana WG, Rellán-Álvarez R, Feldman M, Priest H, Trontin C, Lee T, Jiang H, Baxter I, Mockler TC, Hochholdinger F, Brutnell TP and Dinneny JR (2016) Grasses suppress shoot-borne roots to conserve water during drought. *Proc. Natl. Acad. Sci. U. S. A.* 8861–8866, doi: 10.1073/pnas.1604021113

2) Sebastian J and **Dinneny JR** (in press) *Setaria viridis*: a Genetic Model System for Panicoideae Grass Root Systems Setaria, Springer

3) Rellán-Álvarez R, Lobet G, Lindner H, Pradier P-L, Sebastian J, Yee MC, Geng Y, Trontin C, LaRue T, Schrager A, Haney C, Nieu R, Maloof J, Vogel JP, Dinneny JR (2015) GLO-Roots: an imaging platform enabling multidimensional characterization of soil-grown root systems. *eLife* 2015;10.7554/eLife.07597

*Funding is provided by a grant from the Department of Energy Biological and Environmental Research program (#DE-SC0008769) to T.P.B., T.C.M., I.B. and J.R.D.* 



### Development and application of novel phenotyping techniques to understand the genetic control of productivity and drought traits in the model C4 grass Setaria

Andrew D.B. Leakey<sup>1</sup>\* (leakey@illinois.edu), Darshi Banan<sup>1</sup>, Rachel Paul<sup>1</sup>, Parthiban Prakash<sup>1</sup>, Luke Freyfogle<sup>1</sup>, Maximilian Feldman<sup>2</sup>, Nathan Miller<sup>3</sup>, Edgar Spalding<sup>3</sup>, Ivan Baxter<sup>2,4</sup> and **Thomas P. Brutnell<sup>2</sup>** 

<sup>1</sup>University of Illinois, Urbana-Champaign; <sup>2</sup>Donald Danforth Plant Science Center, St Louis; <sup>3</sup>University of Wisconsin, Madison; and <sup>4</sup>USDA-ARS, St Louis.

### http://foxmillet.org/

Genetically tractable model systems closely related to bioenergy grasses need to be developed to drive the crop improvement required for large scale, ecologically sustainable bioenergy production. *Setaria viridis* is an ideal candidate C4 panacoid grass. The overarching objectives of this large, collaborative project are to utilize genomic, computational and engineering tools to begin the genetic dissection of drought response in *S. viridis*. This will be achieved through: 1) Quantitative trait and association genetics; 2) novel controlled environment and field phenotyping combined with molecular and chemical profiling; 3) development of metabolic and gene networks; 4) development of transformation technologies; 5) reverse genetic testing of candidate genes.

Our ability to cheaply and quickly phenotype large mapping populations of C4 grass crops for complex traits related to productivity and drought tolerance severely limits efforts to understand genotype-to-phenotype associations under field conditions. Here we report the development and application of methods to assess: (1) above-ground biomass production from hemispherical imaging; (2) stomatal patterning from optical tomography; (3) leaf nitrogen status and allometry from hyperspectral reflectance; (4) drought-induced leaf curling from hemispherical imaging; and (5) canopy temperature by infra-red imaging as a proxy for crop water use. We demonstrate that these methods successfully capture the same genotype by environment interactions and reproduce quantitative trait loci analyses as traditional methods that are slower and more expensive. The combination of these advances in phenotyping capability and new knowledge of the genetic architecture of productivity and drought traits creates a research platform that can now be applied to biosystems design of more productive and ecologically sustainable biofuel and bioproduct crops.

This research was funded through Subaward No. 23009-UI, CFDA # 81.049 between University of Illinois and Donald Danforth Plant Science Center Under Prime Agreement No. DE-SC0008769 from Department of Energy.

### Biocontainment in Perennial Grasses: Transgenic Pollen Ablation in *Brachypodium* sylvaticum and *Panicum virgatum*.

**Roger Thilmony**<sup>1\*</sup> (<u>Roger.Thilmony@ars.usda.gov</u>), Bryan Tarape Hernandez<sup>1</sup>, Ray Collier<sup>1</sup>, Jonathan Willis<sup>1</sup>, **Christian Tobias**<sup>1</sup>, **John Vogel**<sup>2</sup>, **Eduardo Blumwald**<sup>3</sup>

<sup>1</sup>USDA-ARS-WRRC, Crop Improvement and Genetics Research Unit, Albany, CA <sup>2</sup>DOE Joint Genome Institute, Walnut Creek, CA <sup>3</sup>Department of Plant Sciences, University of California, Davis, CA

### Project Goal: Develop a gene containment system to minimize transgene flow from transgenic switchgrass.

Although numerous characteristics make switchgrass a desirable biofuels crop, it is also a windpollinated species native to the region where it will be grown. Agricultural cultivation of transgenic switchgrass cultivars with improved traits raises the possibility that engineered traits may move into natural populations with potential unintended environmental impacts. Thus, a means to control transgene escape to wild plant populations is needed prior to commercial release. We are developing a transgene containment system for perennial grasses to address this concern. The ability of novel transformation constructs to block transmission of transgenes via pollen is being evaluated in the model perennial grass Brachypodium sylvaticum and switchgrass. Using Agrobacterium-mediated transformation, we generated transgenic plants that express Barnase (a toxic ribonuclease) under the control of four rice pollen-specific promoters (PS1, PS2, PS3 and OsGEX2). Independent transgenic lines for each construct are being evaluated by pollen staining and/or genetic segregation analyses. Alexander's staining revealed that, relative to wildtype plants, >50% of the pollen collected from the T<sub>0</sub> transgenic plants was dead or deformed. Analysis of selfed T<sub>1</sub> B. sylvaticum progeny showed that transgene heritability was 1:1, consistent with the expected segregation frequency for a male lethal gene, supporting the conclusion that successful ablation of transgenic pollen was achieved in these transgenic plants. We are currently characterizing multiple independent transgenic switchgrass plants carrying the ablation constructs for their effectiveness at mediating transgene containment. Whether the biocontainment constructs alter plant morphology or biomass yield is also being examined.

The project is supported by a U.S. Department of Energy, Office of Science, Biological and Environmental Research, Biosystems Design to Enable Next-Generation Biofuels and Bioproducts grant (DESC0008797).

### Genome Editing of CENH3 in Switchgrass and Brachypodium: A Histone Variant Essential for Centromere Specification

Sangwoong Yoon<sup>1</sup>, Sheyla Aucar<sup>1</sup>, James Horstman<sup>2</sup>, **Eduardo Blumwald**<sup>2</sup>, **Roger Thilmony**<sup>1</sup>, **John Vogel**<sup>3</sup>, **Christian Tobias**<sup>\*1</sup> (christian.tobias@ars.usda.gov)

<sup>1</sup>USDA-ARS-WRRC, Crop Improvement and Genetics Research Unit, 800 Buchanan Street, Albany, CA 94710, USA

<sup>2</sup>Dept. of Plant Sciences, University of California, One Shields Ave, Davis, CA 95616, USA. <sup>3</sup>DOE Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598, USA

### **Project Goals:**

Develop a CENH3-based method for creating doubled haploid grasses and demonstrate it in switchgrass.

### Background

Self-incompatibility and outcrossing behavior in switchgrass prevent effective hybrid development and fixation of gene combinations. This problem is being addressed through biotechnological approaches enabling creation of haploid inducer lines use with doubled haploid breeding strategies. In order to engineer haploid inducer lines in grasses via centromere-mediated genome elimination, we are introducing mutations into both switchgrass and Brachypodium CenH3 genes by genome editing techniques. These mutations could then be rescued or complemented with CenH3 variants similar to those known to induce uniparental genome elimination (1). These as well as genome editing of the COMT gene of Brachypodium that is involved in monolignol synthesis are underway. In order to determine if genome doubling techniques are efficient a population of neo-octaploid `Liberty` switchgrass was also created.

Transformations of switchgrass and Brachypodium with genome editing constructs using both TALEN and CRISPR/Cas9 systems have been conducted and switchgrass lines are being crossed with wild type plants to segregate away the transgene from the desired mutation. With these constructs, target-endonuclease activity was validated either by employing heterologous yeast system, or a GUS reporter gene containing target sequences that interrupted the ORF. GUS activity was detected in onion epidermis after cobombardent of the reporter gene and the binary plant transformation vector. Activity in yeast was detected using a  $\beta$ -galactosidase reporter gene. These assays indicated that target-endonuclase activity could restore an active ORF and that the CRISPR binary vectors were functional at least a transient plant-assay system.

### Brachypodium CenH3 TILLING mutants:

Working with a Brachypodium TILLING resources at INRA, France and a screening center managed there, efforts were made at mutation screening within the CenH3 gene. Ten lines were isolated, all with point mutations in CenH3 that have now been validated by Sanger sequencing. One of these families (5173) contains a predicted Ser to Phe change at position 82 and has reduced stature and fertility. We are now backcrossing this line to WT Bd21-3 to reduce background mutations and analyze further. Other online Brachypodium resources including T-DNA insertion populations and mutant resequencing projects were screened as well but did not yield mutations in CenH3.

### Brachypodium TALEN line A/B 1 is active against it's BdiCenH3 target

We have focused most of our efforts so far on a Brachypodium line that contains a very active TALEN T-DNA as judged by pooled sequencing approaches as well as high resolution melt analysis, and Sanger sequencing. High resolution melt analysis has proved as efficient in our hands as gel-based mutation screening and we have screened for T-DNA presence using *Fok1* and *HptII*-specific primers. We have not isolated any mutations in *CenH3* in the absence of the TALEN T-DNA. Our hypothesis now is that most mutants in BdiCenH3 are embryo-lethal and we have since retransformed this line with altered versions of CenH3 in an effort to obtain viable knockout/substitution lines.

### Switchgrass neo-octoploid lines

Seedling treatment to induce polyploidy could be effective for breeding purposes where genetic exchange between different cytotypes is desired. In this study the cultivar 'Liberty' was chosen because it is of recent hybrid upland/lowland origin, has superior yields, and is adapted to hardiness zones that are considered to be suitable for upland cultivars which are frequently octaploid.

Initial lines were largely cytochimeras containing 4x and 8x sectors (2). Individual octaploid tillers were identified by flow cytometry and allowed to intercross to generate 100% octaploid individuals. Seed harvests indicated that seed derived from octaploid sectors was 19% larger compared to tetraploid sectors and that fertility in these sectors was significantly lower. Non-chimeric octaploid progeny of the treated individuals were confirmed to have approximately twice the number of chromosomes of tetraploid lines although there was high variability in these counts. Induced octaploids had larger pollen and leaf cell size. Guard cell density was found to be significantly lower in induced octaploid compared to tetraploid individuals. An earlier publication found no substantial differences in guard cell density in some naturally occurring octaploid and tetraploid switchgrass populations. Our results showing significant differences should have implications with respect to gas exchange rates and photosynthesis.

This population has been clonally propagated and sent to ARS collaborators at the University of Lincoln, Nebraska. Intercrosses with other octaploid populations are being conducted there as is further phenotypic analysis. One potential benefit is that the larger seed size will improve seedling vigor and stand establishment, as these traits are strongly correlated in switchgrass and other perennial grasses.

### References

- 1. Ravi M, Chan SWL (2010) Haploid plants produced by centromere-mediated genome elimination. Nature 464:615–618. doi: 10.1038/nature08842
- 2. Yoon S, Aucar S, Hernlem BJ, et al (2016) Generation of Octaploid Switchgrass by Seedling Treatment with Mitotic Inhibitors. Bioenerg Res 1–9. doi: 10.1007/s12155-016-9795-2

This material is based upon work supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0008797. This work was also supported by the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) Current Research Information System (CRIS) 2030-21000-023.

### Leveraging *Agave* and *Kalanchoë* Genomics Resources to Transfer Crassulacean Acid Metabolism (CAM) Modules into C<sub>3</sub> Species Using Synthetic Biology Approaches

**Xiaohan Yang**<sup>1</sup>(<u>yangx@ornl.gov</u>), Rongbin Hu<sup>1</sup>, Degao Liu<sup>1</sup>, Henrique Cestari De Paoli, Paul E. Abraham<sup>2</sup>, John Cushman<sup>3</sup>, Anne M. Borland<sup>1,4</sup>, Gerald Tuskan<sup>1</sup>

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; <sup>2</sup>Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN <sup>3</sup>Department of Biochemistry and Molecular Biology, University of Nevada, Reno, NV <sup>4</sup>School of Biology, Newcastle University, Newcastle, UK

### http://cambiodesign.org/

Project Goals: Crassulacean acid metabolism (CAM) is a specialized mode of photosynthesis that features a temporal CO<sub>2</sub> pump with nocturnal CO<sub>2</sub> uptake, facilitates increased water-use efficiency (WUE), and enables CAM plants to inhabit water-limited semi-arid or seasonally dry environments. CAM provides an excellent opportunity for engineering both enhanced WUE and photosynthetic performance into bioenergy crops. This project has two main goals: 1) to identify the CAM-associated genes and gene networks using systems biology approaches and 2) to engineer CAM gene modules into C<sub>3</sub> species using synthetic biology approaches. The success of the project could allow biomass production on semi-arid, abandoned, or marginal agricultural lands.

CAM-into-C<sub>3</sub> engineering requires multiple CAM-related genes to be manipulated in a modular manner, including: 1) a carboxylation module for CO<sub>2</sub> fixation and nocturnal accumulation of malic acid in the vacuole; 2) a decarboxylation module for release of CO<sub>2</sub> from malate; and 3) a stomatal control module for nocturnal stomatal opening and stomatal closure during the daytime (Yang et al. 2015). We have generated rich genomics resources for two important CAM models *Agave* (Abraham et al. 2016) and *Kalanchoë* (phytozome.jgi.doe.gov). Our comparative analysis of protein sequences and gene expression data have identified CAM-related genes in *Agave* and *Kalanchoë*. To characterize the function of CAM-related genes, the CRISPR/Cas9-based genome-editing technology (Liu et al. 2016) was used to create loss-of-function mutants in *K. fedtschenkoi*. Some of the genes related to carboxylation, decarboxylation, and stomatal movement in *Agave* and *Kalanchoë* have been engineered into three C<sub>3</sub> species as described below.

**Engineering of carboxylation module**: Genes related to carboxylation in *Agave* and *Kalanchoë*, including  $\beta$ -type carbonic anhydrase ( $\beta$ -CA), phospho*eno*lpyruvate carboxylase (PEPC), PEPC kinase (PPCK), malate dehydrogenase (MDH) and tonoplast aluminium-activated malate transporter (ALMT), were successfully introduced individually into C<sub>3</sub> species including model plants (i.e., *Arabidopsis* and tobacco) and bioenergy crop *Populus*. The transgenic plants were characterized using RT-PCR and western blot analysis to validate the expression of the transgenes. Phenotypical characterization of the transgenic plants is being conducted.

**Engineering of decarboxylation module**: Genes related to decarboxylation in *Agave* and *Kalanchoë*, including tonoplast dicarboxylate transporter (tDT), pyruvate phosphate dikinase (PPDK) and NADP-dependent malic enzyme (NADP-ME), were successfully introduced into C<sub>3</sub> species including model plants (i.e., *Arabidopsis* and tobacco) and the bioenergy crop *Populus*.

The transgenic plants were characterized using RT-PCR and western blot analysis to validate the expression of the transgenes. Phenotypical characterization of the transgenic plants is being conducted. Also, we made use of transgenic plants (*amiRs*), which show timely increase of malate by  $\approx$ 2-fold at early night, to reprogram decarboxylation using synthetic biology principles. First, we complemented *amiRs*, which are partially deficient in CO<sub>2</sub> release from malate, with a dark-induced promoter driving a functional NADP-ME enzyme and monitored gas exchange during a 20-hour period. Next, we segregated the functional NADP-ME out and reevaluated CO<sub>2</sub> release from malate in the reverted line (amiR-r). Our results demonstrate the robustness of a controller to effectively link 3 inputs, which includes an ON-OFF switch, and generate two distinct outputs for diel control of intracellular CO<sub>2</sub> in leaves. This capability has different applications including the making of future CAM-engineered plants inducible under a specific input (e.g., drought).

**Engineering of stomatal control module**: Two strategies were used to change the stomatal behavior of C<sub>3</sub> plants: 1) transfer of multiple stomatal movement-associated genes from *K. fedtschenkoi* to C<sub>3</sub> plants and 2) synthesizing AND gate genetic circuits based upon the CRISPR-dCas9 system. More than 10 genetic circuits targeting 'CAM-like stomatal movement' were designed, assembled, and transferred into three target C<sub>3</sub> species (i.e., *Arabidopsis*, tobacco, and poplar) via *Agrobacterium*-mediated transformation. Transgenic plants were generated, and identified by PCR, qRT-PCR, and western blot analysis. The effects of these genetic circuits on stomatal movement in the transgenic plants will provide useful information for optimizing the strategy of CAM-into-C<sub>3</sub> engineering. Also, success of this research could facilitate future efforts to engineer other multi-gene traits through synthetic biology approaches.

This material is based upon work supported by the Department of Energy, Office of Science, Genomic Science Program (under award number DE-SC0008834). Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the US Department of Energy (under contract number DE-AC05-000R22725).

### References

- Abraham PE, Yin H, Borland AM, Weighill D, Lim SD, De Paoli HC, Engle N, Jones PC, Agh R, Weston DJ et al. 2016. Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. *Nature Plants* **2**: 16178.
- Liu D, Hu R, Palla KJ, Tuskan GA, Yang X. 2016. Advances and perspectives on the use of CRISPR/Cas9 systems in plant genomics research. *Curr Opin Plant Biol* **30**: 70-77.
- Yang X, Cushman JC, Borland AM, Edwards EJ, Wullschleger SD, Tuskan GA, Owen NA, Griffiths H, Smith JAC, De Paoli HC et al. 2015. A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist* 207: 491-504.

### Engineering Crassulacean Acid Metabolism (CAM) to Improve Water-use Efficiency of Bioenergy Feedstocks

**John C. Cushman<sup>1</sup>\*** (jcushman@unr.edu), Jungmin Ha<sup>1</sup>, Sung Don Lim<sup>1</sup>, Bernard W. M. Wone<sup>1</sup>, Won Cheol Yim<sup>1</sup>, Travis Garcia<sup>1</sup>, Pradeep K. Yerramsetty<sup>1</sup>, Juli Petereit<sup>2</sup>, Richard Tillett<sup>2</sup>, Karen A. Schlauch<sup>2</sup>, Anne M. Borland<sup>3,4</sup>, Alistair Leverett<sup>4</sup>, Louisa V. Dever<sup>5</sup>, Susanna F. Boxall<sup>5</sup>, Jade Waller<sup>5</sup>, Jack Davies<sup>5</sup>, Phaitun Bupphada<sup>5</sup>, Nirja Kadu<sup>5</sup>, James Hartwell<sup>5</sup>, Henrique C. De Paoli<sup>3</sup>, Rongbin Hu<sup>3</sup>, Nancy Engle<sup>3</sup>, Ryan Agh<sup>3</sup>, Degao Liu<sup>3</sup>, Kaitlin Palla<sup>3,6</sup>, Robert C. Moseley<sup>3,6</sup>, Timothy J. Tschaplinski<sup>3</sup>, David J. Weston<sup>3</sup>, Hong Guo<sup>7</sup>, Hao-Bo Guo<sup>7</sup>, Henfu Yin<sup>3</sup>, Deborah Weighill<sup>3,6</sup>, Piet C. Jones<sup>3,6</sup>, Daniel Jacobson<sup>3,6</sup>, Olaf Czarnecki<sup>3</sup>, Anthony C. Bryan<sup>3</sup>, Sara S. Jawdy<sup>3</sup>, Zong-Ming Cheng<sup>8</sup>, Jin-Gui Chen<sup>3</sup>, Stan D. Wullschleger<sup>10</sup>, Timothy Tschaplinski<sup>3</sup>, Paul E. Abraham<sup>9</sup>, Robert L. Hettich<sup>9</sup>, Gerald A. Tuskan<sup>3</sup>, and Xiaohan Yang<sup>3</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, University of Nevada, Reno. <sup>2</sup>Nevada Center for Bioinformatics, University of Nevada, Reno; <sup>3</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; <sup>4</sup>School of Biology, Newcastle University, Newcastle upon Tyne; <sup>5</sup>Department of Plant Sciences, Institute of Integrative Biology, University of Liverpool; <sup>6</sup>Bredesen Center, University of Tennessee, Knoxville, TN<sup>; 7</sup>Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville; <sup>8</sup>Department of Plant Sciences, University of Tennessee, Knoxville; <sup>9</sup>Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN. <sup>10</sup>Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

### http://cambiodesign.org

Project Goals: The long-term goal of this project is to enhance the water-use efficiency (WUE) and adaptability to hotter, drier climates of species that normally perform  $C_3$  photosynthesis by introducing crassulacean acid metabolism (CAM). Photosynthetic performance and WUE will be enhanced in *Arabidopsis* and *Populus* by: 1) defining the genetic basis of key CAM modules in both eudicot and monocot CAM species, 2) characterizing the regulation of 'carboxylation', 'decarboxylation', and 'inverse stomatal control' gene modules of CAM via loss-of-function studies in model CAM species, 3) deploying advanced genome engineering technologies to enable transfer of fully functional CAM modules into  $C_3$  species and 4) analyzing the effects of these transgenic modules on 'stomatal control', CO<sub>2</sub> assimilation and transpiration rates, biomass yield, and WUE in *Arabidopsis* and *Populus*.

Global warming trends are leading to increased terrestrial soil drying, reduced terrestrial net primary production and carbon sinks, global food security and future biofuel production, and the global expansion of drylands that already cover 42% of the earth's surface. In order to offset these negative effects, an increased reliance upon crassulacean acid metabolism (CAM) crops or the introduction of CAM, a water-wise form of photosynthesis, into C<sub>3</sub> food and bioenergy crops might serve as a useful strategy to improve the water-use efficiency (WUE) of sustainable biomass production systems in the future (1). CAM features inverse stomatal behavior, in which stomata are open at night for  $CO_2$  uptake when evapotranspiration rates are reduced compared with the daytime and closed during all or part of the day, thereby maximizing WUE. CAM also exploits a temporal separation of nocturnal  $CO_2$  uptake and fixation by phospho*enol*pyruvate carboxylase (PEPC), which leads to the formation of  $C_4$  organic acids that are stored in the vacuole. The subsequent decarboxylation of these organic acids during the day releases  $CO_2$  and concentrates it around ribulose-1-5-bisphosphate carboxylase/oxygenase (RUBISCO), suppressing photorespiration, while resulting in carbohydrate production *via* the  $C_3$  Calvin–Benson cycle.

Detailed functional and integrative 'omics analyses of several CAM model or crop species including Kalanchoe (2, phytozome.jgi.doe.gov), Mesembryanthemum crystallinum, Agave (3), and pineapple (4), have recently defined the basic genetic requirements for CAM. Both K. fedtschenkoi and M. crystallinum were selected recently as DOE JGI Flagship Genome species. The development of synthetic RNAi-mediated gene silencing strategies targeting multiple genes (5) and CRISPR/Cas9 strategies for precise genome editing (6) are expedient ways to down-regulate, knock-out, or alter the expression of specific gene modules or pathways. Loss-of-function studies of individual enzymes, metabolite transporters, and regulatory proteins or transcription factors are being used to provide critical insights into the basic genetic requirements for CAM. For example, RNAi-mediated gene silencing of specific CAM components, such as mitochondrial NAD-malic enzyme and cytosolic/plastidic pyruvate orthophosphate dikinase revealed not only impaired nocturnal CO<sub>2</sub> uptake, but also reduced circadian clock-controlled phosphorylation of PPC (7). Other studies using RNAi lines of K. fedstchenkoi have shown that the route of nocturnal starch degradation is a key point of divergence between C<sub>3</sub> photosynthesis and CAM species. In C<sub>3</sub> species, hydrolytic starch degradation produced glucose and maltose, which is exported from the chloroplast as substrate for the provision of sucrose for growth. In contrast, phosphorolytic starch degradation in CAM species produces substrates such as glucose-6-phosphate, which is exported from the chloroplast for production of PEP in the cytosol (8). Such information is critical for knowing which genes to select when creating synthetic gene circuits to reconstruct CAM carboxylation and decarboxylation subpathways.

Facile gene stacking strategies for the assembly of a large number of transcription units (TUs) with appropriate circadian and drought-inducible expression patterns are necessary for the genetic reconstitution of facultative CAM into host C<sub>3</sub> species (9). We have created a plantspecific position/adaptor/carrier vector system originally designed for engineering mammalian cells (10) that enables the rapid, reliable, and scalable creation of complex gene circuits using the Gibson isothermal assembly process (11). Design and construction of CAM-specific carboxylation and decarboxylation gene circuits containing 9 and 15 genes has been completed and are in the process of being introduced into Arabidopsis and Poplar. The gene circuits were designed to include mesophyll-specific, drought-inducible, and appropriately timed circadian expression patterns of the transgenes in order to engage the CAM pathway only during waterdeficit stress conditions. Lastly, tissue succulence has been successfully engineering in the C<sub>3</sub> photosynthesis model species A. thaliana in order to increased mesophyll cell size for increased malate storage capacity and reduced intercellular air space to limit internal CO<sub>2</sub> diffusion out of the leaf during the day to favor refixation by ribulose 1,5-bisphosphate carboxylase oxygenase (RUBISCO) and thereby increases the capacity to perform CAM. Added benefits of engineered tissue succulence included increased biomass production, increased WUE, and tolerance to water-deficit (drought) and salinity stress. The combination of engineered CAM and tissue succulence is expected to increase the WUE of bioenergy feedstocks and potentially expand their production into more marginal, abandoned, or semi-arid regions.

### References

1) Yang X, Cushman JC, Borland AM, Erika J. Edwards EJ, Wullschleger SD, Tuskan GA, Owen NA, Griffiths H, Smith JAC, De Paoli HC, Weston DJ, Cottingham R, Hartwell J, Davis SC, Silvera K, Ming R, Schlauch KA, Abraham P, Stewart JR, Guo H-B, Albion RA, Ha J, Lim SD, Wone BWM, Yim WC, Garcia T, Mayer JA, Petereit J, Nair SS, Casey E, Hettich RL, Ceusters J, Ranjan P, Palla KJ, Yin H, Reyes-García C, Andrade JL, Freschi L, Beltrán JD, Dever LV, Boxall SF, Waller J, Davies J, Bupphada P, Kadu N, Winter K, Sage RF, Aguilar CN, Schmutz J, Jenkins J, Holtum, JAC<sup>(2015)</sup> A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist.* 207: 491–504.

2) Hartwell J, Dever LV, Boxall SF (2016). Emerging model systems for functional genomics analysis of crassulacean acid metabolism. *Current Opinion in Plant Biology* 31: 100–108.

3) Abraham PE, Yin H, Borland AM, Weighill D, Lim SD, De Paoli HC, Engle N, Agh R, Weston DJ, Wullschleger SD, Tschaplinski T, Jacobson D, Cushman JC, Hettich RL, Tuskan GA, Yang X (2016) Transcript, protein and metabolite temporal dynamics in the CAM plant *Agave. Nature Plants.* 2:16178.

4) Ming R, Van Buren R, Wai CM, Tang H, Schatz MC, Bowers, JE, Lyons E, Wang M-L, Chen J, Biggers E, Zhang J, Huang L, Zhang L, Miao W, Zhang J, Ye Z, Miao C, Lin Z, Wang H, Zhou H, Yim WC, Priest HD, Zheng C, Woodhouse M, Edger PP, Guyot R, Guo H-B, Guo H, Zheng G, Singh R, Sharma A, Min X, Zheng Y, Lee H, Gurtowski J, Sedlazeck F, Harkess A, McKain MR, Liao Z, Fang J, Liu J, Zhang X, Zhang Q, Hu W, Qin Y, Wang K, Chen L-Y, Shirley N, Lin Y-R, Liu L-Y, Hernandez AG, Wright CL, Bulone V, Tuskan GA, Heath K, Zee F, Moore PH, Sunkar R, Leebens-Mack JH, Mockler T, Bennetzen JL, Freeling M, Sankoff D, Paterson AH, Zhu X, Yang X, Smith, JAC, Cushman JC, Paull RE, Qingyi Yu, Q (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics* 47:1435–1442.

5) Czarnecki O, Bryan AC, Jawdy SS, Yang X, Cheng Z-M, Chen J-G, Tuskan GA (2016). Simultaneous knockdown of six non-family genes using a single synthetic RNAi fragment in *Arabidopsis thaliana*. *Plant Methods* 12:16.

6) Liu D, Hu R, Palla KJ, Tuskan GA, Yang X (2016). Advances and perspectives on the use of CRISPR/Cas9 systems in plant genomics research. *Current Opinion in Plant Biology* 30:70–77.

7) Dever LV, Boxall SF, Knerova J, Hartwell J (2015) Transgenic perturbation of the decarboxylation phase of crassulacean acid metabolism alters physiology and metabolism but only has a small effect on growth. *Plant Physiol*. 167:44–59.

8) Borland AM, Guo H-B, Yang X, Cushman JC (2016). Orchestration of carbohydrate processing for crassulacean acid metabolism. *Current Opinion in Plant Biology* 31: 118–124. doi:10.1016/j.pbi.2016.04.001

(9) De Paoli HC, Tuskan GA, Yang X (2016). An innovative platform for quick and flexible joining of assorted DNA fragments. *Scientific Reports* 6:19278.

(10) Guye P, Li Y, Wroblewsk L, Duportet X, Weiss R (2013) Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Research* 41: e156

(11) Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CAr, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6: 343–345

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Genomic Science Program under Award Number DE-SC0008834. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract Number DE-AC05-000R22725.

### RNase III as a Tool to Manipulate Transcript Stability: Identifying *in vivo* Targets in *Escherichia coli*

Gina Gordon<sup>1</sup>, Jeffrey C. Cameron,<sup>2</sup> Michael D. Engstrom<sup>1</sup>, and **Brian F. Pfleger**<sup>1,\*</sup> (pfleger@engr.wisc.edu)

<sup>1</sup>University of Wisconsin, Madison; <sup>2</sup>University of Colorado, Boulder

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

### **Project Goals:**

For this project we will use RNA-sequencing techniques to study mRNA stability in *Escherichia coli* and *Synechococcus* sp. PCC7002, a sequenced cyanobacteria with promising industrial traits. We will use this method to study how various RNA processing enzymes affect decay of mRNA from a global perspective. We will use the knowledge gained from our studies of mRNA turnover to develop design rules for (de)stabilizing transcripts and apply these rules to ongoing metabolic engineering projects in PCC 7002 for use in photosynthetic biorefineries.

Our specific objectives are:

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

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

CRISPRi is a trans-acting regulator strategy that provides novel opportunities to study essential genes and regulate metabolic pathways. We have adapted the clustered regularly interspersed palindromic repeats (CRISPR) system from *Streptococcus pyogenes* to repress genes *in trans* in the cyanobacterium *Synechococcus sp.* strain PCC 7002 (hereafter PCC 7002) and achieved conditional and titratable repression of heterologous YFP, downregulation of phycobillisomes, creation of a conditional auxotroph, and demonstrated a novel strategy for increasing central carbon flux by conditionally downregulating a key node in nitrogen assimilation. We will describe our system and its future application to studying mRNA stability in PCC 7002.

### References

- Cameron JC, Gordon GC, Pfleger BF. Genetic and genomic analysis of RNases in model cyanobacteria. *Photosynthesis Research*. Oct;126(1):171-83 (2015). DOI: <u>10.1007/s11120-015-0076-2</u>
- 2. Gordon GC, Korosh TC, Cameron JC, Markley AL, Begemann MB, **Pfleger BF**. CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Metabolic Engineering*. Jul 29. pii: S1096-7176(16)30062-3 (**2016**) DOI: <u>10.1016/j.ymben.2016.07.007</u>

### Funding Statement:

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

### Engineering Anaerobic Gut Fungi for Lignocellulose Breakdown

John K. Henske<sup>1</sup>, Sean P. Gilmore<sup>1</sup>, Kevin V. Solomon<sup>1</sup>, Susanna Seppala<sup>1</sup>, **Michelle A. O'Malley** (momalley@engineering.ucsb.edu)<sup>1</sup>\*

<sup>1</sup>Department of Chemical Engineering, University of California, Santa Barbara

http://omalleylab.com

Project Goals: The goal of this project is to engineer anaerobic gut fungi as novel platform organisms for biofuel production from plant material. To accomplish this goal, a panel of anaerobic fungi will be isolated from different herbivores and screened for their ability to degrade lignocellulose. The basic metabolic networks that govern lignocellulose hydrolysis within anaerobic fungi will also be determined, and models will be generated to describe how important enzyme groups are coordinated during breakdown. Using this information, genetic transformation strategies to manipulate gut fungi will be developed, which would endow them with enhanced functionality against a range of industrially relevant substrates. Collectively, this information will establish the molecular framework for anaerobic fungal hydrolysis, and will guide in the development of lignocellulosic biofuels.

Anaerobic fungi are the primary colonizers of biomass within the digestive tract of large herbivores, where they have evolved unique abilities to break down lignin-rich cellulosic biomass through invasive, filamentous growth and the secretion of powerful lignocellulolytic enzymes and enzyme complexes (cellulosomes). Despite these attractive abilities, considerably less genomic and metabolic data exists for gut fungi compared to well-studied anaerobic bacteria and aerobic fungi that hydrolyze cellulose. This presents a significant knowledge gap in understanding gut fungal function, substrate utilization, and metabolic flux, which has prohibited the genetic and functional modification of gut fungi. Our approach combines nextgeneration sequencing with physiological characterization to establish the critical knowledge base to understand lignocellulose breakdown by gut fungi.

At the outset of this project, a fungal enrichment experiment was conducted to isolate biomass-degrading fungi from large herbivores at the Santa Barbara Zoo that would serve as model systems. These isolates have officially been assigned species names, and are indexed in the Index Fungorum (http://www.indexfungorum.org). In the past 2 years, our group has worked with collaborators at the Broad Institute of MIT and Harvard, as well as the Joint Genome Institute (JGI) (via approved CSP projects) to sequence transcriptomes and genomes of the anaerobic fungi *Neocallimastix californiae, Anaeromyces robustus,* and *Piromyces finnis.* Through this work, we found that the number of biomass-degrading genes in the anaerobic fungi was richer than any other microbe yet described in nature. Compared to "higher" (e.g.

more evolved fungi), the more primitive anaerobic fungi are equipped with a rather even distribution of cellulose-degrading enzymes, xylan-degrading enzymes, and accessory enzymes that have been largely lost to evolution. This illustrates not only their powerful biomass-degrading abilities, but also the potential to identify completely novel sequences in their genomes that could be exploited for biotechnology. More recently, we have worked with the JGI to release the first high-quality genomes of the anaerobic gut fungi (manuscript in progress, results available on Mycocosm).

Compared to model organisms, there is a relative dearth of information about the metabolism of anaerobic gut fungi. To address this issue, we used transcriptomic information to identify enzymes from critical metabolic processes for the first time. Transcripts identified in the *de novo* assembled transcriptomes were functionally annotated using a combination of NCBI BLAST, EMBL-EBI InterProScan, and Gene Ontology mapping. Enzymes were identified by their Enzyme Commission (EC) numbers, which were only assigned to transcripts with a high similarity to InterPro protein domains. Combining this information with metabolic maps from KEGG databases, the core metabolic maps were filled in and the sugars that could be metabolized by the fungus were identified. This analysis revealed that the fungus contains all of the enzymes required for glycolysis through ethanol fermentation, but takes an irreversible route as it is missing phosphoglycerate mutase (EC:5.4.2.11), which indicates that the gut fungi cannot perform gluconeogenesis. Furthermore, all enzymes necessary for the metabolism of fructose, mannose, sucrose,  $\alpha$ -galactose, and xylose were present.

Finally, we have taken the first major steps towards a method to stably genetically transform anaerobic gut fungi. Beyond the obvious lack of knowledge of their genomes, the very thick cell wall of anaerobic fungi (greater than 5um) often prevents delivery of any amount of DNA. So far, we have shown that gut fungal zoospores (cells in the early stage of their life cycle) are amenable to transformation by electroporation (DNA uptake during electroshock). This optimized procedure was the result of a design of experiments approach, where electrotransformation, biolistic transformation, and chemical transformation methods were evaluated as applied to fungal zoospores capable of taking in foreign extracellular plasmid DNA, and then are able to express the genetic reporter encoded on the plasmid as they mature. In this case, the genetic reporter is the GusA enzyme, which hydrolyzes the reagent X-gluc, creating an easily observable blue colony formation. Future experiments will leverage this transformation breakthrough to genomically modify the model fungal systems for the first time.

This Project is supported by the Office of Biological and Environmental Research through the DOE Office of Science (Early Career Program)

### Expression of a Hyperthermophilic Endoglucanase in Poplar to Improve Ethanol Production

Yao Xiao<sup>1\*</sup> (yxiao07@syr.edu), Charleson Poovaiah, <sup>1</sup> Xuejun He, <sup>1</sup> and **Heather D.** Coleman

<sup>1</sup>Syracuse University, Syracuse, NY

http://asfaculty.syr.edu/pages/bio/Coleman-Heather.html

Project Goals: This project investigates the effect of *in planta* expression of a hyperthermophilic endoglucanase on plant development and cell wall structure in poplar. The ultimate goal is to produce transgenic biomass that requires less input of commercial cellulases with the same yield of fermentable sugars, thus reducing the cost of bioethanol production.

Ethanol produced from non-food lignocellulosic biomass holds advantages over fossil fuels and corn-derived ethanol for a number of well-documented reasons. During lignocellulosic ethanol production, cellulases are required to hydrolyze cellulose to glucose following by ethanol fermentation. Currently, microbial-derived cellulases are the only available option, but they remain too expensive for successful commercial lignocellulosic ethanol production. Expression of cellulases in the biomass feedstock itself is a potential solution for more affordable ethanol production. This project aims to produce a hyperthermophilic endoglucanase in poplar, thereby improving efficiency of hydrolysis and reducing the cost of cellulosic ethanol production. Transgenic poplar trees harboring a Thermotoga neapolitana endoglucanase (TnCelB) under the control of the Cauliflower Mosaic Virus 35S promoter have been generated and are growing in the greenhouse. Preliminary data in tissue culture shows that poplar-derived TnCelB kept high activity to substrates at 100°C. Transgenic poplar will be assessed for plant growth phenotype, levels of TnCelB expression and activity, and cell wall composition. In addition, the biomass will undergo enzymatic hydrolysis to assess improvements in breakdown due to the expression of the gene. It is anticipated that expression of TnCelB will result in biomass that is more easily digested following thermochemical pretreatment, resulting in an improvement of sugar release and ethanol yield.

Funding statement: This material is based upon work supported by the Department of Energy under Award Number DE-SC0010411.

### **Extreme Inducible Expression of Cellulases in Poplar**

Charleson Poovaiah<sup>1\*</sup> (crpoovai@syr.edu), Yao Xiao,<sup>1</sup> and Heather D. Coleman<sup>1</sup>

<sup>1</sup>Syracuse University, Syracuse, NY

http://asfaculty.syr.edu/pages/bio/Coleman-Heather.html

# Project Goals: The overall goals of the project are to verify in poplar In Plant Activation (INPACT) technology<sup>1</sup>, which enables inducible expression of genes and accumulation of proteins at very high levels *in planta*, and to evaluate the ability of cellulases produced through this technology to hydrolyze cellulose to simple sugars for fermentation.

Cellulolytic enzymes are used to hydrolyze carbohydrates within lignocellulosic biomass into fermentable sugars, and the cost of these enzymes significantly increases the cost of bioethanol. An alternative is the production of these enzymes *in planta* in the lignocellulosic biomass itself. This would allow a single feedstock to play a dual role as both biomass substrate and enzyme provider. The aim of this project is to produce high levels of these cellulolytic enzymes in poplar using In Plant Activation (INPACT) technology. INPACT technology allows for very high inducible expression of recombinant proteins *in planta*. INPACT uses the rolling circle replication seen in Gemini viruses to produce high levels of gene amplification and protein production. We will verify the adaptability of this technology in poplar to accumulate proteins at very high levels. Using this technology, we will express cellulases in poplar with constitutive and tissue specific promoters. Cellulases from three major groups of enzymes, endoglucanases, exoglucanases and  $\beta$ -glucosidases, involved in the hydrolysis of cellulose will be expressed with constitutive and tissue specific promoters.

To date, cellulases from thermophilic organisms have been codon optimized and cloned into INPACT vectors. Constructs harboring the alcohol inducible replication initiation protein (Rep) which allows for the induction of the INPACT system have been successfully transformed into poplar and a mother line selected based on Rep/RepA gene expression and plant growth before and after alcohol treatment. GUS constructs in the split orientation were then super-transformed into the mother line to generate INPACT-GUS double transgenic lines. Independent transgenic poplar lines harboring both the alcohol inducible Rep construct and the split orientation GUS gene were then multiplied to produce clonal lines. Currently these INPACT-GUS lines are being evaluated in the greenhouse for expression in leaf and developing xylem using the GUS reporter system. Transgenic plants with thermostable cellulases in the split orientation have been super-transformed in the Alc-Rep mother line to generate INPACT-Cellulase double transgenic lines, along with corresponding positive and negative control lines. These lines are presently being multiplied for greenhouse evaluation.

### References

1. Dugdale, Benjamin, et al. "In plant activation: an inducible, hyperexpression platform for recombinant protein production in plants." The Plant Cell 25.7 (2013): 2429-2443.

Funding statement: This material is based upon work supported by the Department of Energy under Award Number DE-SC0010411.

### Modified Cell Wall Composition through Expression of an Expansin-Like Protein in Poplar

Yao Xiao,<sup>1</sup> Maria Dombrov,<sup>1</sup> Linnea Ritchie,<sup>1</sup> Charleson Poovaiah, and **Heather D. Coleman**<sup>1\*</sup> (hcoleman@syr.edu)

<sup>1</sup>Syracuse University, Syracuse, NY

http://asfaculty.syr.edu/pages/bio/Coleman-Heather.html

Project goal: One of the challenges associated with the expression of cellulolytic enzymes *in planta* is the alteration of cell wall chemistry, and the resultant negative effects on plant growth and biomass production. Here we express a fungus expansin-like protein in poplar, with the intent of altering cell wall structure without compromising cell wall functions and plant growth.

Degradation of cellulose to glucose for ethanol production from lignocellulosic biomass requires multiple cellulase and accessory enzymes. One of the challenges with expressing these enzymes *in planta* is a potential resultant negative phenotype. By selecting enzymes that are known to disrupt structure without eliciting a 'breakdown' of cellulose, it is anticipated that phenotypes with improved hydrolysis but without detrimental phenotypes may be achieved. One such potential enzyme is the expansin-like protein Swollenin 1 (Swo) from Trichoderma reesei. Swo has been shown to disrupt cell wall structure without a corresponding release in carbohydrate monomers [1], and to work synergistically with hydrolytic cellulases [2]. Our work explores the effect of *in planta* expression of Swo on plant cell wall chemistry and structure, while investigating the potential of using this transgenic biomass in the production of cellulosic ethanol. The gene encoding Swo has been expressed in poplar under either the constitutive Cauliflower Mosaic Virus 35S promoter or the putative vascular tissue specific Subterranean Clover Stunt Virus S7 promoter. Four CaMV lines and three SCSV lines were grown in the greenhouse for four months along with corresponding wildtype controls. While there was little variation in growth phenotype, preliminary cell wall analysis by thermogravimetric analysis suggested large changes in cellulose and hemicellulose levels. Further analysis of cell wall chemistry is ongoing, as is assessment of cell wall crystallinity and amenability of biomass to enzymatic hydrolysis.

### Funding statement: This material is based upon work supported by the Department of Energy under Award Number DE-SC0010411.

- 1. Saloheimo, M., et al., *Swollenin, a Trichoderma reesei protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials.* European Journal of Biochemistry, 2002. **269**(17): p. 4202-4211.
- 2. Gourlay, K., et al., *Swollenin aids in the amorphogenesis step during the enzymatic hydrolysis of pretreated biomass.* Bioresource technology, 2013. **142**: p. 498-503.

#### Automatic generation of genome-scale metabolism and expression models for bacteria

Laura Kutt<sup>1</sup>, Matthew DeJongh<sup>2</sup>, Christopher S. Henry<sup>3</sup>, Ines Thiele<sup>1</sup>, and Ronan MT Fleming<sup>1</sup> (\*ronan.mt.fleming@gmail.com)

<sup>1</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; <sup>2</sup>Hope College, Holland, Michigan. <sup>3</sup>Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois.

http://stanford.edu/group/SOL/multiscale/

Project Goals: The aim of the project is to increase the availability of next-generation genome-scale models that integrate metabolism and macromolecular expression (ME models). As these models can predict the rate of macromolecular synthesis, e.g., protein synthesis, and have wider coverage of cellular processes compared to commonly used genome-scale metabolic models, they facilitate more comprehensive integration of omics data. The overall goal is to make available a large set of draft ME models and enable the community to simulate with ME models on a variety of platforms, including the KBase platform. This will facilitate the extended application of genome-scale ME models in systems biology research.

With growing amounts of omics data, computational models are crucial for derivation of biological insights. Most modelling paradigms use statistical methods to find data patterns. An alternative approach uses comprehensive models built upon prior biological data contextualised using omics data derived from a particular context. Metabolism and Expression models (ME models) integrate transcription and translation processes with metabolic networks and are the new-generation genome-scale models. These models include the metabolic burden associated with expression the proteins required for catalysis of biochemical processes, such as enzymes and translation factors. All the included proteins are produced by the model taking into account the nucleotide and amino acid sequences of the expressed genes. To date, the generation of ME models has been restricted to a few species, e.g., E. coli[1, 2]. Although there are potentially an extensive number of applications for ME models, the lack of readily available draft ME models and suitable solvers has limited their wider application within systems biology. Recently, suitable solvers have been developed [3, 4]. This work describes a pipeline for automatic generation of draft ME models for bacteria using the KBase platform. We demonstrate the utility of this pipeline for generation of draft ME models for 684 strains of bacteria. In addition to examining metabolic fluxes, these models can be utilised for more comprehensive mapping of omics data to examine proteome allocation, calculate core proteome requirements, and to simulate gene-expression profiles under different conditions. We envisage that dissemination of ME models via the KBase will facilitate widespread application of this new generation of genome-scale models to bioenergy research.

### References

- [1] Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, et al. (2012) Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. PLoS ONE 7: e45635.
- [2] Ebrahim A, Brunk E, Tan J, O'Brien EJ, Kim D, et al. (2016) Multi-omic data integration enables discovery of hidden biological regularities. Nature Communications 7: 13091.
- [3] Ma D, Yang L, Fleming RMT, Thiele I, Palsson BO, et al. (2016) Reliable and efficient solution of genome-scale models of Metabolism and macromolecular Expression. arXiv:160600054 [q-bio].
- [4] Yang L, Ma D, Ebrahim A, Lloyd CJ, Saunders MA, et al. (2016) solveME: fast and reliable solution of nonlinear ME models. BMC Bioinformatics 17: 391.

This work was supported by the U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant #DE-SC0010429. Matthew DeJongh supported by a Fulbright Research Scholar Grant for Research in Computer Science at the University of Luxembourg.

### CHRR: Coordinate hit-and-run with rounding for uniform sampling of metabolic networks

Hulda S. Haraldsdóttir<sup>1</sup>, Ben Cousins<sup>2</sup>, German Preciat<sup>1</sup>\* (german.preciat.001@student.uni.lu), Ines Thiele<sup>1</sup>, Santosh Vempala<sup>2</sup> and **Ronan M. T. Fleming**<sup>1</sup>

<sup>1</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg; <sup>2</sup>School of Computer Science, Algorithms and Randomness Center, Georgia Institute of Technology, Atlanta, Georgia

http://stanford.edu/group/SOL/multiscale/

Project Goals: Development of a framework for mass conserved elementary kinetic modelling of metabolic networks [1, 2, 3, 4, 5]. This collaborative project was motivated by a need for data to constrain and validate metabolic models. Sampling algorithms have demonstrated applications in measurement and estimation of kinetic parameters, steady state fluxes and metabolite concentrations for biochemical systems [6, 7, 8].

In constraint-based modelling, physicochemical and biochemical constraints define a set of feasible states of a biochemical network. Steady state mass conservation and limits on substrate uptake constraints are specified by a set of linear equalities and inequalities that define a polyhedral convex set of feasible flux vectors. Uniform sampling of this set provides an unbiased characterisation of the metabolic capabilities of a biochemical network [9]. However, reliable uniform sampling of genome-scale biochemical networks is challenging due to their high dimensionality and inherent anisotropy. Here, we apply a new sampling algorithm, coordinate hit-and-run with rounding (CHRR) [?], to metabolic networks of increasing dimensionality. This algorithm is based upon the provably efficient hit-and-run random walk [10] and crucially, it uses a preprocessing step to round the anisotropic flux set. CHRR provably converges to a uniform stationary sampling distribution and does so several times faster than a popular artificial centering hit-and-run (ACHR) algorithm [11]. We demonstrate the effects of improved convergence rate on predictions of the metabolic capabilities of Bacillus Subtilis [12].

### References

- Sun Y, Fleming RM, Thiele I, Saunders MA (2013) Robust flux balance analysis of multiscale biochemical reaction networks. BMC Bioinformatics 14: 240.
- [2] Noor E, Haraldsdóttir HS, Milo R, Fleming RMT (2013) Consistent Estimation of Gibbs Energy Using Component Contributions. PLoS Comput Biol 9: e1003098.
- [3] Artacho FJA, Fleming RMT (2014) Globally convergent algorithms for finding zeros of duplomonotone mappings. Optimization Letters: 1–16.
- [4] Artacho FJA, Fleming RMT, Vuong PT (2015) Accelerating the DC algorithm for smooth functions. arXiv:150707375 [math, q-bio].
- [5] Haraldsdóttir HS, Fleming RMT (2016) Identification of Conserved Moieties in Metabolic Networks by Graph Theoretical Analysis of Atom Transition Networks. PLOS Computational Biology 12: e1004999.
- Yang L, Mahadevan R, Cluett W (2010) Designing experiments from noisy metabolomics data to refine constraint-based models. In: American Control Conference (ACC), 2010. pp. 5143–5148. doi:10.1109/ACC.2010.5530678.
- [7] Schellenberger J, Zielinski DC, Choi W, Madireddi S, Portnoy V, et al. (2012) Predicting outcomes of steady-state 13c isotope tracing experiments using Monte Carlo sampling. BMC systems biology 6: 9.
- [8] Khodayari A, Zomorrodi AR, Liao JC, Maranas CD (2014) A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data. Metabolic Engineering 25: 50–62.
- [9] Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nature Reviews Microbiology.
- [10] Lovasz L, Vempala S (2006) Hit-and-run from a corner. In: SIAM journal on computing. Society for Industrial and Applied Mathematics, volume 35, pp. 985–1005.
- [11] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290–1307.
- [12] Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. The Journal of Biological Chemistry 282: 28791–28799.

#### Comparative evaluation of atom mapping algorithms for metabolic reactions

German Preciat Gonzalez (\*gapreciat@gmail.com), Lemmer El Assal, Ines Thiele, Hulda Haraldsdóttir, **Ronan M.T. Fleming**(ronan.mt.fleming@gmail.com)

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

#### Project Goals: Comparison of the predictive accuracy of atom mapping algorithms.

The reaction mechanism of each chemical reaction in a metabolic network can be represented as a set of atom mappings, each of which relates an atom in a substrate metabolite to an atom of the same element in a product metabolite. Atom mapping data for metabolic reactions open the door to a growing list of applications [1, 2, 3, 4]. Complete manual acquisition of atom mapping data for a large set of chemical reactions is a laborious process. Many algorithms exist to predict atom mappings. How do their predictions compare to each other and to manually curated atom mapping algorithms [5, 6, 7, 8, 9, 10]. We also compared these predictions to those obtained by manual curation of atom mappings for over five hundred reactions distributed amongst all top level enzyme commission number classes. Five of the evaluated algorithms had similarly high prediction accuracy over 91% when compared to manually curated atom mapped reactions. On average, the accuracy of the prediction was highest for reactions catalysed by oxidoreductases and lowest for reactions catalysed by ligases. In addition to prediction accuracy, the algorithms were evaluated on their availability and advanced features such as the ability to identify equivalent atoms and reaction centres, and the option to map hydrogen atoms. In addition to prediction accuracy, we found that availability and advanced features were fundamental to the selection of an atom mapping algorithm.

### References

- [1] Wiechert W (2001) 13c Metabolic Flux Analysis. Metabolic Engineering 3: 195–206.
- Haraldsdóttir HS, Fleming RMT (2016) Identification of Conserved Moieties in Metabolic Networks by Graph Theoretical Analysis of Atom Transition Networks. PLOS Computational Biology 12: e1004999.
- [3] Pey J, Planes FJ, Beasley JE (2014) Refining carbon flux paths using atomic trace data. Bioinformatics 30: 975–980.
- [4] Kotera M, Okuno Y, Hattori M, Goto S, Kanehisa M (2004) Computational Assignment of the EC Numbers for Genomic-Scale Analysis of Enzymatic Reactions. Journal of the American Chemical Society 126: 16487–16498.
- [5] First EL, Gounaris CE, Floudas CA (2012) Stereochemically Consistent Reaction Mapping and Identification of Multiple Reaction Mechanisms through Integer Linear Optimization. Journal of Chemical Information and Modeling 52: 84–92.
- [6] ChemAxon (2015) Standardizer, was used for structure canonicalization and transformation, JChem 16.1.11.0, 2015, ChemAxon (http://www.chemaxon.com). URL ChemAxon (http://www.chemaxon.com).
- [7] Rahman SA, Torrance G, Baldacci L, Cuesta SM, Fenninger F, et al. (2016) Reaction Decoder Tool (RDT): extracting features from chemical reactions. Bioinformatics 32: 2065–2066.
- [8] Kumar A, Maranas CD (2014) CLCA: Maximum Common Molecular Substructure Queries within the MetRxn Database. Journal of Chemical Information and Modeling 54: 3417–3438.
- [9] Latendresse M, Malerich JP, Travers M, Karp PD (2012) Accurate Atom-Mapping Computation for Biochemical Reactions. Journal of Chemical Information and Modeling 52: 2970–2982.
- [10] Kraut H, Eiblmaier J, Grethe G, Löw P, Matuszczyk H, et al. (2013) Algorithm for reaction classification. Journal of Chemical Information and Modeling 53: 2884–2895.

#### DistributedFBA.jl: High-level, high-performance flux balance analysis in Julia

Laurent Heirendt, Sylvain Arreckx(\*sylvain.arreckx@uni.lu), Ines Thiele, Ronan M.T. Fleming

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

http://stanford.edu/group/SOL/multiscale/

Project Goals: Constraint-Based Reconstruction and Analysis (COBRA) [1] is a widely used approach for modeling genome-scale biochemical networks and integrative analysis of omics data in a network context. For kilo-scale models, flux variability analysis (FVA) can currently be performed efficiently using existing methods [2, 3, 4]. However, these implementations perform best when using only one computing node with a few cores, which becomes a temporal limiting factor when exploring the steady state solution space of larger models. Here, we exploit Julia [5], a high-level, high-performance dynamic programming language, to distribute sets of flux balance analysis problems and compare its performance to existing implementations.

Flux balance analysis, and its variants, are widely used methods for predicting steady-state reaction rates in biochemical reaction networks. The exploration of high dimensional networks with such methods is currently hampered by software performance limitations. *DistributedFBA.jl* is a high-level, high-performance, open-source implementation of flux balance analysis in Julia. It is tailored to solve multiple flux balance analyses on a subset or all the reactions of large and huge-scale networks, on any number of threads or nodes. The code and benchmark data are freely available on *github.com/opencobra/COBRA.jl*. The documentation can be found at *opencobra.github.io/COBRA.jl*.

### References

- Palsson BØ (2015) Systems Biology: Constraint-based Reconstruction and Analysis. Cambridge University Press, 1 edition.
- [2] Gudmundsson S, Thiele I (2010) Computationally efficient flux variability analysis. BMC Bioinformatics 11: 489.
- [3] Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Systems Biology 7: 74.
- [4] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290–1307.
- [5] Bezanson J, Edelman A, Karpinski S, Shah VB (2014) Julia: A Fresh Approach to Numerical Computing. arXiv:14111607
  [cs].

#### Software development of The Constraint-Based Reconstruction and Analysis Toolbox

Sylvain Arreckx(\*sylvain.arreckx@uni.lu), Laurent Heirendt, Ines Thiele, Ronan M.T. Fleming

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

http://stanford.edu/group/SOL/multiscale/

Project Goals: The aim of this project is to increase the robustness and enhance the stability of The Constraint-Based Reconstruction and Analysis Toolbox (The COBRA Toolbox). Since its release in 2011, The COBRA Toolbox v2.0 [1] is widely used for modeling, analysing and predicting a variety of metabolic phenotypes using genome-scale biochemical networks. Recently added new features include novel sampling algorithms, sparse optimisation code, new test functions and tools for integration of metabolomic data, to name a few. With an increasing number of contributions from developers around the world, the code base is evolving at a fast pace. In order to guarantee consistent, stable, and high quality code, the goal is to implement a continuous integration approach to development of The COBRA Toolbox.

The popularity of The COBRA Toolbox, its open-source nature, and the increasing number of collaborators bear the need for a continuous integration server. This automated environment ensure that every change in code, as minor as it may be, is fully tested before being released as part of a stable version. Each contribution is verified automatically, and for each code submission, a comprehensive test suite is run in order to detect bugs and integration errors before release. Consequently, the automated testing environment ensures higher quality code and the release of well tested computer code. The COBRA Toolbox provides researchers with a common high-level interface to a variety of optimization solvers, which allows users to easily select the most effective optimisation algorithm to solve genome-scale modeling problems. Within the continuous integration environment, the code is automatically run with multiple versions of MATLAB and several releases of the solvers, which guarantees its compatibility. The documentation is automatically deployed from the continuous integration server and generated based on the comments within the code. The code and the documentation are freely released as part of the openCOBRA project at github.com/opencobra/cobratoolbox.

### References

 Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290–1307.

### Thermodynamically constraining a genome-scale metabolic model with von Bertalanffy 2.0

Lemmer El Assal, Hulda S. Haraldsdóttir and Ronan M. T. Fleming (\*ronan.mt.fleming@gmail.com)

Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

#### http://stanford.edu/group/SOL/multiscale/

Project Goals: In flux balance analysis of genome-scale stoichiometric models of metabolism, the principal constraints are uptake or secretion rates, the steady state mass conservation assumption and reaction directionality. Von Bertalanffy 2.0 is an algorithmic pipeline for quantitative assignment of reaction directionality in multi-compartmental genome scale models based on an application of the second law of thermodynamics to each reaction, via estimation of thermodynamic properties using the Component Contribution method.

Reaction directionality in metabolic network reconstructions should be consistent with the second law of thermodynamics, which implies that a reaction can only proceed in a direction associated with a negative chemical potential difference. In biochemical systems, the chemical potential difference is termed the (Legendre) transformed reaction Gibbs energy, which is appropriate for a certain in vivo temperature, pH, ionic strength, electrical potential and metabolite concentrations. In a typical metabolic reconstruction process, reaction directionality is assigned based on thermodynamic data such as experimentally measured equilibrium constants, whenever such data is available. Two problems frequently arise: either no thermodynamic data is available in the literature on the reaction in question or thermodynamic data is available, but only for conditions that differ from in vivo conditions. Quantitative assignment of reaction directionality is possible, even if thermodynamic information does not exist for a particular reaction in question by estimating thermodynamic properties of metabolites and metabolic reactions based on structurally similar metabolites and reactions. Von Bertalanffy 2.0 is an extension to The COBRA toolbox [1] that provides a set of MATLAB functions to that enable quantitative assignment of reaction directionality for all reactions in a multi-compartmental, genome-scale metabolic network reconstructions. This estimation is most accurate for reactions that are structurally most similar to reactions where experimental thermochemical data exists. The methodology for estimation of thermodynamic properties is based on recent advancement in estimation of Gibbs Energy using the Component Contribution method<sup>[2]</sup>. This method combines estimation via a Group Contribution method with the more accurate Reactant Contribution method by decomposing each reaction into two parts and applying one of the methods on each of them. This method gives priority to the reactant contributions over group contributions while guaranteeing that all estimations will be consistent, i.e., will not violate the first law of thermodynamics. The result is a thermodynamically constrained genome-scale model of a metabolic network with as confidence intervals provided for all estimates. Dissemination via Von Bertalanffy 2.0 is envisaged to facilitate the wide use of thermodynamic data for a better understanding of metabolism.

### References

- Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6: 1290–1307.
- [2] Noor E, Haraldsdóttir HS, Milo R, Fleming RMT (2013) Consistent Estimation of Gibbs Energy Using Component Contributions. PLoS Comput Biol 9: e1003098.

#### Cardinality optimisation in constraint-based modelling of metabolism

Hoai Minh Le, Hulda S. Haraldsdottir, Tu Vuong Phan, Ines Thiele, **Ronan M.T. Fleming** (\*ronan.mt.fleming@gmail.com) <sup>1</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg.

http://stanford.edu/group/SOL/multiscale/

Project Goals: A growing number of genome-scale biochemical networks are becoming available, especially for metabolism. Although the models can be constructed from omics data, several types of errors can occur that make the quality of the model insufficient. manually testing the accuracy of models exceeding several thousands of metabolites and reactions is impractical. Hence, it is important to develop tools to verify the biological models, or to automatically detect and correct errors. These tools often involve the solution of an optimisation problem with a discrete solution. In general, this problem is computationally intractable at large scale but in specific cases continuous approximation can yield an adequate solution, even for large computational models.

Several biochemical applications can be mathematically formulated as a cardinality optimisation problem [1, 2, 3]. The cardinality optimisation problem refers to an optimisation problem that involves optimising the number of non-zero components of vector, that is, the  $\ell_0$  norm. In general, this cardinality optimisation problem NP-hard. In this work, we investigate a non-convex approximation approach which consists in replacing the  $\ell_0$  norm by non-convex continuous functions. The resulting problem is non-convex and can be reformulated as a Difference of Convex (DC) program. Based on the theory of DC programming and DC Algorithm (DCA), we present a generic algorithm for solving several cardinality optimisation problems arising in constraint-based modelling of biochemical networks. As application of the proposed algorithm, we study three problems in systems biochemistry: the detection of the stoic chiometrically consistent part of a metabolic network, the sparse flux balance analysis and relaxation of constraints to obtain non-zero steady state flux in an otherwise infeasible flux balance analysis problem. Numerical tests on various biochemical networks from a range of species clearly show that our algorithms outperform existing related algorithms.

### References

- Vlassis N, Pacheco MP, Sauter T (2014) Fast reconstruction of compact context-specific metabolic network models. PLoS Comput Biol 10: e1003424.
- [2] Gevorgyan A, Poolman MG, Fell Da (2008) Detection of stoichiometric inconsistencies in biomolecular models. Bioinformatics 24: 2245–51.
- [3] Fleming RM, Vlassis N, Thiele I, Saunders MA (2015) Conditions for duality between fluxes and concentrations in biochemical networks. arXiv preprint arXiv:151202690.

### Title: High Resolution DNA Stable Isotope Probing of Soil Indicates Changes in Microbial Community Metabolism Associated in Disturbance Due to Tillage

C. Koechli\* (cnk29@cornell.edu), N. D. Youngblut, D. H. Buckley

School of Integrative Plant Science, Cornell University, Ithaca

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

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

Bacteria are essential to the cycling and storage of carbon in the soil ecosystem. Tillage decreases soil organic matter content and changes the composition of soil microbial communities. Differences in microbial ecology between no-till vs tilled soils may contribute to differences in organic matter loss pathways, however mechanistic linkages between microbial community structure and function remain unclear in soils.

We employed high resolution DNA stable isotope probing (HR-SIP) to evaluate the temporal dynamics of microbial metabolism associated with the degradation of dissolved (<sup>13</sup>C-xylose) and particulate <sup>(13</sup>C-cellulose)</sup> forms of carbon in soils from a long-term tillage experiment. In addition, we used high throughput sequencing of 16S rRNA gene amplicons to assess seasonal variation in taxon relative abundance in relation to tillage history and pattern of isotope incorporation in HRexperiments. Bacterial communities vary significantly with tillage as expected SIP (PERMANOVA, R2 = 0.14, p = 0.001). No-till soils also had significantly higher rates of soil respiration and <sup>13</sup>C-xylose mineralization, but not <sup>13</sup>C-cellulose mineralization relative to tilled soil. The bacteria that incorporated <sup>13</sup>C xylose initially (days 1 and 3) differed in tilled vs. no-till soils, though similar taxa were ultimately enriched in both soil types over time. In contrast, the bacteria that incorporated <sup>13</sup>C cellulose remained similar between tilled and no-till soils throughout the experiment. The taxa participating in carbon transformations differed as a function of soil management history, with implications for carbon fate. These results suggest that changes in the structure of the microbial community, caused by tillage, affects xylose degradation dynamics but not cellulose degradation dynamics.

Funding statement.

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

### Title: Mapping microbial food web dynamics in soil with high resolution stable isotope probing

Authors: N. D. Youngblut (nyoungb2@gmail.com), C. Koechli, S. Barnett\*, D. H. Buckley

School of Integrative Plant Science, Cornell University, Ithaca

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

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

Soils make up the largest active carbon pool on the planet. Although carbon cycling in soil is largely mediated by microbial life, the specific taxonomic groups that perform each role in the soil microbial food web have not been well resolved. High-resolution stable isotope probing (HR-SIP) leverages highly multiplexed high-throughput 16S rRNA sequencing to simultaneously map *in situ* substrate assimilation dynamics to potentially thousands of finely resolved microbial taxa. We performed an HR-SIP experiment that employed nine <sup>13</sup>C-labeled substrates chosen to represent organic matter present during plant biomass degradation (cellulose, xylose, glucose, glycerol, vanillin, palmitic acid, amino acids, lactate, and oxalate).

We observed a succession of <sup>13</sup>C-substrate respiration and incorporation into bacterial biomass. Relatively labile dissolved substrates (e.g., glucose) were utilized rapidly (days ~1-6), followed by oxalate (days ~6-14), and finally by insoluble substrates such as cellulose and palmitic acid (days ~14-30). The total amount of <sup>13</sup>C respired varied substantially between substrates, from ~35% for vanillin to nearly 100% for lactate. The number <sup>13</sup>C-incorporating taxa ("incorporators") also varied among treatments, with a greater diversity of taxa incorporating insoluble substrates (cellulose or palmitic acid), than dissolved substrates. We find evidence for both generalist and specialist taxa. For example, certain *Gammaproteobacteria* and *Firmicutes* responded rapidly and consumed almost all substrates, though many other taxa in these clades specialized on one or two substrates. In contrast, many *Verrucomicrobia* grew slowly and specialized on insoluble substrates such as <sup>13</sup>C-cellulose and <sup>13</sup>C-palmitic acid. Clustering OTUs by their signal of incorporation in each treatment produced an assortment of groups with differing functional roles in the

carbon cycle (e.g., vanillin or cellulose specialists). Most functionally coherent groups contained multiple phyla. These results suggest pervasive niche partitioning among bacterial taxa in the soil carbon cycle, with partitions for both the relatively transient dissolved organic matter pool and the more persistent particulate organic matter pool. More generally, these findings will help define ecologically relevant taxonomic groups of microbes with coherent functional roles in the soil microbial food web.

#### Funding statement.

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

### Bacterial Traits Linking *Avena* Exudate Chemistry and Rhizosphere Bacterial Community Structure During Root Development

Hee jung Cho\*(heejungcho@lbl.gov)<sup>1,2</sup>, Ulas Karaoz<sup>2</sup>, Kateryna Zhalnina<sup>2</sup>, **Trent Northen<sup>2</sup>**, **Mary Firestone<sup>1,2</sup> and Eoin Brodie**<sup>1,2</sup>

<sup>1</sup>University of California, Berkeley; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, California

**Project Goals:** Plant roots are a major source of soil organic carbon, and the rhizosphere is a hotspot of biological activities mediating multiple nutrient cycles. Plant roots exude varying combinations of organic substrates into soil, including amino acids, sugars, organic acids, fatty acids and secondary metabolites. The rhizosphere of growing plant roots is a dynamic interface for complex rhizospheric interactions, involving both one-to-one interactions and multitrophic interactions. Growing plant roots alter soil bacterial community composition by selecting a subset of bacteria from surrounding soil. Our study site, the Hopland UC research field station in Northern California, is a semi-arid annual grassland that has been dominated by *Avena* plants for over one hundred years, and thus has long history of rhizospheric interactions associated with *Avena* roots. Many studies have demonstrated the selection of soil bacterial taxa by growing plant roots and have elucidated the role of individual exudate compounds. However, we still lack clear understanding on how the changing chemistry of plant root exudates interacts with bacterial traits such as substrate utilization preferences and growth strategies to shape bacterial communities in the rhizosphere.

Mapping metagenomes from the Avena rhizosphere and bulk soil during plant developmental stages to the genomes of thirty nine bacterial isolates and ninety seven genome bins assembled from the soil metagenomes, we classified bacterial isolates and genome bins into three groups representing positive, negative and neutral responders to root growth. Comparative analyses of these genomes was then used to identify bacterial traits that may be essential for growth and survival in the rhizosphere. This demonstrated that multiple genes involved in amino acid, sugar and sulfonate transport, flagellar protein assembly and function, and chemotaxis were more abundant in bacteria that responded positively to growing Avena roots. Fragment recruitment of metagenome reads to isolate genomes was used to identify strain variation and relate recovered isolates to their metagenome counterparts and showed that closely related strains appeared to have divergent responses to root growth. To investigate this we compared the genomes of three closely related *Bradyrhizobium* isolates that exhibited differential responses to root growth (positive or neutral) and highlighted regions unique to the positive responder. These encoded genes for amino acid and carbohydrate transporters, genes coding for xylanase/chitin degradation, and, chemotaxis related proteins amongst others, demonstrating that the acquisition of a relatively small number of genes may result in niche optimization and improved fitness.
Exometabolite analyses were used to identify how root exudates changed across *Avena* developmental stages, and demonstrated that exudate chemistry could be classified into three major patterns of exudation corresponding to early developmental stage, vegetative stages and senescence. Further exometabolomic analyses defined the exudate utilization preferences of bacterial isolates and demonstrated that positive, negative and neutral responders to root growth had distinct preferences for specific components of root exudates. Positive responders showed higher uptake of organic acids, and amino acids that corroborated the greater number of transporters for these substrates encoded in their genomes. In particular, aromatic organic acids including indole-3-acetic acids, salicylic acid, cinnamic acid and nicotinic acid were primarily consumed by the positive responders to root growth, suggesting a significant role of these metabolites in shaping the rhizosphere microbial community. Further work is being carried out to confirm the genetic basis and generality of these rhizosphere traits.

### Funding statement.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Numbers DE-SC0010570 and DE-SC0016247 to UC Berkeley and University of Oklahoma. Part of this work was performed at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231.

### Linking Arctic Lake Sediment Microbial Ecology to Carbon Biogeochemistry

Joanne B. Emerson<sup>1\*</sup> (joanne.b.emerson@gmail.com), **Ruth K. Varner**<sup>2</sup>, Donovan H. Parks<sup>3</sup>, Joel E. Johnson<sup>2</sup>, Rebecca B. Neumann<sup>4</sup>, Ben J. Woodcroft<sup>3</sup>, Caitlin M. Singleton<sup>3</sup>, Rodney Tollerson<sup>1</sup>, Martin Wik<sup>5</sup>, Akosua Owusu-Dommey<sup>6</sup>, Morgan Binder<sup>6</sup>, Nancy L. Freitas<sup>6</sup>, **Patrick M. Crill<sup>5</sup>**, **Scott R. Saleska<sup>6</sup>**, **Gene W. Tyson<sup>3</sup>**, and **Virginia Rich<sup>1</sup>** 

<sup>1</sup>The Ohio State University, Columbus; <sup>2</sup>University of New Hampshire, Durham; <sup>3</sup>University of Queensland, Australia; <sup>4</sup>University of Washington, Seattle; <sup>5</sup>University of Stockholm, Sweden; and <sup>6</sup>University of Arizona, Tucson

### http://isogenie.osu.edu

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

Abstract: Thawing permafrost peatlands are likely to provide a positive feedback to climate change via methane and carbon dioxide emissions. In particular, Arctic post-glacial lakes in these landscapes have been identified as a significant source of methane to the atmosphere, largely through ebullition (bubbling) of microbially produced methane from the sediments. Lakes and ponds are estimated to contribute approximately two-thirds of total natural methane emissions above 50° N latitude, and post-glacial lakes (formed by glacially carved microtopography) cover more than four times the aerial extent of more intensively-studied thermokarst lakes (*i.e.*, those associated with the physical collapse of Arctic ground as it thaws). Both the production and oxidation of methane are under microbial control in lakes, yet the microbiota of Arctic post-glacial lake sediments have not been examined. Here we combined microbial (metagenomic and amplicon), isotopic, and geochemical characterizations across four sediment cores from two post-glacial lakes to identify key microbes contributing to carbon loss. Aerobic and anaerobic methane oxidation, methylotrophy, and methanogenesis were identified as controls on methane loss and correlated with shifts in sediment carbon geochemistry. Reconstructed population genomes revealed the capacity for carbon transformations differing from previous lineage-specific metabolic predictions, including methylotrophy in Candidate Phylum Aminicenantes and fermentation without the capacity for methanogenesis in the Methanomassiliicoccaceae. Collectively, results support the potential for significant microbial controls on carbon cycling in post-glacial Arctic lakes and reveal unexpected roles for abundant microbial lineages in key carbon transformations.

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

### The IsoGenieDB: An Integrated Solution to Cross-Disciplinary Data Management

Benjamin Bolduc<sup>1</sup>\* (bolduc.10@osu.edu);-Mike Palace<sup>2</sup>, Ruth Varner<sup>2</sup>, Gene Tyson<sup>3</sup>, Jeff Chanton<sup>4</sup>, Patrick Crill<sup>5</sup>, **Scott R. Saleska**<sup>6</sup>, Malak M. Tfaily<sup>7</sup>, Carmody K. McCalley<sup>8</sup>,-Matthew B. Sullivan<sup>1</sup> and **Virginia Rich**<sup>1</sup>

<sup>1</sup>Ohio State University, Columbus; <sup>2</sup>University of New Hampshire, Durham; <sup>3</sup>University of Queensland, Australia; <sup>4</sup>Florida State University, Tallahassee; <sup>5</sup>University of Stockholm, Sweden; <sup>6</sup>University of Arizona, Tucson, <sup>7</sup>Pacific Northwest National Laboratory, Richland and <sup>8</sup>Rochester Institute of Technology, Rochester.

https://isogenie-web-dev.asc.ohio-state.edu

http:// isogenie.osu.edu

**Project Goals:** The objective of the IsoGenie2 Project is to discover how microbial communities mediate the fate of carbon in thawing permafrost landscapes under climate change. We are engaged in a systems approach integrating (a) molecular microbial and viral ecology, (b) molecular organic chemistry and stable and radiocarbon isotopes, and (c) state-of-the-art modeling, along an interconnected chronosequence of permafrost thaw and post-glacial lakes in subarctic Sweden. Data management across disciplines and scales is key to success.

Abstract Text: Understanding systems-scale data – from geochemistry measurements of isotopic data, microbial ecological and biochemistry to vegetation surveys and climate data – is essential to identifying statistical relationships, and to model and predict biogeochemical cycling in any system. Integration of and access to this systems-scale data presents an exceptional challenge in data management, exacerbated by technological advancements in all areas of environmental data generation. This includes non-traditional investigative methods such as highresolution drone imaging. The IsoGenieDB – a graph-based database with a web portal – provides an efficient means of not only organizing and storing the highly diverse data types that exist within the IsoGenie Project, but also offers exceptional querying capabilities that can leverage cutting-edge network-based analytics. The results of these analytical methods can reveal underlying or overarching patterns of interaction typically invisible within ecosystems data due to incomplete integration of the relationships within and between data types. The IsoGenieDB therefore seeks to directly solve the often paradoxical challenge of integrating increasingly complex and large datasets while simultaneously allowing end-users the ability to easily access and query against stored data generated by the IsoGenie Team. This data includes extensive, multi-year datasets including meta-omics, high resolution chemical and geochemical data,

vegetation "ground cover," satellite and unmanned aerial imagery and many other data types, spanning temporal (seconds to decades) and spatial (depth profiles, and site locations) data series. This data is augmented by the 100-year historical meteorological measurements of the Abisko Scientific Research Station. A web interface allows non-technical users the ability to query against the database for specific datatypes, temporal or depth-associated information, geospatial grouping or a combination of the above. More sophisticated queries – including network-based methods and pattern recognition algorithms – can be performed by accessing the underlying database server directly. The web interface provides both private and public access, with private data shared among all members of the IsoGenie group, with public data release of bundled cross-disciplinary datasets upon publication.

Taken together, the IsoGenieDB is a novel data management and data interrogation system developed from the ground up, built to more accurately represent the data transformations within all levels of an ecosystem, regardless of data type or origin, e.g. temporal, spatial, geochemistry, imagery, abundances, minutes, hours, etc. We believe the IsoGenieDB will provide (a) a platform for IsoGenie members to explore their data through data and relationships provided by themselves and other collaborators to more fully describe their system and address their particular focuses and goals, and (b) a model for solving such data management challenges in other systems-scale projects, where ease of integration, in addition to data integration itself, can fundamentally change how scientists view interdisciplinary work and approach problems.

*Funded by the DOE Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research, grant DE-SC0010580.* 

### Soil viral ecology along a permafrost thaw gradient

Joanne B. Emerson<sup>1</sup>, Jennifer R. Brum<sup>1</sup>, Simon Roux<sup>1</sup>, Benjamin Bolduc<sup>1</sup>, Ben J. Woodcroft<sup>2</sup>, Caitlin M. Singleton<sup>2</sup>, Joel A. Boyd<sup>2</sup>, Suzanne B. Hodgkins<sup>3</sup>, Rachel M. Wilson<sup>2</sup>, Gareth Trubl<sup>1</sup>, Ho Bin Jang<sup>1</sup>, Patrick M. Crill<sup>4</sup>, Jeffrey P. Chanton<sup>3</sup>, Scott R. Saleska<sup>5</sup>, Gene W. Tyson<sup>2</sup>, Virginia I. Rich<sup>1</sup>, and **Matthew B. Sullivan<sup>1</sup>** (mbsulli@gmail.com)

<sup>1</sup>The Ohio State University, Columbus, OH, <sup>2</sup>Australian Center for Ecogenomics, Brisbane, Australia, <sup>3</sup>Florida State University, Tallahassee, FL, <sup>4</sup>Stockholm University, Stockholm, Sweden, <sup>5</sup>University of Arizona, Tucson, AZ USA

Project URL: https://isogenie-web-dev.asc.ohio-state.edu, http://isogenie.osu.edu

**Project Goals:** The objective of the IsoGenie2 Project is to discover how microbial communities mediate the fate of carbon in thawing permafrost landscapes under climate change. We are engaged in a systems approach integrating (a) molecular microbial and viral ecology, (b) molecular organic chemistry and stable and radiocarbon isotopes, and (c) state-of-the-art modeling, along an interconnected chronosequence of permafrost thaw and post-glacial lakes in subarctic Sweden.

**Abstract:** High-latitude environments are disproportionately impacted by climate change with decadal declines now known in sea ice extent and thickness and permafrost thaw, concomitant with increases in greenhouse gas emissions. Numerous efforts to evaluate the role of microbes in high-latitude soils have revealed that they significantly alter these ecosystem features. However, in spite of large viral impacts on microbial dynamics and metabolic outputs in other environments (e.g., oceans), virtually no knowledge exists for how viruses impact soil microbes and ecosystems. Here we analyzed 201 bulk soil metagenomes along an Arctic permafrost thaw gradient to identify viral genomes and large genome fragments as a basis for developing a population-based ecological understanding of viruses in these systems. A total of 1,907 viral genomes were recovered and taxonomically classified, which nearly triples known genera of prokaryotic soil viruses. Metatranscriptomic analyses of a subset (n=26) of these samples revealed that genes from approximately two-thirds (61%) of these viruses were expressed, suggesting that active infections were captured. Both active and overall viral community composition differed along the thaw gradient, concomitant with shifts in host community composition and biogeochemistry. In silico host prediction linked 35% of the viruses to co-occurring host populations, including biogeochemically relevant microbes such as methanogens and methanotrophs. Further, they enabled unprecedented field-based, lineagespecific virus:host abundance estimates, which revealed dynamic infections across the permafrost thaw gradient for select microbial lineages. Together, these data suggest that viruses are integral to modulating climate-critical peatland soil biogeochemistry.

*Funded by the DOE Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research, grant DE-SC0010580.* 

### Pyramiding genes and alleles for improving energy cane biomass yield

Ching Man Wai<sup>1</sup>, Tyler Jones<sup>2</sup>, Chifumi Nagai<sup>2</sup>, Qingyi Yu<sup>3</sup>, **Ray Ming**<sup>1</sup>\* (rming@life.uiuc.edu)

<sup>1</sup>University of Illinois at Urbana-Champaign, Urbana, Illinois; <sup>2</sup>Hawaii Agriculture Research Center, Kunia, Hawaii; and <sup>3</sup>Texas A&M AgriLife Research, Dallas, Texas

### http://www.life.illinois.edu/ming/LabWebPage/Home.html

Project Goals: Our long term goal is to establish a new paradigm to accelerate energy cane breeding programs and maximize the biomass yield for biofuel production. Our specific objectives are: (1) Phenotyping extreme segregants of the F2 population for exploring the molecular basis of high biomass yield from transgressive segregation; (2) Mapping genes affecting biomass yield by transcriptome sequencing of the extreme segregants; (3) Identifying differentially expressed genes and alleles through analyses of transcriptomes of extreme segregants from the F2 population; (4) Developing gene- and allele-specific markers for implementation of marker-assisted selection in energy cane breeding programs.

As a C4 plant, sugarcane/energy cane has been recognized as one of the world's most efficient crops in converting solar energy into chemical energy. Traditional energy cane and sugarcane breeding via interspecific hybridization and backcrossing to *S. officinarum* improved stress tolerance and recovered high sugar and biomass yield. However, this approach reduced the genetic diversity of sugarcane and energy cane breeding materials and limited the potential maximizing biomass yield. We are developing a new paradigm to accelerate energy cane breeding programs and maximize the biomass yield for biofuel production.

The smallest genome in *Saccharum* is the *S. spontaneum* haploid AP85-441 developed by anther culture. It has a genome size of 3.4 Gb, 34% of a typical 10 Gb hybrid genome, and the smallest chromosome number at 2n = 4x = 32. Pacbio single-molecule real-time (SMRT) sequencing technology was used to sequence this haploid genome at 77X coverage. The *S. officinarum* LA Purple has 2n = 8x = 80 with genome size at 7.6 Gb. This genome was also sequenced at 80x using PacBio. Two F1 mapping populations were developed from crosses between double haploid AP83-108 and its pollen source SES208 and from S. *officinarum* LA Purple and *S. robustum* MOL5829. The individual genomes of these two F1 populations were sequenced at 5X coverage using Illumina HiSeq2500. The assembled draft genome of AP85-441 and LA Purple were used as references for SNP calling to map QTLs affecting biomass yield, and for gene expression analysis of extreme segregants.

An  $F_2$  population with 2616 individuals was created from an interspecific cross between LA Purple x MOL5829 (2n = 80, x = 10). This population showed transgressive segregation with high yielding clones substantially exceeding the biomass yield of both parents. Field trial of the extreme segregants was carried out for three years in Hawaii. Biomass yield of the top 10  $F_2$  clones ranged from 71.5 to 122.1 MT/ha in 12 months, and the best performing clone showed 338% yield increase compared to its high yield parent LA Purple. The bottom 10 F2 clones had estimated biomass yield ranging from 3.4 to 8.2 MT/ha in 12 months, the worst performing clone showed yield decrease to only about 10% of LA Purple yield.

Leaf and stem internodes of high and low biomass F2 extreme segregants were used for RNAseq to decipher the molecular mechanism of rapid plant growth and dry weight accumulation. Gene Ontology terms involved in cell wall metabolism and carbohydrate catabolism were enriched among 3,274 differentially expressed genes between high and low biomass groups. Specifically, up-regulation of cellulose metabolism, pectin degradation and lignin biosynthesis genes were observed in the high biomass group, in conjunction with higher transcript levels of callose metabolic genes and the cell wall loosening enzyme expansin. Furthermore, UDPglucose biosynthesis and sucrose conversion genes were differentially expressed between the two groups. A positive correlation between stem glucose, but not sucrose, levels and dry weight was detected. We thus postulated that the high biomass sugarcane plants rapidly convert sucrose to UDP-glucose, which is the building block of cell wall polymers and callose, in order to maintain the rapid plant growth required for biomass accumulation. The gene interaction of cell wall metabolism, hexose allocation and cell division contributes to biomass yield, expanding our understanding at the molecular level required for energy cane breeding and engineering. Time-series metagenomics of experimentally warmed Alaskan tundra and Oklahoma temperate soils enables fine-resolution assessment of belowground C cycling feedbacks to climate change

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

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

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

**Project goals**: The overall goal of this project is to advance system-level predictive understanding of the feedbacks of belowground microbial communities to multiple climate change factors and their impacts on soil C cycling processes. Regarding this goal, we are pursuing the following objectives: (i) To improve our understanding of soil microbial communities indigenous to temperate and tundra ecosystems through whole-community analysis, and through the classification of novel taxa recovered directly from metagenomes and/or obtained in pure culture; (ii) To determine the microbiological basis underlying temperature sensitivity of soil organic matter decomposition; (iii) To determine the ubiquity of recovered bacterial populations across large geographic regions spanning several hundred kilometers; and (iv) To develop integrated bioinformatics and modeling approaches to scale information across different organizational levels towards predictive understanding of ecosystem responses to multiple climate change factors, which will be collaborated and integrated with K-Base.

**Abstract:** Soils contain more carbon (C) in the form of soil organic matter (SOM) than both aboveground plant and atmospheric pools combined. Higher land temperatures are expected to cause the release of considerable amounts of  $CO_2$  and  $CH_4$  to the atmosphere, primarily through the stimulation of microbial-mediated turnover of SOM. However, the direction, magnitude, and underlying basis of soil ecosystem feedbacks to climate warming remain poorly understood. To this end, we have investigated microbial communities from Alaskan tundra permafrost (AK) and Oklahoma temperate grassland (OK) soils, both of which have been experimentally warmed *insitu* (2 to 4°C above ambient temperature in the field) and under laboratory conditions (15 and 25°C). By combining well-replicated soil metagenomes with continuous environmental monitoring, respiration data, and soil measurements, we hope to gain an improved understanding of microbial responses to climate warming, particularly those involved in the turnover of (SOM) and release/sequestration of greenhouse gases.

Metagenomic datasets representing soils sampled after 1.5 and 5 years of field warming yielded near-complete representation of microbial community 'sequence richness' at AK and OK sites, and revealed that OK soil communities are an order of magnitude more diverse than their tundra counterparts. Sequence assembly and binning techniques allowed for the recovery of several near-complete bacterial population genomes from both ecosystems, most of which represent previously uncharacterized taxa, allowing for prediction of their metabolic lifestyles,

regional prevalence (based also on publically-available datasets from nearby locations), and response to elevated temperatures. Several of the recovered AK populations were regionally ubiquitous, e.g., found at several locations ~100-530 kilometers apart (Johnston et al., 2016). Warming favored bacterial populations encoding diverse metabolisms for recalcitrant and labile SOM degradation, including abundant members of the community (0.25-2% of total), which increased by 30-100% of their original abundance after just 5 years of field warming. Wholecommunity assessment of 5-year AK field samples also revealed a uniform increase in many SOM catabolism pathways coincident with warming, including those for both the labile and recalcitrant fractions of SOM (Johnson et al., in preparation). These results were also consistent with GeoChip functional gene analysis and observations of increased ecosystem respiration reported at an earlier experimental phase (Xue et al., 2016). 5 years of experimental warming at the OK field site altered the functional composition of microbial communities ( $\beta$ -diversity distances) and increased microbial community sequence complexity/diversity.

Metagenomic sequencing and assembly of lab-incubated soils resulted in the recovery of several hundred-population genomes, which collectively represented 25-75% of the total soil community. This data allowed for more resolved associations between SOM-turnover and the community composition (i.e., members responsible for these activities) to be identified. For instance, a correlation coefficient of 0.6 was obtained by relating the presence and abundance of population genomes to soil C and respiration measurements. Also, a correlation coefficient of 0.7 was obtained by relating specific bacterial populations involved in major N-cycling processes (N-fixation, denitrification, etc.) to measured soil N. To further verify these results and obtain model organism for future studies, >600 bacterial cultures have been recovered from AK longterm laboratory incubation soils using dilute nutrient, minimal salt and soil extract media under reduced oxygen stress. The 16S rRNA genes from these isolates have been matched against assembled bins for identification, and Biolog plates are beings used to reveal their associated metabolisms. These efforts have recovered, for example, an Acidobacteria (Terriglobus sp.) organism capable of N-fixation that likely plays a N-cycling role in tundra soils, which are more severely N-limited relative to temperate grassland, constraining the rate of SOM-turnover. The recovery of several N-fixing isolates also complements recent efforts to describe nifH-harboring community members indigenous to the Alaska tundra ecosystem (Penton et al., 2016).

### References

- [1] E. R. Johnston, L. M. Rodríguez-R, C. Luo, M. M. Yuan, L. Wu, Z. He, E. A. Schuur, Y. Luo, J. M. Tiedje, J. Zhou, and K. T. Konstantinidis. (2016). Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska tundra ecosystem. Frontiers in Microbiology. 7:579 doi:10.3389/fmicb.2016.00579
- [2] K. Xue, M. M. Yuan, Z. J. Shi, Y. Qin, Y. Deng, L. Cheng, L. Wu, Z. He, J. D. Van Nostrand, R. Bracho, S. Natali, E. A. G. Schuur, C. Luo, K. T. Konstantinidis, Q. Wang, J. R. Cole, J. M. Tiedje, Y. Luo, and J. Zhou. (2016). Rapid microbial feedbacks reveal vulnerability of tundra soil carbon to climate warming. Nature Climate Change. 6:595-600 doi:10.1038/nclimate2940
- [3] C. R. Penton, C. Yang, L. Wu, Q. Wang, J. Zhang, F. Liu, Y. Qin, Y. Deng, C. L. Hemme, T. Zheng, E. A. G. Schuur, J. M. Tiedje, and J. Zhou. (2016). NifH-Harboring Bacterial Community Composition across an Alaskan Permafrost Thaw Gradient. Frontiers in Microbiology. 7:1894. doi:10.3389/fmicb.2016.01894

Funding statement: Our work is supported by the U.S. DOE Office of Science, Biological and

Environmental Research Division (BER), Genomic Science Program, Award No. DE-SC0010715.

#### Successional Dynamics of Grassland Microbial Communities in Response to Warming, Precipitation Alternation, and Clipping

Xue Guo<sup>1</sup>, Qun Gao<sup>1</sup>, Mengting Yuan<sup>1</sup>, Zhou Shi<sup>1</sup>, Xishu Zhou<sup>1</sup>, **Liyou Wu<sup>1</sup>\*(lwu@rccc.ou.edu)**, **Zhili He**<sup>1</sup>, Joy D. Van Nostrand<sup>1</sup>, Lauren Hale<sup>1</sup>, Daliang Ning<sup>1</sup>, Aifen Zhou<sup>1</sup>, Jiajie Feng<sup>1</sup>, Ying Fu<sup>1</sup>, Chang-Gyo Jung<sup>1</sup>, Tong Yuan<sup>1</sup>, Xuanyu Tao<sup>1</sup>, **Jim Cole<sup>2</sup>**, **James M. Tiedje<sup>2</sup>**, **Edward A.G. Schuur<sup>3</sup>**, **Konstantinos Konstantindis<sup>4</sup>**, **Yiqi Luo<sup>1</sup>**, and **Jizhong Zhou<sup>1</sup>** 

<sup>1</sup>Institute for Environmental Genomics and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK; <sup>2</sup>Center for Microbial Ecology, Michigan State University, East Lansing, MI; <sup>3</sup>Center for Ecosystem Sciences and Society, Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ; <sup>4</sup>Center for Bioinformatics and Computational Genomics and School of Biology, Georgia Institute of Technology, Atlanta, GA

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

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

As a part of the integrated project, here we present results from; i) field experiments established in 2008 in a temperate grassland of central Oklahoma to reveal the influence of elevated temperature, altered precipitation and plant biomass clipping on long-term succession of plant and microbial communities, and ii) modeling ecosystem C pools and C flux under climate warming.

Long-term succession of microbial communities. To understand successional dynamics of microbial communities in response to warming, clipping, and altered precipitation, 264 annual soil samples from 2009 to 2014 were analyzed by sequencing of 16S rRNA genes for bacteria and archaea, ITS regions for fungi, and 18S rRNA genes for other micro-eukaryotes excluding fungi, and by functional gene arrays (GeoChip 5.0). Our analyses indicated that OTU richness for all three marker genes significantly decreased in warmed plots compared to ambient plots, which remained true when clipping and reduced precipitation were coupled with warming. Among all twelve treatments, OTU richness under clipping with double precipitation was the highest for all three marker genes. To understand temporal patterns and their underlying mechanisms within the context of climate change factors, species-time relationships (STRs) were estimated. By examining different soil microbial populations, bacteria, fungi and microeukaryotes in this multifactor global change experiment, we demonstrated that all soil microbial populations exhibit significant (P < 0.05) STRs under various global changes. Secondly, our results showed warming significantly (P < 0.05) stimulated microbial temporal scaling rates (STR exponent) across different organization scales. This is the first demonstration that climate warming has significant impacts on temporal scaling of microbial communities and significantly (P < 0.05) promotes temporal scaling rates of all microbial populations even under the context of various global changes. All of these results indicated that warming plays a dominant role in accelerating temporal scaling rates of soil microbial communities.

Dissimilarity and permutational analyses of variance indicated that global change factors including warming, clipping, half precipitation and double precipitation significantly (P < 0.05) altered the taxonomic composition and phylogenetic structure of different microbial populations. Among these global change factors, climate warming played a dominant role in altering the taxonomic composition and

phylogenetic structure of soil microbial communities in the long-term experiment. Second, since previous "snapshot" studies have well-documented the sensitivities of soil microbial communities to various global change factors, it is anticipated that warming and related climate factors have important effects on the temporal successional patterns of soil microbial communities. However, little is known about how microbial communities develop over time under multiple global change conditions. Our results showed that global change factors and their interactions differently shifted the successional patterns of soil microbial communities. To the best of our knowledge, this is the first demonstration that various global change factors have significant (P < 0.05) impacts on temporal succession patterns of soil microbial communities. Importantly, among these factors, climate warming played a dominant role in accelerating divergent succession of all soil microbial communities as evidenced that experimental warming enhanced microbial temporal divergences under the context of various global changes. These results are also consistent with some recent studies showing that temperature plays a primary role in shaping microbial communities. Third, global changes induced divergent succession of soil microbial communities significantly (P < 0.05) affected ecosystem function processes. Aboveground net plant productivity (ANPP), gross primary productivity (GPP) and soil respirations exhibited significant (P < 0.05) correlation with the changes of microbial community structures across different microbial populations. This could suggest that the global changes induced divergent successions of soil microbial communities have important effects on the terrestrial C and nitrogen (N) cycling. Microbial functional traits under climate changes. As revealed by GeoChip hybridization, a number of microbial functional groups important for C decomposition and N cycling were differentially impacted by warming across 6 years. In the first two years, warming increased (P < 0.05) the relative abundance of genes involved in C degradation, N cycling and phosphorus (P) utilization, while in the third year, warming had no significant effects on these genes. Then, in the most recent two years, these genes decreased in response to the warming treatment (P < 0.05). Mantel test revealed that very strong correlations were observed between community structure and ecosystem functional processes, such as plant GPP and soil total respiration, and the functional gene groups involved in C degradation, N cycling, P utilization and S cycling (P < 0.05). Remarkably, soil heterotrophic respiration (from soil microbes) specifically showed strong correlations with the genes involved in degrading recalcitrant C, including vanillin, lignin and cellulose, which could suggest that the increased heterotrophic respiration may be contributed to the degradation of recalcitrant C. Secondly, precipitation alternations significantly (P < 0.05) affected the phylogenetic compositions of bacterial and fungal communities, but not their functional gene structures. In addition, annual clipping significantly (P < 0.05) changed the abundance of genes involved in C decomposition and N cycling. Cumulative annual clipping effects on functional genes were observed over time. From the second year of operation, significant (P < 0.05) increases of the relative abundances of genes involved in degradation of both labile and recalcitrant C were observed. However, in the last two years, only the genes involved in the degradation of recalcitrant C increased (P < 0.05). Other genes involved in nutrient-cycling processes including N cycling and P utilization also increased (P < 0.05) by annual clipping.

**Modeling of soil C pools and flux.** To explore the effects of warming on the temperature sensitivity of soil respiration (Q<sub>10</sub>), we conducted regression analysis using  $R_s = ae^{bT}$ , where  $R_s$  is soil respiration, T is soil temperature, coefficient a is the intercept of soil respiration when temperature is zero, and coefficient b represents the temperature sensitivity of soil respiration. Warming significantly (P < 0.05) reduced Q10 in every experimental year (from 2009 to 2014), indicating that respiratory acclimation has occurred after 6 years of warming. The variation of Q<sub>10</sub> among different years was also observed with lower Q<sub>10</sub> in 2012 and 2014. This was consistent with the year variation of precipitation, which showed lower precipitation in the year 2012 and 2014. Such results indicate that precipitation can influence the warming effect on respiratory acclimation. To understand the whole ecosystem C dynamic feedback to warming, and the microbial traits affecting this feedback, terrestrial ecosystem (TECO) modeling was performed to simulate different C pools as well as soil respiration under control and warming treatment. Microbial functional within-group similarity (n=4) of each year was then incorporated into model simulation.

Warming significantly (P < 0.001) decreased soil C storage compared with control treatments. This result was consistent with our experimental data that warming significantly (P < 0.05) increased heterotrophic respiration, suggesting warming increased microbial decomposition of soil C, thus could potentially decrease soil C. After incorporating microbial function gene information into the model, the simulation of heterotrophic respiration improved by 18.9%, suggesting that it is important to include microbes global change models.

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

### Ensemble cell-wide kinetic modeling of anaerobic organisms to support fuels and chemicals production

Satyakam Dash<sup>1\*</sup>(sud25@psu.edu), Ali Khodayari<sup>1</sup>, Jilai Zhou<sup>2</sup>, Evert K. Holwerda<sup>2</sup>, Daniel G. Olson<sup>2</sup>, Junyoung O. Park<sup>3</sup>, David Emerson<sup>3</sup>, Benjamin Woolston<sup>3</sup>, Po-wei Chen<sup>4</sup>, James C. Liao<sup>4</sup>, Gregory Stephanopoulos<sup>3</sup>, Lee R. Lynd<sup>2</sup>, and **Costas D. Maranas**<sup>1</sup>

<sup>1</sup>The Pennsylvania State University; <sup>2</sup>Thayer School of Engineering at Dartmouth College; <sup>3</sup>Massachusetts Institute of Technology; <sup>4</sup>University of California, Los Angeles.

http://maranas.che.psu.edu/; https://engineering.dartmouth.edu/lyndlab/; http://bamel.scripts.mit.edu/gns/; http://www.seas.ucla.edu/~liaoj/

Project Goals: The goal of the project is to systematically build dynamic metabolic models of *Clostridium thermocellum* and *Clostridium ljungdahlii* using Ensemble Modeling (EM) paradigm through integrative metabolomic and fluxomic data. These models will be instrumental in exploring genetic interventions for overproduction of biofuel products.

Clostridia possess broad and flexible systems for substrate utilization: Clostridium thermocellum can metabolize cellulose, and *Clostridium ljungdahlii* and *Moorella thermoacetica* can fix CO<sub>2</sub> to produce biofuels. However, their metabolism remains poorly characterized. Here we develop kinetic models of clostridia using the Ensemble Modeling paradigm, which require curated genome-scale metabolic (GSM) models as its foundation. For C. thermocellum, we constructed a second-generation GSM model (iCth446) with 446 genes, 598 metabolites and 660 reactions, along with gene-protein-reaction associations by updating cofactor dependencies, maintenance (GAM and NGAM) values and resolving elemental and charge imbalances. iCth446 model was subsequently used to develop k-ctherm118, a kinetic model of C. thermocellum's central metabolism containing 118 reactions and 92 metabolites with cellobiose as the carbon source under anaerobic growth condition. k-ctherm118 encompasses the cellobiose degradation pathway, glycolysis/ gluconeogenesis, the pentose phosphate (PP) pathway, the TCA cycle, pyruvate metabolism, anaplerotic reactions, alternative carbon metabolism, nucleotide salvage pathway, along with all biomass precursors and 22 substrate level regulatory interactions extracted from BRENDA. The kinetic model parameters were estimated by simultaneously imposing fermentation yield data in lactate, malate, acetate and hydrogen production pathways for 19 measured metabolites spanning a library of 19 distinct single and multiple gene knockout mutants along with 18 intracellular metabolite concentration data for a  $\Delta gldh$  mutant and ten experimentally measured Michaelis-Menten kinetic parameters. k-ctherm118 captures significant metabolic changes including effect of nitrogen limitation on C. thermocellum's metabolism as well as impact of ethanol stress on intracellular metabolite pools due to downregulation of fermentation pathways. Robustness analysis of k-ctherm118 alludes to the presence of a secondary activity of ketol-acid reductoisomerase and possible regulation by valine and/or leucine pool levels.

For the case of *C. ljungdahlii*, we constructed a core metabolic model composed of 79 reactions and 63 metabolites based on a GSM published by Nagarajan *et al.* [1] and including 41 putative substrate-level regulations from BRENDA. The core model accounts for glycolysis, Wood-Ljungdahl pathway, pentose-phosphate pathway and production pathway for major products such as lactate, acetate and ethanol. The TCA cycle is incomplete with missing succinate dehydrogenase, 2-oxoglutarate synthase, and succinyl-CoA synthetase [2]. We have used the wild-type data from the Stephanopoulos group to generate the initial ensemble. The experimental

data included growth rate and the rate of formate, acetate, and ethanol production with CO<sub>2</sub> and/or hexose as carbon source. *C. ljungdahlii* shares its unique metabolic capabilities including CO<sub>2</sub> fixation with another anaerobe of the class clostridia, *M. thermoacetica*. To this end, the Stephanopoulos group performed <sup>13</sup>C isotope tracing and metabolic flux analysis to obtain flux ratios between key metabolic pathways, which further constrain the feasible parameter space. The flux analysis revealed that glucose and CO<sub>2</sub> can be consumed simultaneously with CO<sub>2</sub> generating majority of the TCA cycle intermediates. In addition, integrative metabolomic and fluxomic assays under various culture conditions will be performed on *C. ljungdahlii* and *M. thermoacetica* by the Stephanopoulos group and on *C. thermocellum* by the Liao group, which will be used to parametrize and refine the kinetic models. The constructed kinetic models will ultimately aid in exploring the full metabolic capability of clostridial biofuel production.

*The work was supported by the genomic science grant from Department of Energy, USA (grant # # DE-SC0012377).* 

- 1. Nagarajan, H., et al., *Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii*. Microbial cell factories, 2013. **12**: p. 118.
- 2. Kopke, M., et al., *Clostridium ljungdahlii represents a microbial production platform based on syngas.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(29): p. 13087-92.
- 3. Chang, A., et al., *BRENDA in 2015: exciting developments in its 25th year of existence*. Nucleic Acids Research, 2015. **43**(D1): p. D439-D446.



Figure 1: Core metabolic map of *Clostridium thermocellum* and *Clostridium ljungdahlii*. The green arrows represent the metabolite concentrations which are experimentally measured.

#### k-ecoli457: A genome-scale Escherichia coli kinetic metabolic model satisfying

#### flux data for multiple mutant strains

Ali Khodayari\* (auk241@psu.edu) and Costas D. Maranas

Department of Chemical Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.

http://www.maranasgroup.com

Project Goals: The goal of this effort is to construct a genome-scale kinetic model of *Escherichia coli* metabolism by making use of Ensemble Modeling (EM) concepts. Model parameterization is carried out using multiple flux datasets for different substrates and growth (aerobic vs. anaerobic) conditions.

Kinetic models of metabolism at a genome-scale that faithfully recapitulate the effect of multiple genetic interventions would be transformative in our ability to reliably design novel overproducing microbial strains. Here, we introduce k-ecoli457, a genome-scale kinetic model of Escherichia coli metabolism that satisfies fluxomic data for wild-type and 25 mutant strains under different substrates and growth conditions. Model k-ecoli457 contains 457 reactions, 337 metabolites and 295 substrate-level regulatory interactions. Parameterization is carried out using a genetic algorithm by simultaneously imposing all available fluxomic data (about 30 measured fluxes per mutant). Model predictions were tested against multiple experimentally measured datasets that were not used during model parameterization. These included (i) 898 steady-state metabolite concentrations for twenty of the mutant strains [1-4], (ii) 234 Michaelis-Menten constants (185  $K_m$  and 49  $k_{cat}$  values) from BRENDA and EcoCyc, and (iii) 320 literature reported product yields for designed strains covering 24 different bioproducts. Comparisons revealed that 66% of the predicted metabolite concentrations as well as 51% and 63% of the estimated  $K_m$  and  $k_{cat}$  values, respectively, are within the experimentally reported ranges. Notably, the average relative error of k-ecoli457 predictions for the product yield in 129 out of 320 designed strains is within 20% of the measured values. Stoichiometric model based techniques such as FBA, minimization of metabolic adjustment (MOMA) or maximization of product yield were within 20% of the experimentally reported yield for only 16, 18 and 65 of the designed strains, respectively. Overall, the predicted product yields by k-ecoli457 achieve significantly higher value of correlation with experimental data (i.e., Pearson's correlation coefficient of 0.84) than FBA, MOMA or maximization of product yield (i.e., 0.18, 0.37 and 0.11, respectively). These

results quantitatively demonstrate that k-ecoli457 can reliably be used to predict genetically perturbed *E. coli* phenotypes under different growth conditions with a substantially higher accuracy than any other earlier modeling effort (k-ecoli457 is available for download at <u>http://www.maranasgroup.com</u>).



**Figure** A pictorial representation of the k-ecoli457 model of *E. coli* metabolism. Red X's denote the location of reaction deletions in the mutant datasets. Reactions in the previously developed core model [5] are shown in grey (no flux data) and blue (with flux data) while the additional reactions in k-ecoli457 are shown in green.

*The work was supported by the genomic science grant from Department of Energy, USA (grant# DE-SC0012377).* 

### References

- 1. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A *et al*: **Multiple high-throughput analyses monitor the response of E. coli to perturbations**. *Science* 2007, **316**(5824):593-597.
- 2. Kabir MM, Ho PY, Shimizu K: Effect of ldhA gene deletion on the metabolism of Escherichia coli based on gene expression, enzyme activities, intracellular metabolite concentrations, and metabolic flux distribution. *Biochem Eng J* 2005, **26**(1):1-11.
- 3. Zhao J, Baba T, Mori H, Shimizu K: **Global metabolic response of Escherichia coli to gnd or zwf gene-knockout, based on 13C-labeling experiments and the measurement of enzyme activities**. *Applied microbiology and biotechnology* 2004, **64**(1):91-98.
- 4. Zhao J, Shimizu K: Metabolic flux analysis of Escherichia coli K12 grown on 13C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *Journal of biotechnology* 2003, **101**(2):101-117.
- Khodayari A, Zomorrodi AR, Liao JC, Maranas CD: A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data. *Metabolic engineering* 2014, 25:50-62.

### Robust Non-Oxidative Glycolysis in Escherichia coli

Paul P. Lin<sup>1</sup> (phlin@ucla.edu), Alec J. Jaeger,<sup>1</sup> and James C. Liao<sup>1</sup>

<sup>1</sup>Department of Chemical and Biomolecular Engineering, University of California, Los Angeles

**Project Goals:** This project seeks to construct strains of *Escherichia coli* to rely solely on Nonoxidative glycolysis (NOG) for glucose catabolism during aerobic growth, fundamentally rewiring the central metabolism to favor carbon conservation in acetyl-CoA biosynthesis. This work will also engineer the constructed strains for energy applications. Specifically, we aim to produce an advanced biofuel, n-butanol, with close to complete carbon conservation. Hydrogen or formate will be provided to supply reducing equivalents needed to drive the NADH dependent pathway.

Acetyl-coenzyme A (acetyl-CoA) is a two carbon metabolite and important metabolic precursor to a variety of industrially relevant compounds including biofuels. An ultimate limitation of acetyl-CoA derived biochemical production is the inherent carbon loss when forming acetyl-CoA. Most organisms use some glycolytic variation, commonly the Embden-Meyerhof Pathway (EMP), to initially degrade sugar into pyruvate. Pyruvate, a C3 metabolite, is then decarboxylated to form acetyl-CoA, losing carbon to the environment. This decarboxylation limits the carbon yield to only two molecules of acetyl-CoA from one molecule of hexose, thus inhibiting the economics of any associated bioprocess. A synthetic sugar catabolism pathway, termed non-oxidative glycolysis (NOG), was recently developed to address this problem, as it uses a combination of phosphoketolase dependent cleavage of sugar phosphates and a carbon rearrangement cycle to directly generate three C2 units per hexose in a redox neutral manner. To further expand the applications using NOG, an Escherichia coli strain was constructed to rely solely on NOG for sugar catabolism in this work. Therefore, the resulting strain offers significant potential to be engineered for the production of a variety of acetyl-CoA derived compounds. To implement NOG as a growth pathway, all native sugar degradation pathways, including the EMP, ED and methylglyoxal bypass, were removed, eliminating the cell's ability to grow on sugar as a sole carbon source. In addition, the glyxoylate shunt and gluconeogenesis pathways, which are necessary for the production of essential metabolites using NOG, were upregulated. Following the chromosomal integration of phosphoketolase, an in vitro pathway assay was used to identify limiting core enzymes for C2 metabolite production in cell lysates. Following the overexpression of these enzymes as well as a PTS independent glucose uptake system on an inducible plasmid, the engineered strain was able to develop an improved growth phenotype relative to an uninduced control in minimal glucose media supplemented with acetate. A time course analysis of the growth cultures verified that the induced strain was able to consume glucose while the uninduced strain could not. Additionally, under fermentative conditions this strain was able to produce acetate from glucose at yields approaching the theoretical maximum. Future work on the project will involve understanding the necessity of acetate supplementation for growth in minimal media in order to engineer the cell to grow solely on sugar.

Grant number: U.S. Department of Energy DE-SC0012384

### Protein Acylomes in Fuel-Producing *E. coli* – Changes in Posttranslational Modifications with Different Carbon Sources

David Christensen,<sup>1</sup> Kori D. Dunn,<sup>2</sup> Alexandria K. D'Souza,<sup>3</sup> Christopher V. Rao,<sup>2</sup> Bradford W. Gibson<sup>3</sup>, Alan J. Wolfe,<sup>1</sup> and Birgit Schilling,<sup>3</sup>\* (bschilling@buckinstitute.org)

<sup>1</sup> Loyola University Chicago, Maywood, IL

<sup>2</sup> University of Illinois at Urbana-Champaign, Urbana IL

<sup>3</sup> Buck Institute for Research on Aging, Novato, CA

Project Goals: The goal of this project is to gain deeper insights into posttranslational remodeling of engineered microorganisms, and specifically their protein acylomes under different growth conditions. Lysine acetylation, for example is a common post-translational modification that eukaryotes, archaea, and bacteria employ to regulate protein activity. We and others have recently shown that lysine acetylation predominantly targets metabolic enzymes – in fact, most metabolic enzymes are subject to lysine acetylation. Due to the dynamic nature of protein acetylation and deacetylation mechanisms in the cell, lysine acetylation can likely be considered a global mechanism to regulate metabolism in response to their energy and redox status. Here, we are comparing acetylation status under different growth conditions, using different sugar supplementation strategies and different *E. coli* strains. The significance of this work is that it will address a fundamental gap in our understanding of bacterial metabolism and identify new approaches for overcoming the problems associated with the production of advanced biofuels.

In our recent study, we reported that glucose-regulated lysine acetylation (Kac) was predominant in central metabolic pathways and overlapped with acetyl phosphate-regulated acetylation sites (1). We proposed that acetyl phosphate-dependent acetylation across hundreds of proteins is a response to carbon flux that can regulate central metabolism. Since then, we have extended our investigations to examine the effect on both protein expression levels and the alterations in acetylation status across the *E. coli* proteome when it is provided with alternative carbon sources beyond glucose.

Typically, for mass spectrometric studies protein fractions are either analyzed on the protein level to assess protein expression changes, or protein fractions are subjected to antibody-based affinityenrichment of acetyllysine-containing peptides and are then measured by quantitative mass spectrometry. To monitor changes in protein acetylation and expression, we used a novel label-free and data-independent acquisition (DIA or SWATH) approach on a SCIEX TripleTOF 6600 LC/MS system that we have modified for this purpose (2). In any of our acetylation studies, we typically identify >3,000 unique acetylation sites. Subsequent quantification then reveals sites that show statistically significant relative changes in acetylation abundance between different growth conditions.

In this study, we grew wild-type *E. coli* cells in minimal medium supplemented either xylose (Xyl) or glucose (Glc) as the sole carbon source at either 0.4% or 4% sugar. Using SWATH, we identified hundreds of acetylation sites in more than 150 proteins that showed robust and statistically significant increases in acetylation when the cells were grown with the larger amount of sugar, independent of the supplemented sugar source (Glc or Xyl). These sites overlap substantially with those that are acetyl-phosphate-sensitive. Together, these results support our hypothesis that acetyl-phosphate-dependent protein acetylation is a response to carbon flux and not exposure to a specific carbon source.

The 'regulated' acetylated proteins included many involved in multiple central metabolic processes, e.g., glycolysis/gluconeogenesis, pyruvate metabolism, and the TCA cycle. A functional annotation analysis for regulated acetylated proteins showed a statistically relevant enrichment for the ontology term "generation of precursor and metabolites and energy." To visualize acetylation changes in the context of these protein pathways, we developed novel visual display algorithms. These results are supportive of the hypothesis that protein acetylation could regulate central metabolism in response to carbon flux.

We have been particularly interested in two acetylation sites of XylA, the first enzyme in xylose catabolism, as we determined regulation both on the protein and acetylation site levels. Most intriguingly, acetylation of two sites (K17 and K381) behaved oppositely when growth on 4% glucose was compared to growth on 4% xylose. Differential regulation is supportive of the hypothesis that acetylation may dictate enzyme function.

#### References

1. B Schilling, D Christensen, R Davis, AK Sahu, LI Hu, A Walker-Peddakotla, DJ Sorensen, B Zemaitatis, BW Gibson, AJ Wolfe. Protein acetylation dynamics in response to carbon overflow in *Escherichia coli*, Molecular Microbiology, **2015**, 98(5), 847-63.

2. MJ Rardin, B Schilling, LY Cheng, BX MacLean, DJ Sorensen, AK Sahu, MJ MacCoss, O Vitek, BW Gibson. MS1 Peptide Ion Intensity Chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving Post Acquisition Analysis of Proteomic Experiments. Mol Cell Proteomics **2015**, 14(9), 2405-19

#### Funding Statement:

This work was supported by the U.S Dept. of Energy, Office of Science, Office of Biological & Environmental Research (PI: Rao, DE-SC0012443), grant title 'The Systems Biology of Protein Acetylation in Fuel-Producing Microorganisms'. We also acknowledge the support of the NIH shared instrumentation program for the use of the SCIEX TripleTOF 6600 mass spectrometer (1S10 OD016281).

### Predicting multiple gene targets for optimal oil production in acetic acid metabolism of *Yarrowia lipolytica* by ensemble modeling

Po-Wei Chen<sup>1</sup>, Gregory Stephanopoulos<sup>2</sup> and James C. Liao<sup>1</sup>

<sup>1</sup>Department of Chemical and Biomolecular Engineering, University of California, Los Angeles

<sup>2</sup>Department of Chemical Engineering, Massachusetts Institute of Technology

#### Abstract

Ensemble modeling (EM) has been applied on lipids overproduction of glucose feeding and acetic acid feeding metabolic pathways in *Yarrowia lipolytica*. Since the optimization of glucose feeding strain has been well explored, our team has focused on identifying gene targets to further increase lipids yield in acetic acid feeding pathway system. EM based strain optimization simulation has provided five gene overexpression (OE) and knockdown (KD) target, respectively. The five overexpression targets are, oxoglutarate dehydrogenase, acetyl-CoA carboxylase, isocitrate dehydrogenase, ATP transport, adenylate kinase, respiration; the five knockdown targets are, isocitrate lyase, malate synthase, succinate dehydrogenase, fumarase, malic enzyme NADH mi and pyruvate transport. Among these single gene targets, overexpressing oxoglutarate dehydrogenase can boost lipids yield to 8% more than the reference state. Recently, a new EM strategy has been applied to further increase lipids yield through multiple gene manipulation/control. All possible combination of double and triple gene OE/KD selected from top 3 single OE and KD targets mentioned above has been simulated and compared with single gene OE/KD results. The result suggested that the following multiple gene control can further increase lipids yield to around 10% above the reference state.

- Overexpress oxoglutarate dehydrogenase + knockdown malate synthase
- Overexpress oxoglutarate dehydrogenase + knockdown isocitrate lyase
- Overexpress oxoglutarate dehydrogenase + knockdown isocitrate lyase & malate synthase

Interestingly, among all the targets including changing one, two and three and four gene activities at a time, the strategies of changing two genes shown highest yield. The quadruple gene control strategy: OE oxoglutarate dehydrogenase + OE acetyl-CoA carboxylase + KD isocitrate lyase + KD malate synthase shown similar yield to OE oxoglutarate dehydrogenase. The triple gene control strategy: OE acetyl-CoA carboxylase + KD isocitrate lyase + KD malate synthase even shown lower lipids yield as 7%, which is lower than the best single gene control strategy. Overall, the EM predicted result not only shown the next potential acetic acid feeding strain(s), but also demonstrated the value for aiding designs of sophisticated multiple gene manipulation.

### Amino Acids Are Preferred Over Glucose and Other Sugars in *Escherichia Coli* by a Novel Mechanism of Carbon Catabolite Repression

James S Orr<sup>1</sup>\* (jsorr3@illinois.edu), David Christensen<sup>2</sup>, Alan J Wolfe,<sup>2</sup> and **Christopher V Rao**<sup>1</sup>

<sup>1</sup>University of Illinois at Urbana-Champaign; <sup>2</sup>Loyola University Chicago

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

When we measure protein acetylation in *E. coli*, we routinely grow the cells in tryptone broth (pH 7) supplemented with glucose prior to analysis by liquid chromatography. During the course of these experiments, we observed that consumption of glucose is delayed. In particular, we found that the cells reached an OD600 of  $\sim$ 1 before they started to consume the glucose. Subsequent mass spectrometry analysis demonstrated that *E. coli* consumes multiple amino acids (serine, aspartate, and threonine) before it begins to consume glucose. Similar results were also observed with lactose, arabinose, and glycerol, where again sugar consumption is delayed by amino acids. Of the carbon sources tested, only pyruvate consumption is not delayed.

The unique carbon source selection provides adequate ratios of nutrients for growth. All of the preferred amino acids (serine, aspartate, and threonine) enter metabolism through pyruvate, leading to gluconeogenic growth. This mode of growth appears to inhibit the metabolism of glucose (and other sugars) initially, and allows simultaneous access to carbon and nitrogen. In support of this mechanism, we found that pyruvate also inhibits the uptake of glucose. Curiously, adding ammonium (the preferred nitrogen source) to the growth medium does not affect the preference for amino acids.

Phosphoenolpyruvate (PEP) would be synthesized. PEP initiates the phosphotransferase system (PTS), the major pathway of sugar transport. However, PEP has been shown to inhibit glycolytic enzymes such as phosphofructokinase. Feedback to glycolysis would reduce sugar flux in this case. The PEP:pyruvate ratio indirectly controls sugar uptake through the PTS. Accumulation of pyruvate from the preferred amino acids would lead to less active PTS enzymes to transport sugars. Finally, the TCA intermediate  $\alpha$ -ketoglutarate (aKG) would accumulate. aKG has been shown to reduce sugar uptake via inhibition of the PTS and likely contributes to the delay in sugar uptake.

In conclusion, we have serendipitously identified a new fact of *E. coli* physiology that may translate to other species of bacteria. The results are significant as glucose is normally thought of as the preferred carbon source for *E. coli*. However, our results demonstrate that easily consumed amino acids can be preferred over glucose. Furthermore, they demonstrate that metabolic regulation in *E. coli* is more complex than previously thought.

### References

Doucette, C. D., Schwab, D. J., Wingreen, N. S., & Rabinowitz, J. D. (2011). α-Ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition. *Nature Chemical Biology*, 7(12), 894–901.
Hogema, B. M., Arents, J. C., Bader, R., Eijkemans, K., Yoshida, H., Aiba, H., & Postma, P. W. (1998). Inducer exclusion in Escherichia coli by non-PTS substrates : the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIA Glc, *30*, 487–498.
Ogawa, T., Mori, H., Tomita, M., & Yoshino, M. (2007). Inhibitory effect of phosphoenolpyruvate on

glycolytic enzymes in Escherichia coli, 158, 159–163.

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

### Glycolysis Balances Enzyme Efficiency and Metabolic Adaptivity

Junyoung O. Park<sup>1,2\*</sup> (jopark@princeton.edu), Monica H. Wei,<sup>3</sup> Lukas B. Tanner,<sup>1</sup> Daniel Amador-Noguez,<sup>1</sup> Sophia Hsin-Jung Li,<sup>4</sup> and **Joshua D. Rabinowitz**<sup>1,3</sup>

<sup>1</sup>Lewis-Sigler Institute for Integrative Genomics; <sup>2</sup>Department of Chemical and Biological Engineering; <sup>3</sup>Department of Chemistry; and <sup>4</sup>Department of Molecular Biology, Princeton University, Princeton, NJ

### Project Goal: To track thermodynamic changes in glycolysis using isotope tracers

Rapid glycolysis during slow growth is a desirable feature for industrial biofuel production. In practice, however, glycolysis tends to slow down together with growth. Here we set out to develop and apply, in fast and slow-growing cells, isotope tracer methods for measuring the Gibbs energy of reaction ( $\Delta$ G) of glycolytic reactions.

To this end, we selected glucose tracers harboring either <sup>2</sup>H at the fifth carbon,  $[5-{}^{2}H_{1}]$ , or <sup>13</sup>C at the first two carbons,  $[1,2-{}^{13}C_{2}]$ . These tracers generate labeling patterns across glycolytic intermediates that depend on the pathway's forward-to-backward flux ratio  $(J^{+}/J^{-})$  and therefore  $\Delta G = -RT \ln(J^{+}/J^{-})$ . Quantitative methods for integrating data from the two tracers to reveal reaction thermodynamics will be described.

Using these tracers, we show that rapid upregulation of glycolysis in *Escherichia coli* is accomplished by increasing the pathway's thermodynamic driving force. Specifically, in fastgrowing cells, we observed  $\Delta G < -2kJ/mol$  (i.e. forward flux > 2.2× backward flux) for most glycolytic reactions, reflecting efficient enzyme usage with enzymes mainly catalyzing the forward productive reaction. On the other hand, in nitrogen-limited cells with reduced glycolytic flux, in lower glycolysis, we observed  $\Delta G \sim 0$  (forward flux  $\approx$  backward flux). Such a nearequilibrium situation is energy-efficient but enzyme-inefficient. The likely evolutionary benefit of inefficient enzyme usage became manifest upon nitrogen upshift: by shifting from reversible to forward-driven thermodynamics, rapid glycolysis and growth rate were restored within minutes, without requiring increased enzyme levels or activity. Thus, nutrient-rich cells can run glycolysis at near maximal enzyme capacity, whereas nutrient-limited cells sacrifice enzyme efficiency for fast adaptation. The <sup>2</sup>H- and <sup>13</sup>C-tracer methods developed here should be broadly useful for understanding glycolytic thermodynamics and regulation across strains, species, and environmental conditions.

### References

1. Park, J. O., Rubin, S. A., Xu, Y. F., Amador-Noguez, D., Fan, J., Shlomi, T., and Rabinowitz, J. D. (2016). Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nature Chemical Biology* 12(7): 482-489.

2. Park, J. O., Wei, M. H., Tanner, L. B., Amador-Noguez, D., Li, S., and Rabinowitz, J. D. (to be submitted). Glycolysis balances enzyme efficiency and metabolic adaptivity.

Funding: DOE DE-SC0012461

### Systems-Level Analysis of Mechanisms Regulating Yeast Metabolic Flux

Sean R. Hackett,<sup>1</sup> Vito R.T. Zanotelli,<sup>5</sup> Wenxin Xu,<sup>1,2</sup> Jonathan Goya,<sup>1</sup> Junyoung O. Park,<sup>1</sup> David H. Perlman,<sup>2</sup> Patrick A. Gibney,<sup>1,3</sup> David Botstein,<sup>1,3</sup> John D. Storey,<sup>1,3,4</sup> and Joshua D. Rabinowitz (joshr@princeton.edu; presenter)<sup>1,2</sup>

<sup>1</sup>Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA. <sup>2</sup>Department of Chemistry, Princeton University, Princeton, NJ 08544, USA. <sup>3</sup>Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. <sup>4</sup>Center for Statistics and Machine Learning, Princeton University, Princeton, NJ 08544, USA. <sup>5</sup> Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland.

# Project Goals: To apply the integrated power of modern 'omic technologies to identify physiologically relevant metabolic regulation in a systematic, large-scale, and quantitative manner in Baker's yeast and (going forward) also Clostridia.

INTRODUCTION: Metabolism is among the most strongly conserved processes across all domains of life and is crucial for both bioengineering and disease research, yet we still have an unclear understanding of how metabolic fluxes are determined. Qualitatively, this deficiency involves missing knowledge of enzyme regulators. Quantitatively, it involves limited understanding of the relative contributions of enzyme and metabolite concentrations in controlling flux across physiological states. Addressing these gaps has been challenging because in vitro biochemical approaches lack the physiological context, whereas models of cellular metabolic dynamics have limited capacity for identifying or quantitating specific regulatory events because of model complexity.

RATIONALE: Flux through individual metabolic reactions is directly determined by the concentrations of enzyme, substrates, products, and any allosteric regulators, as captured quantitatively by Michaelis-Menten kinetics. Experimental variation of reaction species in vitro allows for the inference of regulators and reaction equation kinetic parameters. Analogously, physiological changes in flux entail a change in reaction species that can be used to determine reaction equations based on cellular data. This requires measurement across multiple biological conditions of flux, enzyme concentrations, and metabolite concentrations. We reasoned that chemostat cultures could be used to induce predictable and strong flux changes, with changes in enzymes and metabolites measurable by proteomics and metabolomics. By directly relating cellular flux to the chemical species that determine it, we can carry out regulatory inference at the level of single metabolic reactions using cellular data. An important benefit is that the physiological significance of any identified regulator is implicit from its role in determining cellular flux.

RESULTS: Here we describe an approach for Systematic Identification of Meaningful Metabolic Enzyme Regulation (SIMMER). We measured fluxes, and metabolite and enzyme concentrations, in each of 25 yeast chemostats. For each of 56 reactions for which the flux, enzyme, and substrates were measured, we determined whether variation in measured flux could be explained by simple Michaelis-Menten kinetics. We also evaluated alternative models of each reaction's kinetics that included a suite of allosteric regulators drawn from across all organisms. For 46 reactions, we were able to identify a useful kinetic model, with 17 reactions not including any regulation and 29 reactions being regulated by one to two allosteric effectors. Three previously unrecognized cross-pathway regulatory interactions were validated biochemically. These included inhibition of pyruvate kinase by citrate, which accumulated and thereby curtailed glycolytic outflow in nitrogen-limited yeast. For well-supported reaction forms, we were able to determine the extent to which nutrient-based changes in flux were determined by changes in the concentrations of individual reactants, products, enzymes, or allosteric effectors. Overall, substrate concentrations were the strongest driver of the net rates of cellular metabolic reactions, with metabolite concentrations collectively having more than double the physiological impact of enzymes.

CONCLUSION: By connecting changes in flux to their root cause, SIMMER parallels classic in vitro approaches, but it allows testing under physiological conditions of numerous regulators of many reactions simultaneously. Its application to yeast showed that changes in flux across nutrient conditions are predominantly due to metabolite, not enzyme, levels. Thus, yeast metabolism is substantially self-regulating.

This work was funded by the U.S. Department of Energy Award DE-SC0012461 and Office of Science Graduate Fellowship DE-AC05-06OR23100 to S.R.H. We also acknowledge support from Agilent Technologies (Thought Leader Awards to J.D.R. and D.B.).

### Systems biology towards a continuous platform for biofuels production: Engineering an environmentally-isolated *Bacillus* strain for biofuel production and recovery under supercritical CO<sub>2</sub>.

Jason T. Boock<sup>2</sup>, Adam J. E. Freedman<sup>1</sup>, Geoffrey Tompsett<sup>3</sup>, Kristala L. J. Prather<sup>2</sup>, Michael T. Timko<sup>3</sup> and Janelle R. Thompson<sup>\*1</sup> (jthompson@mit.edu)

<sup>1</sup>Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge MA; <sup>2</sup>Chemical Engineering, Massachusetts Institute of Technology, Cambridge MA; <sup>3</sup>Chemical Engineering, Worcester Polytechnic Institute, Worcester, MA

Project Goals: We are developing *Bacillus megaterium* as a host for continuous biofuel production coupled with *in situ* product extraction by supercritical  $CO_2$ (sc $CO_2$ ) stripping. We employ a cross-disciplinary collaborative approach to achieve the following (1) Develop a supercritical  $CO_2$  tolerant strain of *B. megaterium* into a bioproduction host for biofuels (2) Engineer de novo pathways for biosynthesis of longer chain fuels in *B. megaterium* and (3) Develop and model a two-phase stripping chemostat for continuous biosynthesis and in situ extraction of biofuels using sc $CO_2$  as a sustainable extractive solvent. The abstract below reports on progress towards goal (2).

Medium-chain, branched alcohols have advantages as fuels over molecules such as ethanol due to their higher energy density and lower research octane number, enabling their use as drop-in fuels with properties similar to conventional gasoline. End-product toxicity is one limitation to their production biologically, resulting in low titers that require energy intensive separations. In situ product extraction offers a solution to product toxicity by removing compounds from the culture medium as they are generated. We intend to utilize the unique thermodynamic and transport properties of supercritical  $CO_2$  (scCO<sub>2</sub>) in a bi-phasic bioreactor to strip off the alcohol products made by an scCO<sub>2</sub>tolerant strain of Bacillus megaterium (SR7) that has been engineered to produce biofuels. Additionally,  $scCO_2$  is a sustainable solvent, which when depressurized leaves the alcohol product at high concentrations, eliminating costly downstream distillation. Furthermore, scCO<sub>2</sub> has broad microbial lethality, providing a contamination-free reactor for the SR7 to grow in. In this work we have begun developing the genetic tools to engineer SR7 for isobutanol production. Additionally, we are modifying a biofuel pathway to increase the bioproduction 4-methyl-pentanol, which will be later implemented in the scCO<sub>2</sub>-tolerant *Bacillus*.

Developing tools to modify SR7 began through characterizing the physiology of the organism. Through a screen of 71 carbon sources, we identified ones that enabled high growth. One of these compounds, L-malic acid, when used in combination with a previously optimized protoplast transformation method, resulted in at least a 10-fold increase in transformation efficiency. After successful plasmid transformation, we have developed two inducible and one constitutive promoter for heterologous protein expression in SR7. Two of these promoters have not previously been demonstrated in *Bacillus megaterium*. Further, we have found xylose-inducible expression of protein

under  $scCO_2$ , providing evidence that the promoter is functional under these harsh growth conditions and suggesting heterologous production of biofuels under  $scCO_2$  is possible. We are currently looking to expand the genetic toolbox for SR7 by genomic incorporation of heterologous genes and developing strategies to knockout or knockdown endogenous enzymes.

We began to engineer SR7 to produce isobutanol by introducing a two-enzyme (2ketoisovalerate decarboxylase (KivD) and alcohol dehydrogenase (Adh)) pathway. The two-step conversion occurs at approximately 70-80% from 2-ketoisovalerate when grown aerobically; however, the intermediate aldehyde was found to accumulate at short culture times. Due to the high partition coefficient for the aldehyde to the  $scCO_2$  phase, five alternative homologous alcohol dehydrogenases were tested to eliminate its buildup. A variant was identified that lowered the build-up of isobutyraldehyde and resulted in conversion of 2-ketoisovalerate to isobutanol above 85% in aerobic cultures. The optimized, two-step conversion of 2-ketoisovalerate to isobutanol was tested in SR7 grown under scCO<sub>2</sub> with and without xylose induction. For two out of five of the induced cultures, over 20% of the 2-ketoisovalerate was converted to isobutanol, while no isobutanol was observed in the non-induced cultures. Furthermore, we have developed a five-enzyme pathway that converts pyruvate to isobutanol, and have found production of 400 mg/L isobutanol from glucose in SR7 when grown aerobically. However, this pathway has shown genetic instability and lower titers than expected, which we are modifying by genomic incorporation and using the previously developed promoters.

In addition to production of isobutanol in SR7, we have engineered a 4-methyl pentanol (4MP) production pathway in *E. coli*. Due to the limited genetic tools in SR7, its low transformation efficiency, and complexity of the 4MP pathway, we decided to engineer it for increased production in *E. coli* first. We identified the build up of isobutyrate as a potential bottleneck in the pathway. We tested three homologous propionate-CoA transferases and identified one that showed at least a 1.5-fold increase in 4MP titers. We have begun to create a more redox-balanced pathway by identifying homologous enzymes that have preference for NADH over NADPH. Lastly, we are analyzing a set of 40 keto-aryl-CoA reductases for their specificity for branched chain substrates as well as NADH cofactor preference. We intend to use the findings from engineering the 4MP pathway in *E. coli* to inform the incorporation of the pathway in SR7 to increase its likelihood of success.

### *This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0012555).*

## Systems biology towards a continuous platform for biofuels production: Heterologous gene expression and isobutanol synthesis in *B. megaterium* SR7 and biofuel extraction under supercritical CO<sub>2</sub>.

Adam J. E. Freedman<sup>1</sup>, Jason T. Boock<sup>2</sup>, Geoffrey Tompsett<sup>3</sup>, Kristala L. J. Prather<sup>2</sup>, Michael T. Timko<sup>3</sup> and Janelle R. Thompson<sup>\*1</sup> (jthompson@mit.edu)

<sup>1</sup>Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge MA; <sup>2</sup>Chemical Engineering, Massachusetts Institute of Technology, Cambridge MA; <sup>3</sup>Chemical Engineering, Worcester Polytechnic Institute, Worcester, MA

Project Goals: We are developing *Bacillus megaterium* as a host for continuous biofuel production coupled with *in situ* product extraction by supercritical  $CO_2$  (sc $CO_2$ ) stripping. We employ a cross-disciplinary collaborative approach to achieve the following: (1) Develop a supercritical  $CO_2$  tolerant strain of *B. megaterium* into a bioproduction host for biofuels (2) Engineer de novo pathways for biosynthesis of longer chain fuels in *B. megaterium* and (3) Develop and model a two-phase stripping chemostat for continuous biosynthesis and in situ extraction of biofuels using sc $CO_2$  as a sustainable extractive solvent. The abstract below reports on progress towards goals (1) and (3).

Conventional microbial biofuel production systems are subject to several long-standing challenges associated with bioreactor culturing and compound harvesting, including contamination, end product toxicity, and energy-intensive downstream bioproduct purification. However, utilizing the sustainable solvent supercritical carbon dioxide (scCO<sub>2</sub>) for in situ extraction of hydrophobic bioproducts would uniquely enable relief of end-product toxicity while creating a lethal environment for non-target organisms. In addition, depressurization of scCO<sub>2</sub> enables collection of purified, dehydrated biofuel products, eliminating the need for further downstream distillation. While previous studies have successfully demonstrated a broad diversity of *in vitro* biocatalytic reactions utilizing scCO<sub>2</sub> as a solvent and/or substrate, scCO<sub>2</sub> has previously been considered inaccessible to in vivo product biosynthesis due to its sterilizing effects on most microbes. Therefore, a bioprospecting approach was utilized in an attempt to isolate scCO<sub>2</sub>-resistant microbial strains through enrichment culture and serial passaging of deep subsurface fluids from the natural McElmo Dome scCO<sub>2</sub> reservoir. This approach enabled isolation of multiple strains, of which strain Bacillus megaterium SR7 demonstrated superior growth upon inoculation as endospores under scCO<sub>2</sub> conditions. During the first project year we showed that SR7 growth under scCO<sub>2</sub> was significantly improved by chemical induction of spore germination by media amendment with amino acid L-alanine, developed a genetic system for heterologous gene expression in SR7 during growth under scCO<sub>2</sub>, and engineered strain SR7xKY to produce isobutanol by introducing a two-enzyme (2-ketoisovalerate decarboxylase (KivD) and alcohol dehydrogenase (Adh)) pathway and feeding substrate 2-ketoisovalerate ( $\alpha$ -KIV).

Over the previous project year (Year 2) we have validated heterologous gene expression in strain SR7 under  $scCO_2$  using the LacZ reporter system, demonstrated isobutanol expression during growth under  $scCO_2$  in 10 mL batch bioreactors, scaled up SR7 growth to larger reactor volumes with improved design for real-time monitoring or subsampling, and characterized butanol extraction efficiency over a range of reactor operating conditions. Average LacZ specific activity (µmol/min\*mg) values in transformed strain SR7xL cultures grown under  $scCO_2$  demonstrate

that heterologous enzyme production increased 13.1-fold by induction with xylose (p = 0.026), indicating potential for multi-gene heterologous pathway expression under scCO<sub>2</sub>. This hypothesis was verified upon incubating transformed biofuel strain SR7xKY under scCO<sub>2</sub>, which resulted in generation of up to 93.5 mg/L isobutanol and 29.7 mg/L isopentanol from 580.6 mg/L (5 mM) of substrate  $\alpha$ -KIV, indicating a yield of 21.2%. Direct extraction of 5.2% of the total isobutanol product by the scCO<sub>2</sub> represents the first demonstration of heterologous multi-enzyme expression, biofuel generation, and product harvesting in a single scCO<sub>2</sub>-exposed bioreactor.

After demonstrating biofuel production and extraction in bench scale (10 mL) biphasic batch reactors, efforts were initiated to achieve consistent growth for robust biofuel production at increased scales (25 mL and 0.3 L) in biphasic reactors with improved design to allow for continuous visual assessment of culture turbidity (25 mL reactor with sapphire view window) or subsampling (custom-built 300 mL capacity bioreactor with a Rushton impeller stirring at 300 rpm). At the 0.3 L scale, cultures of SR7 in optimized semi-defined minimal media under scCO<sub>2</sub> headspace demonstrated cell growth after 4 days, reaching concentrations above 10<sup>8</sup> cells/mL. However, profiles of cell density, metabolite production, and glucose consumption in fluids removed from the reactor indicated that growth was localized to a recessed, unmixed region at the base of the reactor that remained subject to full pressurization. Subsequent investigation of unmixed SR7 cultures under scCO<sub>2</sub> in the 25 mL capacity batch bioreactor also revealed turbidity after 4 days. Observations using epifluorescence microscopy revealed vegetative cells dispersed in the bulk media as well as cell aggregates encased in suspected biofilm extrapolymeric substances (EPS). We are currently examining the role of increased surface area to volume ratios, oxygen amendments, and modified stirring regimes on spore germination and outgrowth in the bulk media at the 25 mL and 0.3 L scales.

Upon the establishment of SR7 growth at the 0.3 L scale, we will introduce SR7 transformed with a biofuel pathway into the 0.3 L capacity reactor for compound generation and extraction by semi-continuous scCO<sub>2</sub> flow-through. In advance of anticipated short-to-medium chain alcohol bioproduction, we investigated the extraction of butanol, pentanol, and hexanol within the 0.3 L reactor under abiotic conditions. The objective was to examine the extraction efficiency in the 0.3 L reactor, which has not previously been established. We found that extraction rates and efficiency improved with alcohol chain length (i.e., hexanol > pentanol > butanol), broadly consistent with their equilibrium behavior. We modeled the extraction using the 2-film approximation to estimate mass transport coefficients; these were in reasonable agreement with literature values obtained for ethanol extraction. Correlations developed for liquid-liquid extraction agreed within approximately 25% with the estimated values of mass transfer coefficient. In addition, we modeled the energy efficiency of extraction for different CO<sub>2</sub>/alcohol ratios. We found that extraction rates in the fermenter would be sufficient for industrial purposes; however, extraction efficiency (which was on the order of 100 moles CO<sub>2</sub> required per mole of alcohol extracted) were not industrially practical. Instead, a multi-stage extractor design will be needed to achieve industrially relevant extraction efficiencies. Overall, progress associated with biotic culturing and abiotic biofuel extraction provides a clear basis for continuing development of this sustainable, integrated biofuel production and extraction technology.

*This work was supported by the Office of Biological and Environmental Research in the DOE Office of Science (DE-SC0012555).* 

### A Systems Biology and Pond Culture-Based Understanding and Improvement of Metabolic Processes Related to Productivity in Diverse Microalgal Classes for Viable Biofuel Production.

Jürgen Polle<sup>1</sup>\* (jpolle@brooklyn.cuny.edu), Michael Huesemann,<sup>2</sup> and Mark Hildebrand<sup>3</sup>

<sup>1</sup>Brooklyn College of the City University of New York, Brooklyn, NY; <sup>2</sup>Marine Sciences Laboratory, Pacific Northwest National Laboratory, Sequim, WA; and <sup>3</sup>Scripps Institution of Oceanography, UCSD, San Diego, CA.

# Project Goals: Understand the fundamental molecular mechanisms involved in maximizing growth rate and lipid accumulation in diverse classes of microalgae under authentic diurnal conditions to enhance production capabilities for biofuels.

Microalgal mass culture in open ponds can provide feedstock for biofuels production. However, major gaps exist in the understanding of the effect of metabolic topology on cellular carbon partitioning and its regulation with regards to productivity. Our research develops two emerging platform green algal species for biofuels production as model systems: *Acutodesmus (Scenedesmus) obliquus* and *Coelastrella* sp. Our Systems Biology approach includes genomics-based investigations with climate-simulated pond culturing to identify species specific versus general green algal mechanisms underlying performance in realistic biofuels production scenarios.

For A. obliguus, a draft genome was obtained in collaboration with Dr. Starkenburg (Los Alamos National Laboratory) and Dr. Magnuson (Pacific Northwest National Laboratory). The genome is publicly available at https://greenhouse.lanl.gov/greenhouse/organisms/ (the Greenhouse page of LANL). We annotated the genome and created a metabolic map for the carbon core metabolism of A. obliquus, which was used as a foundation for identification of potential bottlenecks and regulatory mechanisms regarding carbon partitioning in this green alga. Based on comparative genomics studies including information from higher plants, we identified several potential bottlenecks and regulatory steps within the carbon core metabolism that could be relevant for biomass productivity and carbon partitioning. One example is the chloroplast localized Triose Phosphase Isomerase (TPI) for which only one gene exists in many green algae, but due to a whole genome duplication (WGD) two gene copies were found in A. obliquus. This is the first report for a WGD in a green alga and possibly valuable regarding green algal productivity, because ancient WGDs in higher plants resulted in gene duplications that underlie important productivity traits in crop species. Our broader comparative genomics analysis regarding the TPI of species from different clades within in green algae strongly indicated major differences in carbon flux topology within the cells, which can be linked to the number of TPI isoenzymes and their cellular localization. In addition, pathways were identified for biosynthesis of Trehalose and Fructose-2,6BP, both essential regulatory metabolites in carbohydrate metabolism. Transcriptional studies are ongoing to determine links between specific gene

expression with daily productivity changes in pond cultures. *Omics* data aquired for *A. obliquus* are now being compared to *Coelastrella* sp., for which a draft genome was obtained and annotated.

Regarding pond mass cultures, an important result is that during the linear growth phase already in the late afternoon biomass productivity in ponds was strongly reduced in both green algae *A. obliquus* and *Coelastrella* sp. Analysis of transcriptomic data (dawn versus dusk) in both algae corroborated up-regulation of genes coding for enzymes involved in glycolysis and the Krebs cycle versus down-regulation of genes coding for the Calvin-Benson-Bassham cycle before onset of the dark period. Our results confirm at the molecular level that light-limitation resulting in respiratory losses of biomass already occurring during the light period is a major factor that has to be addressed for improvement of metabolic processes if increases in productivity of green algae are to be achieve through metabolic engineering. Concurrent creation (UV-mutagenesis) and testing of mutants of both algae for biomass productivities resulted in novel strains that demonstrated higher productivities in laboratory cultures.

In summary, our comparative systems approach regarding carbon core topology allowed us to identify the TPI as a general molecular factor in determining cellular carbon flux among different classes of green algae. In green algae, the TPI is an example on class-wide design differences regarding processing of carbon. We also determined that in pond cultures *A. obliquus* performed significantly better than *Coelastrella* sp., but concomitantly in both species early onset of respiratory activity in the late afternoon is preventing higher daily biomass productivities for both species. Our carbon core metabolic maps will enable us to perform a comparative investigation into regulation of gene expression of carbon core enzymes in *A. obliquus* and *Coelastrella* sp. in pond cultures. Such systems analysis should allow linking of phenotypic biomass productivities of both algae to similarities and differences in their carbon flux topology.

*Our project is supported by the Office of Biological and Environmental Research in the DOE Office of Science through Award #DE-SC0012556.* 

### The Effect of Carbon Flux Topology and Synchronized Culture Growth on Microalgal Productivity

**Mark Hildebrand<sup>1</sup>** \* (mhildebrand@ucsd.edu), Juergen Polle<sup>2</sup>, Michael Huesemann<sup>3</sup> <sup>1</sup> Scripps Institution of Oceanography, University of California, San Diego, San Diego, CA <sup>2</sup> CUNY, Brooklyn College, Brooklyn, NY

<sup>3</sup> Marine Sciences Laboratory, Pacific Northwest National Laboratory, Sequim, WA

### **Project Goals**

The overarching goal of this project is to understand and manipulate fundamental molecular mechanisms involved in maximizing growth rate and lipid accumulation in diverse classes of microalgae under authentic diurnal conditions to enhance production capabilities for biofuels. There are three primary targets to achieve this goal. The first is to advance the development of promising new model organisms relevant to biofuels production, which we are doing by focusing on highly productive strains of *Acutodesmus obliquus* and the diatom *Cyclotella cryptica*. The second is to clarify the features and organization of metabolic pathways in the study organisms, particularly with regard to carbon flux. The third is to evaluate growth and productivity under simulated outdoor cultivation conditions, to understand how cellular processes, particularly during synchronized cell division, affect productivity. Microalgae experience sinusoidal variation in light and temperature when grown outdoors, which is not typically considered in a lab setting, but which has a substantial influence on cellular processes.

### Abstract

I. Elucidation and clarification of metabolic pathways in green algae and diatoms.

Using bioinformatic analysis and experimental approaches, we have gained substantial knowledge about carbon flux metabolism in our study organisms, which relates on a larger scale to two major classes of microalgae. We have generated the first metabolic map of core carbon metabolism in *A. obliquus*, which highlights the major routes of carbon processing. We have completed determination of the genome sequence of *C. cryptica*, and identified differences in carbon flux topology, particularly in processing of pyruvate, compared with the model species *Thalassiosira pseudonana*, that could explain the superior ability of *C. cryptica* to accumulate lipid. We have used a combined bioinformatic and experimental approach to clarify the process of photorespiration in *T. pseudonana*. Photorespiration is the second most important carbon flux pathway in algae. The data indicate that diatoms do not follow the pathway as elucidated in higher plants, and rather than recycle photorespiratory products back into the chloroplast to be re-fixed by RubisCO, they convert the products mostly into amino acids.

These analyses enabled us to perform comparative analyses of carbon flux pathways in different green algal species and different species of diatoms. Unexpectedly, predicted intracellular targeting data indicate substantial differences in carbon flux comparing six different green algal species that have sequenced genomes. One fundamental concept is that even though starch is stored in the chloroplast of most species, in some cases the pathways of synthesis and/or breakdown of starch have been relocated to the cytoplasm. This may affect the efficiency of carbon flux, and temporal preference to undergo cellular division. A comparative analysis of three diatom species also indicates substantial differences in carbon flux topology, which suggests active evolution to optimize arrangements by moving portions of pathways between different cellular compartments.

An overarching conclusion from these analyses relates to the preferred time of cell division of green algae and diatoms during the day; the former tend to divide at night because sufficient carbohydrate stores must be generated during the day, and the latter tend to divide during the day, because they rely more on direct photosynthetically fixed carbon for division than stored carbohydrate. In addition to the temporal effect on division, there are implications related to the extent of nighttime respiration, which also contributes to productivity.

II. Characterization of performance and productivity parameters under simulated outdoor conditions.

We have taken two approaches in this regard, performing 1) small scale experiments using Phenometrics photobioreactors (PBRs), and 2) large scale experiments using the PNNL simulated outdoor raceway ponds.

PBR experiments were done on *C. cryptica* using a 12:12 l:d cycle, 2000 uE maximum light intensity, with a temperature variation between 15-26°C. The data indicate that cell cycle synchronization occurred, with the majority of the population dividing once per day beginning in the early afternoon. Changes in the optical density at 750 nm (OD750) preceded changes in culture density, indicating that OD750 is not a completely accurate proxy for culture density. It was an excellent proxy for ash free dry weight (AFDW). Lipid content, as triacylglycerol (TAG) increased until prior to mitosis, decreased as cell membranes were synthesized, then increased subsequently. The data indicate that optimal times of the day to harvest cultures for biomass or for lipid differed substantially.

We have completed two sets of runs in duplicate each of *C. cryptica* in the PNNL simulated outdoor raceway systems, with a focus on extensive sampling at two different times, 1) on a day when the cultures became limited for silicon, and 2) on a day when no limitation occurred. Conditions were 12:12 l:d cycle, 15-26°C and maximum light intensity of 1200-1400 umol•m<sup>-2</sup>•s<sup>-1</sup>. The cultures synchronized, with division occurring during the day. Productivity during both runs was impressive, the average productivity from lag into stationary phase for run #1 was 18g AFDW/m<sup>2</sup>/d, and for run#2 was 20 g AFDW/m<sup>2</sup>/d. Productivity during the two peak days for each was 28-30 g AFDW/m<sup>2</sup>/d for run #1, and 39-42 g AFDW/m<sup>2</sup>/d for run #2. The sensitivity of productivity to mean light intensity was monitored, showing a relatively linear response. The mean light intensity and mean dissolved oxygen (DO) levels were highly correlated. DO measurements indicated that nighttime respiration was minimal, which could relate to the preferred division time during the day. Variation in cellular TAG content was monitored throughout the runs. Samples for RNA extraction were taken, and will be processed for transcriptomic analyses to investigate underlying cellular responses to progression through the diurnal cycle.

In comparison to *C. cryptica*, under identical pond cultivation conditions *A. obliquus* had a much lower average biomass productivity of about 12g AFDW/m<sup>-2</sup>/d. This drastically lower productivity of the green alga *A. obliquus* may be attributable to the above mentioned species differences, separation of carbohydrate accumulation during the day from cell division at night, thus forcing the green alga to loose biomass due to respiratory activity for energy provision for the energetically costly cell division process during the night. Different temperature optimum for cultivation could also contribute to lower productivity for *A. obliquus*.

*Our project is supported by the Office of Biological and Environmental Research in the DOE Office of Science through Award #DE-SC0012556.* 

### Spatiotemporal Transcriptomics of *Populus* Growth in Response to Daylength and Nutrient Availability

Hua Bai<sup>1,\*</sup> (hbai@vt.edu), Rita Teixeira<sup>1</sup>, Jason A Holliday,<sup>1</sup> and Amy M Brunner<sup>1</sup>

<sup>1</sup>Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg VA 24061

Project Goals: The overall goal of this project is to elucidate the regulatory networks that govern transitions between growth and quiescence in response to two abiotic conditions – daylength and nutrient availability. Studying these two abiotic signals will enable identification of commonalities and differences among different environmental signaling networks that converge on the *Populus* central shoot growth regulator FT2.

In woody plants, a variety of unfavorable abiotic conditions ultimately induce shoot apical and cambial growth cessation and terminal budset, a quiescent state that has classically been defined as ecodormancy. If the limiting factor is supplied, trees will rapidly resume shoot growth. We conducted experimental treatments in which daylength and nutrient availability were reduced to induce quiescence and then increased to resume growth. These experiments lasted several weeks with sampling of relevant tissues (shoot apex/bud, leaf, root, and vascular cambium) at 11 time points. While bud set and dormancy induction in response to short days is well-studied, and proceeded as expected, less is known about the dynamics of nutrient-mediated growth cessation. We found that under nutrient deficiency, the active shoot apex transitioned more rapidly to the final budset stage, but quiescence of the cambial meristem was delayed. Moreover, these plants quickly resumed growth with subsequent fertilization. mRNA libraries were prepared from each of three biological replicates of the shoot apex, leaf and root samples/time points, resulting in 234 libraries sequenced on an Illumina HiSeq 2500 instrument, and we are currently in the process of mapping and calculating transcript abundance metrics for these data. Laser capture microdissection is being used to isolate the cambial zone at phenotypically distinct time points for mRNA libraries. In parallel, we are preparing small RNA and degradome libraries for a subset of the samples, which will provide an additional layer of transcriptional regulatory information. These data will collectively be used to develop transcriptional networks that describe spatially and temporally explicit regulatory relationships for both daylength and nutrient-mediated quiescence. Functional relationships among key network regulators will be further characterized through transgenic manipulation.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012574

### Synthetic microbial ecology for biofuel production from lignocellulose

Meghan Thommes<sup>1</sup>\* (mthommes@bu.edu), Andrea Lubbe<sup>2</sup>\* (alubbe@lbl.gov), Jessica Lee<sup>3,4,5</sup>\* (jessicalee@uidaho.edu), Melisa Osborne<sup>6</sup>, Arion Stettner<sup>1</sup>, Ilija Dukovski<sup>6</sup>, Alyssa Baugh<sup>3</sup>, Nicholas Shevalier<sup>3</sup>, Sergey Stolyar<sup>3,4,5</sup>, Christopher Marx<sup>3,4,5</sup>, Trent Northen<sup>2</sup> and **Daniel Segrè**<sup>1,6,7</sup>

<sup>1</sup>Dept. of Biomedical Engineering, Boston University, Boston, Massachusetts; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, California; <sup>3</sup>Dept. of Biological Sciences, <sup>4</sup>Institute for Bioinformatics and Evolutionary Studies, and <sup>5</sup>Center for Modeling Complex Interactions, University of Idaho, Moscow, Idaho; <sup>6</sup>Bioinformatics Program, Boston University, Boston, Massachusetts; <sup>7</sup>Dept. of Biology, Boston University, Boston, Massachusetts

Project Goals: The goal of our research is to implement a high-throughput pipeline for the systematic, computationally-driven study and optimization of microbial interactions and their effect on lignocellulose degradation and biofuel production. We combine multiple approaches, including computer modeling of ecosystem-level microbial metabolism, mass spectrometry of metabolites, genetic engineering, and experimental evolution.

The conversion of lignocellulose to biofuels is a complex process composed of many steps. Instead of engineering one microorganism to perform every function, our project assigns microbes to specific functions: (hemi)cellulose degradation, lignin degradation, removal and detoxification of methoxy groups on lignin, and the production of biofuel precursors (Figure 1). In order to test and predict the capacity of organisms to grow together and perform their specific tasks, we are using an array of tools: **growth experiments** to determine the best medium components, **metabolomics** to measure metabolite production and consumption, **flux balance analysis** (FBA) to model metabolite exchange between community members, and **genetic engineering and experimental evolution** to develop strains with desired traits.



Figure 1. Synthetic microbial community design.
We chose each organism based on its metabolic capacity, oxygen requirement, pathogenicity, and genome availability. All community members grow aerobically, are not pathogenic, and have had their genome sequenced. *Cellulomonas fimi* is a soil-dwelling bacterium that degrades (hemi)cellulose, *Pseudomonas putida* is a soil-dwelling bacterium that degrades lignin, *Methylobacterium extorquens* is a soil-dwelling bacterium that can grow on C1 compounds such as formaldehyde, and *Yarrowia lipolytica* is an oleaginous yeast that can produce high yields of lipids. *C. fimi* is the only microbe that lacks a published metabolic model<sup>1–3</sup>; however, we built and gap-filled a draft model using KBase and have manually added reactions based on our experimental observations. Since lignocellulose is very complex and the models lack many of the carbon sources found in lignocellulose, we are using Biolog plates with our own protocol to scan for growth and utilization of additional carbon sources to be added to the models.

Free formaldehyde produced by the demethoxylation of lignin monomers can inhibit microbial growth due to its toxicity. Although *M. extorquens* can utilize formaldehyde as a central metabolic intermediate, it cannot naturally remove and detoxify the methoxy groups on lignin. In order to demethoyxlate and cleave complex aromatics, we have genetically engineered M. extorquens to express the vanABK operon from Methylobacterium nodulans. We have tested he capacity of several species to grow together, and obtained preliminary data on potential interactions between community members as a function of environmental conditions. We observed the growth of C. fimi, P. putida, and M. extorquens on vanillic acid and cellobiose, which serve as model compounds for lignin and cellulose, respectively. C. fimi grows very poorly in the presence of P. putida, but grows slightly better and consumes cellobiose faster in the presence of *M. extorquens* (irrespective of whether *P. putida* is present). *M. extorquens* grows better in the presence of C. fimi. This is probably because C. fimi provides a growth substrate for *M. extorquens* (since *M. extorquens* cannot use vanillic acid or cellobiose as a carbon source). The growth of P. putida was not affected by the presence of C. fimi or M. extorquens. However, the presence of M. extorquens decreases the amount of formaldehyde produced by P. putida growing on vanillic acid. We also observed the growth of C. fimi, P. putida, and Y. lipolytica on ionic liquid treated switchgrass, with slightly different patterns of interactions between specific organism pairs. Ongoing work is aimed at processing growth data, metabolic measurements and model predictions to converge towards a functioning prototype of the complete consortium.

#### References

- 1. Nogales, J., Palsson, B. Ø. & Thiele, I. BMC Syst. Biol. 2, 79 (2008).
- 2. Peyraud, R. et al. BMC Syst. Biol. 5, 189 (2011).
- 3. Kavšček, M., Bhutada, G., Madl, T. & Natter, K. BMC Syst. Biol. 9, 72 (2015).

This work was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy, Award No. DE-SC0012627, and the Office of Science Early Career Research Program, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.

#### Genomics of Energy Sorghum's Water Use Efficiency/Drought Resilience

John Mullet,<sup>1</sup>\*, Presenter, Principal Investigator (<u>jmullet@tamu.edu</u>), Brock Weers,<sup>1</sup>, William Rooney,<sup>1</sup> and Marc Libault<sup>2</sup>

<sup>1</sup>Texas A&M University, College Station, Texas; <sup>2</sup>University of Oklahoma, Norman, Oklahoma

**Project Goals:** The overall goal of this research is to increase the biomass yield and resilience of C4 bioenergy grasses such as energy sorghum by improving water use efficiency, soil water extraction, and drought resilience.

Field data on the growth of energy sorghum leaves, stems and root systems was collected during a ~150-200 day growing season under well watered and water limiting conditions. Energy sorghum root systems grow throughout the long duration vegetative phase accumulating ~10+ Mg/ha of dry root biomass in addition to ~40 Mg of shoot biomass under good growing conditions. The grain sorghum APSIM crop model was modified and used to predict energy sorghum biomass accumulation and soil water extraction dynamics, and to investigate traits that could improve productivity in water limited environments. VPD-limited transpiration, increased leaf angle, and deep rooting systems were found to improve water capture, water use efficiency and resilience. The genetic basis of variation in root system size and morphology in the field was investigated by excavating, washing, and imaging root systems using WINRHIZO. QTL for root system size, root angle, nodal root number, and branching were identified through analysis of RIL populations. Genetic and environmental factors modulating soil water extraction, water use efficiency and water deficit resilience were examined with greater precision using rhizotronlysimeters filled with field soil in controlled environments. Diverse sorghum genotypes and RIL populations were screened using the rhizotron-lysimeters revealing significant variation in total root system size, root angle, root morphometrics, water use efficiency and root growth responses to water deficit. This system was used to screen RIL populations and to identify QTL that modify a wide range of root traits. A sorghum transcriptome atlas was constructed in collaboration with JGI. Root systems differentially expressed a large number aquaporins among other genes that aid root function. Variation in sorghum root hair biology was analyzed using an ultrasound aeroponic system developed by Marc Libault. Lateral root number increased in response to water deficit in this system, as observed in the field, and in rhizotron-lysimeters. To enhance the microscopic observation and measurement of root hair cells, sorghum root systems were stained with 4% trypan blue solution to maximize contrast. Root hair cell density and morphology was recorded from 5 different locations on the root system using Image J software. Analysis of sorghum RIL populations revealed significant differences in root hair length and growth in response to water deficit. A root hair transcriptome and epigenome resource is being developed in collaboration with JGI.

*Funding statement.* This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012629

#### Systems Biology of Autotrophic-Heterotrophic Symbionts for Bioenergy: Constraint-Based Community Modeling Reveals Condition-Dependent Interactions

Cristal Zuniga<sup>1\*</sup> (crzuniga@eng.ucsd.edu), Geng Yu<sup>3</sup>, Chien-Ting Li<sup>3</sup>, Michael Guarnieri<sup>2</sup>, George A. Oyler<sup>3</sup>, **Michael Betenbaugh<sup>3</sup>**, and **Karsten Zengler<sup>1</sup>** 

<sup>1</sup>University of California, San Diego; <sup>2</sup>National Renewable Energy Laboratory, Golden; <sup>3</sup>Johns Hopkins University, Baltimore

The goal of this proposal is to combine autotrophs and heterotrophs as a novel sustainable symbiotic platform for the production of biofuel and its precursors. Photosynthetic microorganisms are providing substrates and oxygen to the heterotrophs, which in exchange will produce  $CO_2$  for carbon fixation. Synthetic communities of cyanobacterium-bacterium, cyanobacterium-fungus, and fungus-algae pairs were studied through genome-scale metabolic modeling. As part of this project, we also reconstructed the model for the algae *Chlorella vulgaris* UTEX 395 (*i*CZ843), which was experimentally validated.

Six co-culture metabolic models for all proposed co-cultures were constructed and used to predict growth phenotypes. The models enabled the systematic characterization of co-cultures and provided insight into the interaction of the autotrophic and heterotrophic symbiotic communities. Defined growth conditions changes or genetic manipulations vielded improved metabolic phenotypes for these co-cultures. Furthermore, the metabolic models allow elucidation of the nature of specific interactions between individual members of the co-culture (e.g. commensalism, mutualism, competition, predator-prey, etc). In order to validate the predicted results, studies were designed to vigorously examine the predictions. Based on experimental observations, we corroborated our predictions about culture medium and genetic modifications for the co-culture pair of Chlorella vulgaris UTEX 395 and Saccharomyces cerevisiae S288c. The co-culture model was reconstructed using individual genome-scale metabolic models integrated into a combined model and by including a shared metabolite pool (SMP). The community model *i*CZ-Cv-Sc(1748) contains 14 unique compartments (cytoplasm, mitochondria, peroxisome, nucleus, endoplasmic reticulum, vacuole, Golgi apparatus, glyoxysome, chloroplast, thylakoid and extracellular space). Metabolites that can potentially be shared by C. vulgaris and S. cerevisiae were defined using BIOLOG data. The Constraint-Based Reconstruction and Analysis of Communities (COBRAcom) toolbox was developed in the framework of this proposal. COBRAcom allows obtaining single model's statistics, and provide tools for the reconstruction of community models. Additionally, COBRAcom contains test functions for several co-culture characterization applications, such as a) prediction and fitting of growth rates and population proportions (constraints-based choice to achieve experimental growth rates); b) determination of possible interactions (theoretical interchange of metabolites, SMP analysis); c) co-culture medium optimization (robustness analysis); d) syntrophic pathway inclusion (metadata contextualization); e) gene essentiality (knock-out analysis for population); and f) gene interactions within the population members.

The model successfully predicts growth rates observed by experiments. While *S. cerevisiae* uptakes  $O_2$  and provides  $CO_2$  when nitrate is the nitrogen source in medium, both members split the available glucose

and grow as mutualists. When  $NH_4$  was added to the culture medium *S. cerevisiae* dominated the coculture and outcompeted *C. vulgaris*, changing the type of interaction.

The SMP analysis showed that under mutualistic co-culture conditions, the growth of *S. cerevisiae* and *C. vulgaris* is mediated by exchange of built-in metabolites. Validation of these metabolites exchange by targeted metabolomics is under way. We also evaluated the effect of 1,748 single gene deletions for monocultures and the co-culture. The experimental growth rates match with predictions, e.g. yeast's mutant alters both Chlorella and yeast growth rates. When another gene is deleted, differential effects on yeast and Chlorella growth are observed. In some cases, deleting a gene can offer a growth advantage in co-culture, including an improved growth phenotype for the participating species.

These examples demonstrate how co-culture metabolic models can accurately predict the behavior of heterogeneous co-culture pairs thus improving production phenotype for bioproduction.

Supported by DE-SC0012658

### Aromatic tolerance and utilization in adapted *Rhodococcus opacus* strains for lignin bioconversion

William R. Henson<sup>1</sup>, Tayte P. Campbell<sup>2,3</sup>\* (tayte.campbell@wustl.edu), Aki Yoneda<sup>3,4</sup>, Yu Gao<sup>1</sup>, Drew DeLorenzo<sup>1</sup>, Bertram Berla<sup>3</sup>, Soo Ji Kim<sup>1</sup>, **Marcus Foston<sup>1</sup>**, **Tae Seok Moon<sup>1</sup>**, and **Gautam Dantas**<sup>3,4,5,6</sup>

<sup>1</sup>Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, St. Louis, MO, 63130; <sup>2</sup>Department of Plant and Microbial Biosciences, Washington University in St. Louis, St. Louis, MO, 63130; <sup>3</sup>Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>4</sup>Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>5</sup>Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, 63130; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, St. Louis, MO, 63130; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, MO, 63108; <sup>6</sup>Department Microbiology, Washington Univ

# Project Goals: We aim to characterize and improve *Rhodococcus opacus* PD630 as a chassis for conversion of lignocellulosic biomass, specifically thermochemically depolymerized lignin (i.e., aromatics), to valuable products. This approach can add value to a previously under-utilized feedstock (lignin) and improve titers and productivities in lignocellulosic biorefineries.

Lignocellulosic biomass is a renewable feedstock that can be converted to valuable products. During biomass depolymerization, growth inhibitors can be generated which limit product yields, titers, and productivities, and the non-sugar component (i.e., lignin) is not used during fermentation. To increase tolerance to growth inhibitors and incorporate utilization of lignin for production of valuable products, we have explored tolerance to and utilization of lignin-derived aromatics by *Rhodococcus opacus* PD630, a bacterial strain that can accumulate triacylglycerols (biodiesel precursor) up to ~80% of its cell dry weight in nitrogen-limited culture conditions. Specifically, we performed adaptive evolution on lignin model compounds to improve the native aromatic tolerance of *R. opacus* and characterized adapted strains by multi-omics approaches. *R.* opacus strains adapted for growth on single and multiple lignin model compounds demonstrated improved growth and lipid accumulation compared to the wild-type strain. In addition, mixtures of adapted strains demonstrated improved growth on depolymerized lignin compared to the wildtype strain. Whole genome sequencing, RNA-Sequencing, <sup>13</sup>C-fingerprinting, and metabolomic approaches have identified possible mechanisms for improved aromatic tolerance and utilization such as upregulation of degradation pathways and transporters. We will present our progress towards the development of *R. opacus* as a microbial cell factory.

#### References

1. WD Hollinshead, WR Henson, M Abernathy, TS Moon and YJ Tang. Rapid Metabolic Analysis of *Rhodococcus opacus* PD630 via parallel <sup>13</sup>C-Metabolite Fingerprinting, Biotechnol. Bioeng. 113, 91-100 (2016)

2. Yoneda A, Henson WR, Goldner NK, Park KJ, Forsberg KJ, Kim SJ, Pesesky MW, Foston M, Dantas G, Moon TS. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630, Nucleic Acids Res. 18, 2240-2254 (2016)

This work is funded by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research, Biological System Sciences Division, award #DE-SC0012705.

#### Spatial Connectomics to Identify Agents Relevant to Lignocellulose Deconstruction in Fungi

Jiwei Zhang<sup>1</sup>, Gerald N. Presley<sup>1</sup>, Kevin Silverstein<sup>1</sup>, Melania Figueroa<sup>1</sup>, Katarina Sweeney<sup>1</sup>, Kenneth E. Hammel<sup>2</sup>, Christopher J. Hunt<sup>2</sup>, Ellen A. Panisko<sup>3</sup>, and **Jonathan S. Schilling**<sup>1</sup>

<sup>1</sup>University of Minnesota, Saint Paul, <sup>2</sup>U.S. Forest Products Laboratory (FPL); <sup>3</sup>Pacific Northwest National Laboratory

URL: http://schillinglab.cfans.umn.edu

**Project Goals:** Our goal is to discover which genes are differentially up-regulated across the mycelia of brown rot wood-degrading fungi *in planta*, particularly at the leading edge of wood decomposition. These unique fungi accomplish what we have difficulty achieving – energy from plant biomass. To do this, brown rot fungi apparently couple an oxidative pretreatment step with enzymatic saccharification in discrete space, via partitioning of reactions. We have previously shown evidence that these are governed by differential expression, but with genes putative, transport out of hyphae unclear, and the secretome poorly studied, historically, a comprehensive approach is needed that can also limit ('winnow') data sets from powerful global analytical tools to focus on the genes and pathways that matter. To address this, we are comparing global expression profiles among mycelial regions and matching what we see to what they do, in terms of changing wood physiochemistry. To map a 'connectome' in a wood-fungal interaction, specifically, we must also match gene expression patterns with the extracellular secretome and with physiochemical wood modifications.

**Objective 1** – **Zone localization:** Use the wood wafer design to resolve a depolymerization zone in *P. placenta* near the mycelial leading edge, and optimize RNA extraction for thin-sectioning.

**Objective 2 - Fungal connectomics:** Co-localize gene expression with the secretome and with relevant physiochemical modifications made to the wood, e.g. hemicellulose loss, porosity changes.

**Objective 3 - Clade comparisons:** Compare key zones among brown rot clades, in context of white rot same-clade ancestors, to target universal 'brown rot' genes and candidates for bioprocessing.

#### Abstract:

Some fungi are uniquely able to deconstruct lignocellulose, and their mechanisms have potential biofuels applications. A key hindrance to harnessing these mechanisms has been their spatial complexity. Our past work has shown that differentiated networks of hyphae that penetrate wood are not metabolically uniform, with critical reactions occurring near the hyphal front. Standard analyses of these fungi from artificial media or from colonized wood ground en masse fail to distinguish expression of key gene products in localized regions along growing hyphae.

Our focus for this research is specifically on brown rot fungi, a more recently evolved decay fungal group (relative to white rot) that circumvents the lignin barrier to extract sugars from lignocellulose. The genetic basis for how this capacity evolved away from white rot multiple times remains unknown, despite the modern options to align the compare brown rot and white rot fungal genomes. Our new collaboration aims to focus omics techniques to map and integrate expression over networks of wood-degrading fungal hyphae *in planta*. A similar approach, 'connectomics,' has been used to map the human nervous system, and its application here is timely. First, wood-degrading fungal genomes are an emerging resource, particularly brown rot functional types.

Second, we recently optimized a thin-section wood set-up that can finely resolve reaction zones along an advancing mycelium. Within these zones, we can employ deep omics approaches without the typical noise of whole-sample homogenization. By co-localizing gene expression, secretions, and wood modifications, we can prioritize the genes most useful for application.

Our goal is to discover which genes are differentially up-regulated across the mycelia of wooddegrading fungi *in planta*, particularly at the leading edge of wood decomposition. To do this, we need to compare global expression profiles among mycelial regions. To map a 'connectome' in a wood-fungal interaction, specifically, we must also match gene expression patterns with the extracellular secretome and with physiochemical wood modifications. Given this potential for substrate-fungus feedback, we will cross-check genes using separate clade representatives for brown rot fungi alongside their white rot ancestors, harnessing the JGI MycoCosm portal and several key resources and expertise at the USDA Forest Products Laboratory and the Pacific Northwest National Laboratory.

#### **References:**

Presley, G.N., Zhang, J., Schilling, J.S.\* (online first) A genomics-informed study of oxalate and cellulase regulation by brown rot wood-degrading fungi. Fungal Genetics & Biology doi.org/10.1016/j.fgb.2016.08.004.

Zhang, J., Presley, G.N., Hammel, K.E., Menke, J.R., Hu, D., Ryu, J.S., Orr, G., Schilling, J.S. \*(2016) Localizing gene regulation reveals a staggered wood decay mechanism for the brown rot fungus *Postia placenta*. Proceedings of the National Academy of Sciences 113: 10968-10973.

Kaffenberger, J.T. and Schilling, J.S. (online first) Comparing lignocellulose physiochemistry after decomposition by brown rot fungi with distinct evolutionary origins. Environmental Microbiology doi:10.1111/1462-2920.12615.

Schilling, J.S., Kaffenberger, J.T., Liew, F.J., Song, Z. (2015) Signature wood modifications that reveal decomposer community history. PLoS ONE 10(3): e0120679. doi:10.1371/journal.pone.0120679.

Schilling, J.S., Duncan, S.M., Presley, G.N., Filley, T.R., Jurgens, J.A, and Blanchette, R.A. (2013) Colocalizing incipient reactions in wood degraded by the brown rot fungus *Postia placenta*. International Biodeterioration and Biodegradation 83: 56-62.

Schilling, J.S., Ai, J., Blanchette, R.A., Duncan, S.M., Filley, T.R., and Tschirner, U.W. (2012) Lignocellulose modifications by brown rot fungi and their effects, as pretreatments, on cellulolysis. Bioresource Technology 116: 147-154.

#### Funding statement

This work was supported by the U.S. Department of Energy Office of Science (Early Career Grant DE-SC0004012 from the Office of Biological and Ecological Research (BER) to J.S.S.; and BER Grant DE-SC0012742). Confocal microscopy is funded by user Facility grant #48607 at the Environmental Molecular Sciences Laboratory of Pacific Northwest National Laboratory (to J.S.S.).

### Distinct growth patterns and time-dependent secretome alterations by two taxonomically divergent brown rot fungi

Gerald N. Presley<sup>1</sup>, Jiwei Zhang<sup>1</sup>, and Jonathan S. Schilling<sup>1</sup>

<sup>1</sup>University of Minnesota, Saint Paul

URL: http://schillinglab.cfans.umn.edu

**Project Goals:** The overarching goal, within which this study's aims fit, is to discover which genes are differentially up-regulated across the mycelia of brown rot wood-degrading fungi *in planta*, particularly at the leading edge of wood decomposition. These unique fungi accomplish what we have difficulty achieving – energy from plant biomass. To do this, brown rot fungi apparently couple an oxidative pretreatment step with enzymatic saccharification in discrete space, via partitioning of reactions. We have previously shown evidence that these are governed by differential expression, but with genes putative, transport out of hyphae unclear, and the secretome poorly studied, historically, a comprehensive approach is needed that can also limit ('winnow') data sets from powerful global analytical tools to focus on the genes and pathways that matter. To address this, we are comparing global expression profiles among mycelial regions and matching what we see to what they do, in terms of changing wood physiochemistry. To map a 'connectome' in a wood-fungal interaction, specifically, we must also match gene expression patterns with the extracellular secretome and with physiochemical wood modifications.

Specific to this project, the **Objectives** were 1) to focus efforts on the secretome of brown rot fungi in planta in order to match gene expression patterns generated in whole-transcriptome efforts and 2) to expand our work on *Postia placenta* to include fungi from other distinct clades. Brown rot fungi have a polyphyletic history, and we make a significant assumption when extrapolating from *one* brown rot fungus to *all* brown rot fungi.

#### Abstract:

Brown rot fungi are wood-degrading fungi that employ both oxidative and hydrolytic mechanisms to degrade wood. Hydroxyl radicals that facilitate the oxidative component are powerful nonselective oxidants and are incompatible in discrete space with hydrolytic enzymes. Differential gene expression has been implicated in Postia placenta to segregate these reactions, but it is unclear if this two-step mechanism varies in other brown rot fungi with different traits and life history strategies, and that occupy different niches in nature. We used proteomics to analyze a progression of wood decay on thin wafers, using brown rot fungi with significant taxonomic and niche distance - Serpula lacrymans (Boletales; 'dry rot' lumber decay) and Gloeophyllum trabeum (order Gloeophyllales; slash, downed wood). Both fungi produced greater oxidoreductase diversity upon wood colonization and greater glycoside hydrolase activity later, consistent with a two-step mechanism. The two fungi invested very differently, however, in terms of growth (infrastructure) versus protein secretion (resource capture), with ergosterol/extracted protein ratio increased 7x with S. lacrymans than with G. trabeum. In line with their native substrate associations, hemicellulase specific activities were dominated by mannanase in S. lacrymans and by xylanase in G. trabeum. Consistent with previous observations, S. lacrymans did not produce GH 6 cellobiohydrolases (CBH) in this study, despite belonging to the order Boletales which is distinguished among brown rot fungi by having CBH genes. This work suggests that distantly

related brown rot fungi employ staggered mechanisms to degrade wood, but that the underlying strategies vary among taxa.

#### **References:**

Presley, G., Schilling, J.S. (in revision) Distinct growth and protein regulation strategies in two distantly-related brown rot wood-degrading fungi. Applied & Environmental Microbiology - AEM02987-16.

Presley, G.N., Zhang, J., Schilling, J.S.\* (online first) A genomics-informed study of oxalate and cellulase regulation by brown rot wood-degrading fungi. Fungal Genetics & Biology doi.org/10.1016/j.fgb.2016.08.004.

Zhang, J., Presley, G.N., Hammel, K.E., Menke, J.R., Hu, D., Ryu, J.S., Orr, G., Schilling, J.S. \*(2016) Localizing gene regulation reveals a staggered wood decay mechanism for the brown rot fungus *Postia placenta*. Proceedings of the National Academy of Sciences 113: 10968-10973.

Kaffenberger, J.T. and Schilling, J.S. (online first) Comparing lignocellulose physiochemistry after decomposition by brown rot fungi with distinct evolutionary origins. Environmental Microbiology doi:10.1111/1462-2920.12615.

Schilling, J.S., Kaffenberger, J.T., Liew, F.J., Song, Z. (2015) Signature wood modifications that reveal decomposer community history. PLoS ONE 10(3): e0120679. doi:10.1371/journal.pone.0120679.

Schilling, J.S., Duncan, S.M., Presley, G.N., Filley, T.R., Jurgens, J.A, and Blanchette, R.A. (2013) Colocalizing incipient reactions in wood degraded by the brown rot fungus *Postia placenta*. International Biodeterioration and Biodegradation 83: 56-62.

Schilling, J.S., Ai, J., Blanchette, R.A., Duncan, S.M., Filley, T.R., and Tschirner, U.W. (2012) Lignocellulose modifications by brown rot fungi and their effects, as pretreatments, on cellulolysis. Bioresource Technology 116: 147-154.

#### Funding statement

This work was supported by the U.S. Department of Energy Office of Science (DE-SC0012742 from the Office of Biological and Ecological Research (BER) to J.S.S.)

### Systems Analysis of the Physiological and Molecular Mechanisms of Sorghum Nitrogen Use Efficiency, Water Use Efficiency and Interactions with the Soil Microbiome

**Daniel P. Schachtman**<sup>1</sup> (Daniel.Schachtman@unl.edu), Rebecca Bart<sup>2</sup>, Thomas Brutnell<sup>2</sup>, Daniel Chitwood<sup>2</sup>, Asaph Cousins<sup>3</sup>, Jeffrey Dangl<sup>4</sup>, Ismail Dweikat<sup>1</sup>, Andrea Eveland<sup>2</sup>, Maria Harrison <sup>5</sup>, Stephen Kresovich<sup>6</sup>, Peng Liu<sup>7</sup>, <sup>2</sup>Todd Mockler, Jessica Prenni<sup>8</sup>, Balasaheb Sonawane<sup>3</sup>, Susannah Tringe<sup>9</sup>, and Arthur Zygielbaum<sup>1</sup>

<sup>1</sup>University of Nebraska-Lincoln; <sup>2</sup>Donald Danforth Plant Science Center; <sup>3</sup>Washington State University; <sup>4</sup>UNC-Chapel Hill; <sup>5</sup>Boyce Thompson Institute; <sup>6</sup>Clemson University; <sup>7</sup>Iowa State University; <sup>8</sup>Colorado State University; <sup>9</sup>Joint Genome Institute-DOE

#### **Project Goals:**

**The overall project goal** is to establish a foundational, systems-level understanding of plant, microbial, and environmental interactions that will lead to strategies for enhancing growth and sustainability of sorghum through genetic and microbial adaptations to water and nitrogen limited environments.

#### The specific goals are:

- 1. Conduct deep surveys of root microbiomes concurrent with phenotypic characterizations of a diverse panel of sorghum genotypes to define the associated microbes and the most productive lines under drought and low nitrogen conditions.
- 2. Associate systems-level genotypic, microbial, and environmental effects with improved sorghum performance using robust statistical approaches.
- 3. Develop culture collections of sorghum root/leaf associated microbes.
- 4. Test hypotheses regarding Gsorghum x Gmicrobe x E interactions in controlled environments.
- 5. Validate physiological mechanisms, map genetic loci for stress tolerance, and determine the persistence of optimal microbial strains under greenhouse and field conditions.

To compete in the biofuel energy market, cellulosic feedstocks will need to be high yielding and carbon neutral or negative while requiring low inputs. To avoid competition with existing food production systems, these crops will also need to be grown on marginal lands. This will require the introduction of novel traits to increase abiotic stress tolerance associated with marginal soils. This project is using multiple interdisciplinary approaches in varied settings – including the laboratory, controlled environments, and the field – to identify plant genes and sorghum associated microbes that will enhance the sustainable production of sorghum as a biofuel feedstock.

In 2016 the team conducted large field experiments and two large controlled environment studies on 30 sorghum varieties with several aims including: finding the most water and nitrogen use efficient germplasm, studying how the genotype of sorghum alters the microbiome, how low nitrogen and water deficit alters the microbiome, how transcript and metabolite profiles are influenced by abiotic stress and genotype. The field data has been processed and will be presented. DNA from roots, rhizosphere and soil has been extracted and

is currently be analyzed at JGI using iTAGS. Root and leaf material have been sent to the Danforth Center and Colorado State University for transcript and metabolite profiling. This poster will contain a description of the field trials, the biomass and height data as well as stable isotope data and plant nitrogen content.

This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant no DE-SC0014395

#### Harvesting the Root Microbiome of Grasses toward Sustainable Increases in Crop Production

Mucyn TS<sup>1</sup> (tmucyn@live.unc.edu), Breakfield NW<sup>1,2</sup>, Paredes SH<sup>1</sup>, Yourstone SM<sup>1</sup>, Hunter  $C^{1,3}$ , Schachtman DP<sup>4</sup>, Dangl JL<sup>1</sup>.

<sup>1</sup> University of North Carolina, Chapel Hill, NC; <sup>2</sup> Current address, NewLeaf Symbiotics, St Louis, MO; <sup>3</sup> Current address, BASF, Cary, NC; <sup>4</sup> University of Nebraska, Lincoln, NE.

http://labs.bio.unc.edu/Dangl/projects/

#### **Project Goals:**

The global demand for increased food commodities and energy supply highlight the necessity to further improve crop productivity. A variety of plant functions and traits are co-dependent on the microbial communities that exist within and around them. Characterizing, manipulating and /or mimicking the plant microbiota provide a promising path toward increasing crop production. The extent to which the host selects its root microbial community remains poorly understood, especially across vastly divergent species. Grasses constitute our main source of food, feed, and bioenergy, but how their microbiomes relate to, or may differ from other plant microbiomes remain unclear.

In order to determine the influence of plant host species on the assembling of both the rhizospheric (RZ) and endophytic (EC) microbiome, we monitored the bacterial root microbiome of Arabidopsis thaliana Col-0, Brachypodium distachyon Bd21 and Setaria viridis A10-1 through 16S amplicon sequencing. For a rigorous comparison, all three species were grown under identical conditions in the same wild soil. We observed, as expected, that bacterial communities vary mainly according to plant compartment (rhizosphere or endophytic compartment [EC]) and soil digs, and that all three plant species assemble an extremely similar bacterial endophytic communities. However, unweighted UniFrac distance analysis indicates that Brachypodium distachyon (Bd) and Setaria viridis (Sv) share a slightly more similar microbial EC communities compared to Arabidopsis thaliana (At). The Brachypodium EC compartment was especially marked by the presence of *Tenericutes*. The relative abundance of *Tenericutes* was much lower in Setaria and null in Arabidopsis. In an attempt to refine the variation of the bacterial root microbiome across species from the phylum level to the Operational TaxonomicUnit level, we established a stringent regularized generalized linear model to identify differentially abundant (DA) OTUs across species, and across compartments. One fourth of the DA OTUs identified were assigned to the Anaeroplasmatales order, which belongs to the phylum Tenericute, and these were found primarily more abundant in Bd EC samples when compared to At EC samples, thus confirming the robustness of our model. Most of the nonredundant DA OTUs identified in our data set are of very low abundance. Our result indicate that the bacterial EC root microbiome across plant species, even widely divergent species is highly similar at the 16S level and that the minor differences observed are mostly attributed to low

abundance OTUs. In addition, we are conducting in parallel fungal censuses as fungal communities associated with plant roots are also central to the host health, survival and growth. Fungal and bacterial communities most likely act in concert to influence plant heath and development. Interestingly, the *Tenericute* OTUs detected in our 16S census cluster with sequences from Molicutes-related bacteria (Mre) known to associate with arbuscular mycorrhizal fungus (AMF), and are mainly absent in Arabidopsis samples, in accordance with its lack of association with AMF.

We determined that the root microbiome is relatively conserved across plant species however it remains to be assessed if the microbiome may also present functional conservation across species. The Dangl lab has previously isolated ~600 bacterial strains from the rhizoplane and endophytic compartments of *Arabidopsis thaliana* and evaluated the effect of these isolates on plant health under various starvation stresses. Isolates shown to promote growth of Arabidopsis or alleviate nutrient stress response are being tested on *Setaria viridis*. However, our root-associated bacterial isolate collection was derived from plants grown in soils with relatively high nutrient values. To better the relationship between the microbiome and the plant nutritional stress response, and increase our odds to identify bacterial strains promoting plant health under nutrient stress we established a field experiment is which Setaria was grown under low and high Nitrogen side by side with various Sorghum accessions. The root (rhizoplane + EC) and, rhizosphere of those plants, as well as bulk soil, have been collected for bacterial isolation and 16S census.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0014395

### Increasing Growth Yield and Decreasing Acetylation in *Escherichia coli* by Optimizing the Carbon-to-Magnesium Ratio in Peptide-based Media.

David Christensen,<sup>1\*</sup> (dchristensen1@luc.edu), James Orr,<sup>2</sup> Christopher V. Rao,<sup>2</sup> and Alan Wolfe<sup>1</sup>

<sup>1</sup>Loyola University, Chicago, IL; <sup>2</sup>University of Illinois at Urbana-Champaign

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

 $N^{\epsilon}$ -lysine acetylation is a posttranslational modification that occurs within all three domains of life. The acetylation reaction occurs through the donation of an acetyl group from a donor molecule onto a susceptible lysine of a protein or peptide. This modification neutralizes the positive charge of the lysine side chain and increases its size. Acetylation of residues required for catalytic function can render an enzyme inactive. Additionally, neutralization of the positive charge can disrupt salt bridges necessary for protein-protein interactions. In *E. coli*, acetylation is known to be catalyzed by two mechanisms. One, the canonical enzymatic mechanism, utilizes the only known lysine acetyltransferase, YfiQ, to catalyze the donation of the acetyl group of an acCoA molecule onto a lysine. The other and more predominant mechanism employs acetyl phosphate (acP), the intermediate of the acetate fermentation (AckA-Pta) pathway, to donate its acetyl group onto proteins non-enzymatically. Therefore, conditions that promote acetate fermentation invariably lead to protein acetylation.

Previously, our lab and others have found that *E. coli* grown in carbon excess leads to high acetylation levels due to the production of acetate<sup>1,2</sup>. While studying protein acetylation in buffered tryptone broth supplemented with glucose (TB7/glucose), we observed that *Escherichia coli* did not fully consume glucose prior to stationary phase. However, when we supplemented this medium with magnesium, the glucose was completely consumed during exponential growth with concomitant increases in cell number and biomass but reduced cell size. Similar results were observed with other sugars and other peptide-based media, including lysogeny broth. Buffering was found to be necessary for the full magnesium growth yield increase because without buffering, acidification of the medium due to acetate excretion inhibited growth. Magnesium also limited cell growth for *Vibrio fischeri* and *Bacillus subtilis* in TB7/glucose.

Finally, magnesium supplementation reduced protein acetylation. Based on these results, we conclude that growth in peptide-based media is magnesium limited. We further conclude that magnesium supplementation can be used to tune protein acetylation without genetic manipulation. These results have the potential to reduce potentially deleterious acetylated isoforms of recombinant proteins without negatively affecting cell growth.

#### References

1. Weinert, B.T., et al., *Acetyl-phosphate is a critical determinant of lysine acetylation in E. coli*. Mol Cell, 2013. **51**(2): p. 265-72.

2. Kuhn, M.L., et al., Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. PLoS One, 2014. 9(4): p. e94816.

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

#### SYSTEMS LEVEL STUDY OF A NOVEL FAST-GROWING CYANOBACTERIAL STRAIN FOR NEXT GENERATION BIOFUEL PRODUCTION

Justin Ungerer<sup>1</sup>\* (JustinUngerer@wustl.edu), Cory Knoot<sup>1</sup>, Saratram Gopalakrishnan<sup>2</sup>, Lin Wang<sup>2</sup>, Mary Abernathy<sup>3</sup>, Whitney Hollinshead<sup>3</sup>, **Yinjie J. Tang<sup>3</sup>**, **Costas D. Maranas<sup>2</sup>**, and **Himadri B. Pakrasi<sup>1</sup>** 

<sup>1</sup>Department of Biology, Washington University, St. Louis, MO; <sup>2</sup>Department of Chemical Engineering, The Pennsylvania State University, University Park, PA; <sup>3</sup>Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO

https://sites.wustl.edu/photosynthbio/

Project Goal: The overall objective of this project is to use integrated systems biology and synthetic biology approaches to develop *Synechococcus* 2973, a fast growing cyanobacterial strain, as a platform organism for photobiological production of advanced biofuels and other useful chemicals. The project aims to improve the understanding of metabolic processes in this microbe through *in silico and in vivo* analysis followed by experimental validations to enable efficient strain design. Overall the goals are to (a) develop a genetic tool kit that will enable facile metabolic engineering of this strain, (b) measure photosynthetic parameters to identify factors that are critical to rapid growth, (c) reconstruct a genome-scale metabolic model of *Synechococcus* 2973, (d) to develop a carbon-mapping model of this strain, and (e) to better understand its phenotypical properties by isotopically nonstationary metabolic flux analysis (INST-MFA).

Cyanobacteria have great potential as green biofactories because they grow on carbon dioxide and sunlight alone which reduces greenhouse gas emissions and moves society away from dependence on petroleum-based products. Unfortunately, cyanobacteria display growth rates that are much slower than conventionally-used heterotrophic biofactories such as *E. coli* and yeast. This leads to an inherently-lower productivity from the cyanobacteria. The Pakrasi lab have recently identified a cyanobacterial strain, *Synechococcus elongatus* UTEX 2973, with significant industrial potential. *Synechococcus* 2973 exhibits autotrophic biomass productivity comparable to that of heterotrophs such as yeast. Under conditions of high light and high  $CO_2$ this strain has a doubling time of 1.5 hours. Genome sequencing revealed that *Synechococcus* 2973 is a close relative of the slower-growing strain, *Synechococcus elongatus* PCC 7942. The strains differ by 55 SNPs (only 35 of which are in coding regions), a 188 kb inversion and a 7.5 kb deletion.

The Pakrasi lab has developed an advanced CRISPR toolkit for *Synechococcus* 2973 to allow us to investigate the difference between the two strains in more detail. Using CRISPR technology, we performed a mutational analysis of *Synechococcus* 2973 to identify the subset of the 55 SNPs that grant the ability for rapid autotrophic growth. Four mutations were identified as essential for the fast-growth phenotype. Engineering these mutations individually into *Synechococcus* 7942 increases its growth rate; while combinations of the mutations exhibit additive effects. Three of the four mutations that we found to increase growth rate were also identified as factors that may contribute to rapid growth in the genome-scale model (GSM) that was developed by the Maranas group, suggesting that the GSM is highly valid.

The Maranas lab has developed a composite GSM for both Synechococcus 7942 and Synechococcus 2973 (iSyf686). This model serves as both a foundation for further modeling efforts as well as a means to interrogate possible factors contributing to the fast-growth phenotype. The inclusion of constraints based on experimental measurements of CO<sub>2</sub> uptake resulted in a ratio of the growth rates of Synechococcus 2973 to Synechococcus 7942 of 2.03, which nearly recapitulates the *in vivo* growth rate ratio of 2.13. The model was also used to identify four ORFs in Synechococcus 2973 with SNPs whose associated reactions have higher achievable fluxes, three of which were independently identified experimentally to contribute to the fast-growth phenotype. Additional insights into the biology of the fast-growing phenotype of Synechococcus 2973 can be obtained upon quantification of intracellular fluxes using isotopic non-stationary metabolic flux analysis (MFA). The Maranas lab has constructed the mapping models imSyn711 and imSyf608 based on GSMs for Synechocystis sp. PCC 6803 (iSyn731) and Synechococcus 7942 (iSyf686) respectively, which highlight cyanobacteria-specific carbon skeleton rearrangements. Upon flux quantification via genome-scale non-stationary MFA using imSyn711, we found that the oxidative branch of the pentose phosphate pathway was inactive, leaving photosynthetic light reaction as the sole source of reducing equivalents for anabolic processes. The model also predicts that a fraction of the serine pool was synthesized directly from 3-phosphoglycerate, whereas glycine was synthesized predominantly via photorespiration.

In the Tang lab, isotopically nonstationary metabolic flux analysis (INST-MFA), biomass compositional analysis, and metabolite profiling were performed comparing *Synechococcus* 2973 and *Synechococcus* 7942. The outcomes indicate a highly effective metabolism in *Synechococcus* 2973 compared to other model cyanobacteria. First, the flux maps demonstrate strong Calvin cycle, photorespiration, and pyruvate kinase activity, but minimal flux through malic enzyme and oxidative pentose phosphate pathways. Second, anabolism drains intermediate pools from central pathways under high light conditions, while central metabolism accumulates metabolites under suboptimal light (*i.e.*, energy metabolism, rather than carbon fixation pathways, constrains fast cyanobacterial growth). Third, *Synechococcus* 2973 shows similar genetic background and flux ratios to *Synechococcus* 7942, but exhibited greater carbon assimilatory and photorespiratory flux, less accumulation of glycogen, and potentially metabolite channeling that together result in increased biomass growth. Finally, *Synechococcus* 2973 has weak flux through a linear TCA pathway and small pool sizes of acetyl-CoA/TCA intermediates under all growth conditions. Such metabolic features support a photosynthesis platform to produce valuable products from its sugar phosphate pathways.

We are currently exploring the potential for *Synechococcus* 2973 as a production vehicle for alkane biofuels. All cyanobacteria utilize one of two primary biosynthetic pathways to synthesize  $C_{15}$  to  $C_{19}$  alkanes or alkenes from fatty acid precursors. We are studying these pathways in *Synechococcus* 7942 and 2973 in order to understand the best means to engineer the cells to produce more of these valuable metabolites. The biochemical and genetic studies are being coupled with insights gained from the GSMs and metabolic flux analyses.

These studies have been supported by funding from the Office of Biological and Environmental Research in the DOE Office of Science to HBP, YJT and CDM.

### Integrated and Dynamic Multispectroscopic *In Situ* Imaging of Plant Metabolism at the Level of Subcellular Compartments

Geng Ding<sup>1</sup>\* (gengding@iastate.edu), Diane Bassham,<sup>1</sup> Robert Houk,<sup>1</sup> Young-Jin Lee,<sup>1</sup> Jacob Petrich,<sup>1,2</sup> Emily Smith,<sup>1,2</sup> Arthur Winter,<sup>1</sup> Eve Wurtele,<sup>1</sup> and **Basil Nikolau**<sup>1</sup>

<sup>1</sup>Iowa State University, Ames, IA; and <sup>2</sup>The Ames Laboratory, Ames, IA

http://www.insituimaging.org/

Project Goals: This multi-disciplinary team is developing new integrated multi-spectral imaging technologies that will assess and quantitatively model metabolic processes that are non-symmetrically distributed at the cellular and subcellular levels of plant organs. The imaging technology is being developed in the context of computational capabilities that will integrate multi-spectral images with genome scale modeling and thus contribute to the better understanding how biomass-based biofuel producing metabolic pathways are interconnected and controlled within topological constraints in spatially defined membrane-bounded regions within plant cells.

This project is developing and applying integrated molecular imaging technologies that can be used to monitor membrane lipid remodeling. Understanding the remodeling of membrane lipid topology in plant cells has major consequence in optimizing plant biomass productivity. The integrated molecular imaging technologies is being developed in the biological context of autophagy that remodels membrane lipid topologies that control spatially defined subcellular regions within plant cells and optimizes plant biomass productivity in response to environmental signals. Genetic stocks that will enable the dissection of membrane lipid dynamics have been identified and analyzed to identify specific target lipid molecules for molecular imaging. Analytic technologies for imaging these specific target lipid molecules via fluorescence, Raman and mass spectroscopy have been established. These imaging technologies are being developed in the context of computational capabilities that will integrate the multi-spectral images with genome scale models. We have established an infrastructure that ensures transparent collaboration among the students of different collaborating groups. Defined tasks to fulfill the goals of the project, and these are expanded as follows.

Task 1: Genetic and biochemical analysis of defined autophagy and lipid metabolism genes. The initial analyses focused on identifying specific lipid molecules that are affected by the autophagy induced dynamics of cellular membranes. The rationale being that these specific lipid molecules will be targeted for imaging via the technologies that will be developed in Tasks 2-4.

Task 2: Develop and apply in situ optical imaging platforms. The team has demonstrated the first-of-its-kind, red-releasing photocage. Photocages are compounds that release a cargo or generate a change in a signal when exposed to light. For use in plant systems, it is desirable to design photocages that release their cargo when exposed to a range of visible light wavelengths

that are not interfered by the endogenous optically active biomolecules (e.g., chlorophylls, carotenoids). This technology should enable multiple cargos to be released independently with different wavelengths of light or to increase the penetration depth of the light that generates the signal. Longer wavelengths of light are associated with deeper penetration depths in tissues, so a red-release photocage has this benefit. The team synthesized and demonstrated the use of the red photocage in a biological system.

Task 3: Chemical synthesis and tuning of self-destructing fluorophores for the in situ visualization of dynamic events. A new class of fluorescent chemical imaging probes capable of in situ imaging have been synthesized. In particular, we have designed and synthesized a new class of photocages derived from BODIPY dyes capable of dynamic fluorescence imaging using visible light. These probes release compounds with visible light irradiation with wavelengths spanning the visible and entering the near-IR.

Task 4: Spatial mapping of metabolites via mass-spectrometry. We have focused on optimizing the performance of atmospheric pressure mass spectrometry imaging (MSI) and are applying this imaging technology to spatially map specific lipid-metabolites. These optimizations have reduced the laser spot size to approximately 50  $\mu$ m, as compared to the 125  $\mu$ m spot size that was available at the start of the project. In addition, we have focused on finding a technique that can be used to integrate information from different imaging platforms, i.e. mass spectrometry, Raman and fluorescence microscopy, and optical microscopy, which is essential for multimodal image comparison. This is especially important because MSI has far lower image resolution than optical imaging techniques.

Task 5: Develop computational imaging visualization platform. We are establishing a database with a visualization platform that the broader research community in different fields of chemistry, chemical engineering, biochemistry, and biology can readily access, comprehend and explore the data obtained from the combined and diverse analytical chemistries in the context of the biological materials under study.

*This work is supported by the Department of Energy, Office of Science, Bioimaging Technology Program under award number DE-SC0014038.* 

### Monitoring fluxes of atmospherically-reactive gases (CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O) during the conversion of grasslands into a biofuel crop (*Panicum virgatum*)

Arthur Escalas<sup>1</sup>, Colin Bates<sup>1</sup>, Liyou Wu<sup>1</sup>, Don Herman<sup>2,3</sup>, Yuan Wang<sup>4</sup>, Lauren Hale<sup>1</sup>, Chi Myoung-Hwan<sup>4</sup>, **Malay Saha<sup>4</sup>**, **Kelly Craven<sup>4</sup>**, **Jennifer Pett-Ridge<sup>5</sup>**, **Mary Firestone**<sup>2,3</sup> and **Jizhong Zhou**<sup>1,3</sup>\* (jzhou@rccc.ou.edu)

<sup>1</sup>University of Oklahoma, Norman, Oklahoma; <sup>2</sup>University of California, Berkeley, California; <sup>3</sup>Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California; <sup>4</sup>Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, Oklahoma; <sup>5</sup>Nuclear and Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, California

#### URL: <u>http://www.ou.edu/ieg/</u>

Project Goals: Our project, *Establishment to senescence: plant-microbe and microbe-microbe interactions mediate switchgrass sustainability*, aims to understand the bases of switchgrass productivity in marginal soils by dissecting key molecular mechanisms that differentiate soil organisms associated with superior switchgrass genotypes adapted to a range of resource limitations. We hypothesize that successful establishment and sustainable cultivation of switchgrass in marginal soils is in part enabled by beneficial plant-microbial interactions, and that key ecosystem services ranging from C sequestration, increased soil fertility, and reduced trace gas production result from networks of plant-microbial interactions. We seek a mechanistic understanding of the interaction networks occurring within the switchgrass rhizosphere and their effects on ecosystem sustainability.

Switchgrass (*Panicum virgatum L.*), a perennial grass native to the tallgrass prairie, is one of the most promising bioenergy crops in the U.S. It's successful cultivation on marginal soils unsuitable for traditional agricultural crops has been identified as an important goal to meet the US Department of Energy's goal to replace 30% of petroleum-based transportation fuels with biofuel by 2030. In order to fully evaluate the sustainability of switchgrass-based biofuel production, we need to assess the consequences of transforming "natural" ecosystems such as grasslands into bioenergy crops. To that end field measurements of greenhouse gases (GHG) are needed to estimate the net GHG balance of biofuel production. Bioenergy crops in general are known for influencing the soil in which they grow and switchgrass in particular has been shown to potentially increase carbon sequestration in soils. Ultimately, this can further reduce GHG emissions associated with substituting renewable energy for fossil energy.

We present the results of 6 months of monitoring trace gas (CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O) fluxes during the establishment of switchgrass in two sites characteristic of the US southern plains. In each site, we compared background grassland with growing switchgrass cultivated using low-management practices. Additionally, we are interested in characterizing the effects of switchgrass establishment on the soil physical-chemical parameters along with soil microbial communities (bacterial and fungal). Thus, the

main objectives of this study are to determine the effects of transforming a low-productivity grassland into a switchgrass field in terms of (i) soil physico-chemical characteristics, (ii) fluxes of GHG (CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O) and (iii) structure of the soil microbiome (bacteria and fungi). Ultimately, we will relate these three components to better understand and characterize the dynamics of GHG during the conversion of grassland into bioenergy crops.

Our two study sites, located in southern Oklahoma, represent marginal lands and exhibit differences in soil physical and chemical variables (e.g. texture, organic matter content, pH). In both sites, two plots (27x22m) were delimited: one consists of a "fallow" in which natural vegetation is not disturbed and the other contains 500 switchgrass plants (Alamo variety) planted a meter apart in a honeycomb design. The fluxes of three major trace gases, carbon dioxide (CO2), methane (CH4), and nitrous oxide (N2O) were measured using a recently developed cavity ring-down spectroscope (CRDS). This transportable instrument can measure concentrations down to part-per-billion for each gases species and provides high resolution measurements (every 2s). To capture both spatial and temporal variability, we used a non-steady-state chamber approach in a highly-replicated design (21 chambers per plot) along with a monthly sampling frequency. Simultaneously with trace gases sampling, we collected soil samples for physical-chemical characterization: moisture, pH, NH<sub>4</sub>, NO<sub>3</sub>, TN, TC, organic matter. Additionally, our two sites are located close to two meteorological towers from the Oklahoma Mesonet network, which provide us five-minute resolution data for more than twenty variables including precipitations, air and soil temperature and solar radiation. All these data will be combined with data characterizing composition of the microbial communities (by Illumina MiSeq sequencing of 16S and ITS marker genes for bacteria and fungi, respectively).

Preliminary results suggest differences in the trace gases flux dynamics across the two sites along with as well as a strong effect of the plant. Besides providing a better understanding of the effect of grassland transformation into switchgrass fields, our data will be used to estimate greenhouse gas budgets and model the switchgrass ecosystem.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number DE-SC0014079 to UC Berkeley, University of Oklahoma, and Samuel Noble Foundation, Lawrence Livermore National Lab and Lawrence Berkeley National Lab.

### Plant-microbe and microbe-microbe interactions mediate switchgrass sustainability: following rhizosphere microbial communities during switchgrass establishment

Yuan Wang<sup>1</sup>, Erin Nuccio<sup>2</sup>, Josh Barbour<sup>1</sup>, Arthur Escalas<sup>3</sup>, Colin Bates<sup>3</sup>, Lauren Hale<sup>3</sup>, Na Ding<sup>1</sup>, Katerina Estera<sup>4</sup>, Nameer Baker<sup>4</sup>, Javier Ceja Navarro<sup>6</sup>, Abelardo Arellano<sup>6</sup>, Kateryna Zhalnina<sup>5</sup>, **Eoin Brodie<sup>6</sup>**, **Trent Northen<sup>5</sup>**, **Wolf Scheible<sup>1</sup>**, **Michael Udvardi<sup>1</sup>**, **Jennifer Pett-Ridge,<sup>2</sup> Malay Saha<sup>1</sup>**, **Jizhong Zhou,<sup>3</sup> Mary Firestone<sup>4</sup>** and **Kelly Craven<sup>1\*</sup>** (kdcraven@noble.org)

<sup>1</sup>Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, Oklahoma; <sup>2</sup>Lawrence Livermore National Laboratory, Livermore, California; <sup>3</sup>University of Oklahoma, Norman, Oklahoma; <sup>4</sup>University of California, Berkeley, California and <sup>5</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, <sup>6</sup>Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, California.

Project Goals: Switchgrass (SG; *Panicum virgatum* L.) is a perennial grass native to the tallgrass prairie and one of the most promising bioenergy crops in the U.S., with potential to provide high-yield biomass on marginal soils unsuitable for traditional agricultural crops. A persistent concern for bioenergy cultivation of SG with low-input management, is improving seedling establishment and resistance to abiotic and biotic stresses. We hypothesize that successful establishment and sustainable cultivation of SG in marginal soils is in part enabled by beneficial plant-microbial interactions. We are investigating the succession of rhizosphere microbial communities associated with high- and low-performing SG plants grown in nutrient-limited soils at two Oklahoma field sites. The outcome of this research will provide a better genomic basis for SG cultivation in marginal soils, and will expand our knowledge of the interactions between soil microbiomes, plants and ecosystems.

In the soils surrounding roots (rhizosphere), biotic, chemical and physical drivers enrich for specific bacterial and fungal communities. These organisms play multiple roles, and some may benefit plant productivity via help with nutrient acquisition, water uptake and pathogen suppression. We are investigating the composition, function and succession of rhizosphere microbial communities during plant cultivation in an effort to better understand the plant-microbiome interactions that enable plant survival and adaptation under stressed conditions, such as drought and nutrient deficiency.

To study the establishment phase plant-soil microbiome characteristics of SG growing in 'marginal' nutrient or water-limited soils, we selected two Noble Foundation research farms, both remnants of the Dust Bowl Era in Oklahoma. One, Red River Farm has sandy loam soil low in NO<sub>3</sub>-N and organic matter; the other, Third Street Farm, has a silt loam soil with relatively low phosphorus availability. Five hundred Alamo AP13 non-clonal seedlings were planted into each field in May-June of 2016. Other than hand weeding during the summer, no management or water/ nutrients was supplied to the fields. Thirty plants were randomly selected for multi- time point monitoring of rhizosphere and bulk soil samples to study community succession during SG establishment. These same plants were non-destructively sampled over the first growing season at 5 time points: early and late vegetative growth, reproductive growth, maximal growth, and senescence. To test the association between rhizosphere communities and SG growth, plant growth parameters such as plant height, tiller numbers and flowering time were measured; biomass yield was assessed at the end of growth season. Within each plot plant growth was

highly variable, with between 20-143 tillers per plant. Overall, plants at the Red River site had better growth and higher biomass yield than at Third Street.

To better understand factors influencing the establishment and functional properties of rhizosphere microbial communities, we conducted amplicon sequencing of marker genes specific to bacteria, fungi and soil eukaryotes, and monitored physio-chemical characteristics of soil beneath targeted plants, including gravimetric moisture, pH, and NO<sub>3</sub>/NH<sub>4</sub>.We are structuring our analyses to assess differences in environmental drivers, community assembly, and succession patterns in both high- and low- performing plants on our two soils of different nutrient status. A subset of samples of particular interest will be characterized using shotgun metagenomics and GeoChip hybridization to further characterize the microbial functional properties.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number DE-SC0014079 to the UC Berkeley, the Samuel Roberts Noble Foundation, the University of Oklahoma, the Lawrence Livermore National Laboratory and the Lawrence Berkeley National Laboratory. Part of this work was performed at Lawrence Berkeley National Lab under contract DE-AC02-05CH11231 and at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

#### Dissecting the Chemistry of Switchgrass-Microbe Interactions Using Cultivation, Exometabolomics and Mass Spectrometry Imaging

Kateryna Zhalnina<sup>1</sup>\* (kzhalnina@lbl.gov), Jian Gao,<sup>1</sup> Nasim Masoori,<sup>2</sup> Yuan Wang,<sup>3</sup> Erin Nuccio,<sup>4</sup> Rachel Neurath,<sup>5</sup> Chi Myoung-Hwan,<sup>3</sup> Tina Ding,<sup>3</sup> Lauren Hale,<sup>6</sup> Arthur Escalas,<sup>6</sup> Malay Saha,<sup>3</sup> Dominique Loque,<sup>2</sup> **Zhizhong Zhou**,<sup>6</sup> **Kelly Craven**,<sup>3</sup> **Jennifer Pett-Ridge**,<sup>4</sup> **Mary Firestone**, <sup>1,5</sup> **Eoin Brodie**<sup>1,5</sup> and **Trent Northen**<sup>1</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, California; <sup>2</sup>Joint BioEnergy Institute, Emeryville, California; <sup>3</sup> The Samuel Roberts Noble Foundation, Ardmore, Oklahoma; <sup>4</sup>Lawrence Livermore National Laboratory, Livermore, California; <sup>5</sup>University of California, Berkeley; <sup>6</sup>The University of Oklahoma, Norman

Project Goals: Our project works towards a fundamental understanding of the molecular mechanisms driving beneficial plant-microbial interactions in superior switchgrass genotypes adapted to a range of resource limitations. Plant-microbe interactions are examined during establishment to gain insight into how symbiotic and associative microbes improve plant performance and carbon stabilization in marginal soils. We will combine focused (single plant-microbe pairing) and 'community' systems biology approaches to examine the complex interplay among plants, microbes, and their physio-chemical environment.

Plant-soil-microbial interactions regulate the availability of nutrients and carbon stabilization in soil. Plants exude a diverse range of compounds into the soil surrounding their roots; these exudates are thought to attract and support microorganisms that may improve plant nutrient acquisition, drought tolerance, and resistance to pathogens. Here we used mass spectrometry-based metabolomics to identify key chemical mechanisms that drive bidirectional plant-microbe interactions in soil. These include nutrient uptake, metabolite exchange, and their patterns of regulation that influence plant productivity, adaptability to environmental change and the stability of carbon in soil.

Taking a multi-scale approach including field, greenhouse and highly controlled lab experiments our goal is to identify the chemical mechanisms that underlie plant-soil-microbial relationships. We are comparing switchgrass growth and exudation properties across these scales; creating a library of rhizosphere microbial isolates and combining these with plants in controlled systems to dissect their interactions and to link biology of specific microorganisms to the physiology and biochemistry of switchgrass.

Field, switchgrass exudation and soil metabolites. To identify how exudation of switchgrass changes during plant development in soil, we have collected samples from the rhizosphere of switchgrass and bulk soil during the first six months of switchgrass establishment (June-November) in two different field locations in Oklahoma. Currently these metabolites are being analyzed using mass spectrometry based metabolomics. Metabolite-microbe-mineral interactions are essential for processes involved in carbon stabilization in soil. To gain insights into these processes, we are examining soil metabolites across various soil depths and will use these samples to assess how the profile changes after several years of switchgrass cultivation.

Acquired data will be used to inform our lab experiments in highly controlled systems and to link back the results from the lab to the environmental conditions in the field.

**Switchgrass rhizosphere isolates**. We have now isolated more than 1000 heterotrophic bacteria from the rhizosphere of switchgrass and bulk soil using five media with different nutrient content to cover a variety of nutritional niches that may exist in the root zone. These isolates have been isolated in pure culture and preserved in glycerol stocks. Currently we are identifying representatives from this collection, which will be genome sequenced and used for further experiments.

**Switchgrass exudation, hydroponics**. Root exudation plays a major role in how plants alter the soil environment and define its rhizosphere microbiome. To complement our field analyses, we are analyzing switchgrass exudates during early plant development using a hydroponic approach.

**Switchgrass mesocosm studies.** We are using laboratory ecosystems to visualize bacteria and root growth and also enable collection and measurement of metabolites. Currently we are testing switchgrass growth conditions, exudation and response to the microbial inoculation using this platform. In future we plan to use the same laboratory systems to examine specific microorganisms isolated from the rhizosphere of switchgrass.

This research is based upon work supported by the U.S. Department of Energy Office of Science award DE-SC0014079 to UC Berkeley, the Samuel Noble Foundation, University of Oklahoma. Work at Lawrence Berkeley National Laboratory was performed under U.S. Department of Energy Contract No. DE-AC02-05CH11231. Work at Lawrence Livermore National Lab was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344.

#### Faunal Population Dynamics Throughout Switchgrass Developmental Stages

Javier A. Ceja-Navarro (JCNavarro@lbl.gov)<sup>1\*</sup>, Abelardo Arellano<sup>1</sup>, Nhu Nguyen<sup>2</sup>, Jennifer Pett-Ridge<sup>4</sup>, Mary Firestone<sup>3</sup>, Eoin Brodie<sup>1,3</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>2</sup>University of Hawaii, Honolulu, HI, <sup>3</sup>University of California, Berkeley, CA; <sup>4</sup>Lawrence Livermore National Laboratory, Livermore, CA.

Project Goals: The project, *Establishment to senescence: plant-microbe and microbe-microbe interactions mediate switchgrass sustainability*, aims to understand the bases of switchgrass productivity and the biotic and abiotic factors that mediate its establishment in marginal soils. With this research, we seek to define the multitrophic interactions that occur in the rhizosphere of superior switchgrass genotypes adapted to resource limitations and to elucidate how these interactions facilitate plant establishment.

Switchgrass (SG; *Panicum virgatum* L.), is one of the most promising bioenergy crops in the U.S., with the potential to provide high yield biomass while also improving the physical and hydraulic properties of marginal soils unsuitable for traditional crops. A persistent concern for bioenergy cultivation of SG, with low-input management, is improving seedling establishment and resistance to biotic and abiotic stress, a goal that may be reached by enabling beneficial interactions between the plant and the trophic compartments of biological activity in soil. This project is applying a holistic approach to define and understand the trophic interactions that occur during plant establishment and plant developmental stages, by considering that populations of nematodes, protists, and arthropods are critical elements of biological activity that together with free-living and plant-associated microbes define soil and plant productivity.

Our goal is to build robust mathematical models of soil trophic networks that extend beyond bacteria and fungi, and thus require information on faunal community composition, body size, and feeding preference amongst other parameters. As a first step for the study of the multitrophic interactions that mediate SG establishment and productivity, we are analyzing faunal dynamics during SG developmental stages using two primary approaches **1.** Funnelomics: where faunal populations are screened using traditional methods. With this approach populations of arthropods and nematodes are physically isolated from soil samples using Berlese funnels, and wet sieving with sucrose gradient centrifugation respectively. Arthropods are individually imaged, measured, and their DNA extracted for identification. Nematodes are subsampled and pooled into populations for DNA extraction, sequencing, and identification. **2.** Molecular Fauna: where bulk molecular approaches are used to obtain a general inventory of faunal populations to begin to establish hypothetical nodes in these trophic networks. Here, ten grams of soil are used for DNA extraction using in-house protocols that target arthropods, nematodes, and protists, and the extracted DNA is analyzed using metabarcoding (universal metazoan primers), and for the

quantification of different faunal populations using quantitative PCR (qPCR) with primers designed to be group-specific based on the results of "Funnelomics".

Soil samples were collected from SG planted and fallow plots over 1, 2, 6 and 7 months postseedling planting. During each sampling point four plants (SG treatment) and four fallow locations, were randomly selected and soil was collected to a depth of 20 cm within a 10-squared cm area. Fifty grams of soil were stored at -80°C for subsequent DNA extraction, 300 g of soil used for arthropod extraction, and 50 g for nematode isolation. A total of 250 individual arthropods have been individually imaged/measured and a collection of 36 pools of nematodes obtained. Protocols for the extraction of nucleic acids from large soil samples have been optimized and the DNA extracted.

A critical step for the identification of metazoan nodes in a trophic network is to identify appropriate PCR primers that will target the largest number of metazoan groups. We have tested different combinations of markers and primer combinations for cytochrome oxidase I, and 18S rRNA which are the markers most commonly used for metazoan identification. These primers were tested against different arthropod groups to identify primers that provide the best coverage and that will be used as 'universal primers' for subsequent analyses.

Screening of the isolated arthropod specimens has identified nine species of mites belonging to the order Mesostigmata (free-living predatory mites); eight species of mites from the order Oribatida (plant and fungal feeders); five different species of springtails from the families Onychiuridae and Isotomidae (fungal and root-feeders). Different specimens identified as members of family Agrypninae (including leaf-feeding beetles that feed on roots during their larval stages), together with arthropods from the infraclass Paraneoptera (Halothrips sp., Caliothrips sp., and Liposcelis sp.), some of which are known plant pathogens that feed and breed in the flowers of grasses and agricultural crops. Analyses of nematode populations and molecular fauna experiments are in progress.

Results to date have identified important arthropods groups whose presence relates to different developmental stages of SG. The emergence of these faunal groups may have consequences for plant productivity, rhizosphere microbial communities and the dynamics of other faunal nodes in the soil trophic network. The approaches developed here will provide the foundation for molecular approaches to quantitative study of soil trophic networks, also with potential applications as diagnostic tools to identify and intervene for the early control of plant pathogenic arthropods and nematodes in agricultural systems.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number DE-SC0014079. Work was performed at Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

#### EPICON: Epigenetic Control of Drought Response in Sorghum

### Peggy G. Lemaux (lemauxpg@berkeley.edu)<sup>1\*</sup>, Jeffrey Dahlberg<sup>2</sup>, Devin Coleman-Derr<sup>1,3</sup>, Robert Hutmacher<sup>4</sup>, Christer Jansson<sup>5</sup>, Elizabeth Purdom<sup>1</sup>, John Taylor<sup>1</sup>, Axel Visel<sup>6</sup>, and John Vogel<sup>1,6</sup>

<sup>1</sup> University of California, Berkeley CA; <sup>2</sup>University of California, Kearney Ag Research & Extension Center, Parlier CA; <sup>3</sup>USDA-ARS Plant Gene Expression Center, Albany CA; <sup>4</sup>University of California, West Side Research & Extension Center, Five Points CA; <sup>5</sup>EMSL-Pacific Northwest National Laboratory, Richland WA; and <sup>6</sup>DOE Joint Genome Institute, Walnut Creek CA.

Project Goals: EPICON researchers are studying epigenetic control mechanisms in the temporal and spatial response to water-limiting conditions in field-grown, pre-flowering and post-flowering drought-tolerant *Sorghum bicolor* varieties. Investigators are also searching for changes in microbial communities, including bacterial and fungal microbiomes, in soil, rhizosphere, roots and leaves associated with drought-stressed sorghum,. The primary focus of these efforts is to understand the roles that epigenetic signals play in acclimation to and recovery from drought. We will investigate effects of this abiotic stress on transcriptional networks and molecular profiles *in planta*, using a wide array of analytical tools, *i.e.*, RNA-Seq, smRNA-Seq, ChIP-Seq, BS-Seq, LC-Orbitrap-MS, and GC-MS. Similarly, function of microbial populations will be inferred from results of metagenomics and metatranscriptomics. Based on outcomes, we will identify genes and molecular markers to devise genetic strategies for improving drought tolerance in sorghum and other crops. From the cumulative data, we will develop models to better understand, predict and control the coordinated role of epigenetics and the phytobiome in sorghum's response to drought.

Genetic manipulation of crops via classical breeding and mutagenesis has focused primarily on using modifications of the plant's DNA sequence to increase the presence or activity of desirable traits. However, there is increasing published research indicating that environmental responses and plant development are also mediated by epigenetics, the process by which heritable changes in phenotype and/or gene expression are accomplished without changing the underlying DNA sequence. With particular relevance to the EPICON project, epigenetic changes have been shown to play a major role in regulating plant responses to drought, an abiotic stress likely to increase in frequency and severity in the future due to climate change, resulting in major challenges for world agriculture. We hypothesize that exposure of plants to abiotic stresses, including water limitation, triggers cascades of epigenetic changes. These include remodeling of chromatin, the network of DNA, RNA and proteins making up chromosomes, coupled with related changes in regulatory mechanisms, including small non-coding RNAs.

Efforts of EPICON researchers are concentrated on discovering the temporal and spatial influence epigenetic signals play in acclimation to and recovery from drought by identifying effects on individual transcription factors and/or transcriptional networks. To achieve this goal, we are studying the responses of two *Sorghum bicolor* cultivars that differ in their tolerance to pre-flowering and post-flowering drought stress. Sorghum, a widely cultivated cereal recognized for its drought and flood tolerance, offers notable advantages as a bioenergy feedstock because of its relatively reduced environmental footprint compared to its close relative, maize. In our experimental design, drought conditions are being imposed on sorghum in the field in California's Central Valley, where lack of summer rainfall makes well-controlled studies in drought nurseries possible. In the first year of field experiments, the two cultivars were planted in a split plot design in triplicate, along with a fourth fail-safe replicate; both varieties were subjected to a normal watering regime, as well as pre-flowering and post-flowering drought treatments. Phenotypic measurements were conducted throughout the growth cycle (early June - late September) to chart, for example, the flowering time, growth, and grain and biomass yield. Initial statistical analyses of agronomic data indicate that pre- and post-flowering drought stress impacts various agronomic characteristics,

compared to plants experiencing normal watering conditions. Pre-flowering drought stress impacted the yield of both BTx642 and RTx430; however, RTx430 showed a greater tolerance to pre-flowering stress compared to BTx642. This finding is consistent with the reported pre-flowering drought stress tolerance in RTx430. BTx642, a post-flowering, stay-green variety, had yields equal to or slightly higher in drought-stressed plots than BTx642 under controlled watering conditions, while RTx430 showed a reduction in yield. Pre-flowering and post-flowering stress negatively impacted both varieties, compared to the plants grown under normal watering conditions, with regard to other agronomic measurements, including height, flowering and biomass.

For molecular analyses, leaf, root and soil samples were taken weekly at precise intervals during sorghum development. Leaf and root samples were collected to track spatiotemporal changes in epigenetic, transcriptomic, metabolomic and proteomic footprints, that are currently being analyzed using RNA-Seq, smRNA-Seq, ChIP-Seq, BS-Seq, LC-Orbitrap MS, and GC-MS. As potential molecular mechanisms and specific targets are identified *in planta*, engineering of genes and pathways in sorghum will be used to validate suggested findings. In addition, shifts in the composition and activity of sorghum-associated and soil-based bacterial and fungal communities throughout the drought period are being monitored to determine if changes in membership or functional capacity of microbes in the soil, rhizosphere, roots and leaves correlate with epigenetic, transcriptional or metabolomic variation in the plant. Changes in the microbiome are being identified using Illumina iTag, amplicon sequencing of ribosomal repeats (16S rRNA for bacteria and ITS2 for fungi). Function is being inferred via shotgun metagenomic and metatranscriptomic analysis of rhizosphere communities.

Once compiled, analysis of the entire data set, year by year and over three years, will provide a more thorough understanding of the processes responsible for restructuring the metabolic and regulatory landscape of sorghum during drought and the relationship of these processes to drought tolerance. These efforts will lead to achievement of our ultimate goal – to identify key transcriptional regulators and pathways that control drought tolerance and to characterize their mechanisms of action, both *in planta* and potentially their relationship with the associated microbial communities. Additionally, these efforts should uncover phenotypic biomarkers associated with drought tolerance, which can be used to monitor and to correlate with phenotypic changes in large breeding populations. Identified genetic targets and their regulatory pathways will be utilized in future efforts to improve growth and biomass production of sorghum and other crops under water-limiting conditions, using transgenic and genome-editing approaches.

The EPICON project is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

### Switchgrass Growth and Transcriptomic Responses to Nitrogen Availability and the Rhizosphere Microbiome

Alan Bowsher,<sup>1\*</sup> (bowsher1@msu.edu), Maren Friesen,<sup>1</sup> Lisa Tiemann,<sup>1</sup> and Sarah Evans<sup>2</sup>

<sup>1</sup>Michigan State University, East Lansing, MI; <sup>2</sup>Kellogg Biological Station, Hickory Corners, MI.

http://rhizosphere.msu.edu/

## Project Goals: The main goal of the over-arching project, "Connecting nitrogen transformations mediated by the rhizosphere microbiome to perennial cropping system productivity in marginal lands", is to better understand the connections between plant and microbial genomics, transcriptomics, and productivity in the context of varying soil nitrogen availability.

Switchgrass (*Panicum virgatum* L.) has received substantial research attention as a potential bioenergy crop. In particular, switchgrass is noted for its relatively low soil fertility requirements, allowing the possibility of production in marginal lands to avoid competition with existing crops. In addition to the greenhouse gas benefits resulting from the use of switchgrass as an alternative to fossil fuels, soil fertility management strategies utilizing biologically-fixed rather than industrially-fixed N sources could help to maximize those benefits. Although it has been suggested that interactions with N<sub>2</sub>-fixing microbes in the rhizosphere may increase nitrogen acquisition by switchgrass (Bahulikar et al. 2014), few empirical studies have investigated these interactions or the factors influencing them. Here, we conducted a greenhouse study of switchgrass variety "Cave-in-Rock", factorially-manipulating nitrogen availability as well as the presence/absence of marginal land soil microbes. We predicted that if soil microbes improve N availability in the switchgrass rhizosphere, any beneficial effects of soil microbes would be most evident in the low N treatment. Measurement of aboveground biomass, leaf traits, and root morphology is currently underway. We will also assess root transcriptomes to characterize switchgrass responses to nitrogen deficiency and microbial inoculation, and soil metagenomes in order to characterize the microbes functioning in the switchgrass rhizosphere. Lastly, we are characterizing root exudates of switchgrass grown in sterile culture under both high and low nitrogen conditions to identify potential signaling molecules or growth substrates for rhizosphere microbes. We expect the results of this study to provide insight into possible mechanisms supporting productivity of switchgrass in low-fertility soils.

#### References

1. Bahulikar RA, Torres-Jerez I, Worley E, Craven K, and Udvardi MK. 2014. Diversity of nitrogen-fixing bacteria associated with switchgrass in the native tallgrass prairie of Northern Oklahoma. Applied and Environmental Microbiology 80: 5636-5643.

The Project entitled "Connecting nitrogen transformations mediated by the rhizosphere microbiome to perennial cropping system productivity in marginal lands" is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

#### The Contribution of Alternative Nitrogenases to Nitrogen Fixation in Switchgrass Rhizospheres on Marginal Lands

Norman JS,<sup>1,\*</sup> (jsnorman7@gmail.com), West WE,<sup>1</sup> Tiemann LK,<sup>1</sup> Friesen ML<sup>1</sup> and Evans SE<sup>1</sup>.

<sup>1</sup>Michigan State University, East Lansing, MI

#### **Project URL:**

http://rhizosphere.msu.edu

#### Project Goals (applicable to the MMPRNT project):

 What is the Impact of the Rhizosphere Microbiome on Switchgrass Nitrogen Status?
Does the Chemistry and Rate of Root Exudation Influence Rhizosphere Nitrogen Transformations?
How are Switchgrass Functional Traits Influenced by Nitrogen Availability and Microbiome Interactions?
Are Plant-Rhizosphere Linkages Generalizable Across Temporal and Spatial Scales?

#### Abstract text (specific to this investigation):

The productivity of switchgrass, a common biofuel crop, is limited by nitrogen (N) availability in marginal land cropping systems. Free-living diazotrophs, naturally-occurring microorganisms capable of biological N fixation, have the potential to increase switchgrass productivity without fertilizer N inputs by increasing reactive N availability in rhizosphere soils. Classical nitrogenase enzymes require Molybdenum (Mo) to function; this element could therefore limit observed rates of soil biological N fixation in Mo-depleted soils. However, alternative nitrogenase enzymes which operate without Mo, can be expressed by some freeliving diazotrophs as well. In this study we investigated the role that alternative nitrogenases play in biological N fixation in the rhizosphere of switchgrass grown without fertilizer N inputs in marginal land cropping systems. Recent work by Bellenger et al. (2014) has shown that the activity of alternative nitrogenase enzymes can be investigated by measuring the ratio between observed rates of acetylene reduction and <sup>15</sup>N<sub>2</sub> incorporation, two commonly-used assays for nitrogenase activity in soil. We measured both <sup>15</sup>N<sub>2</sub> incorporation and acetylene reduction in soils collected from switchgrass rhizospheres in three marginal land sites in Michigan, which were established as part of the Great Lakes Bioenergy Research Center Marginal Land Experiment. We found evidence for substantial alternative nitrogenase activity in soils collected from a single site (Lux Arbor), but not in soils from other sites tested. Ongoing research is investigating the edaphic factors that favor biological N fixation by alternative nitrogenases in the switchgrass rhizosphere. Furthermore, we are investigating the abundance of free-living diazotrophs and the genetic capacity of these organisms to express alternative nitrogenase enzymes across sites using quantitative PCR.

#### **References:**

Bellenger JP, Xu Y, Zhang X, Morel FMM, Kraepiel AML. (2014). Possible contribution of alternative nitrogenases to nitrogen fixation by asymbiotic N<sub>2</sub>-fixing bacteria in soils. *Soil Biology & Biochemistry* **69:** 413-420.

#### Funding Acknowledgments:

This Project was funded by a grant from the US Department of Energy (DE-SC0014108).

### Physiological and Molecular-Genetic Characterization of Basal Resistance in Sorghum

Jennifer Kimball<sup>1</sup>, Thalita Tuleski<sup>2,3</sup>, Xinye Zhang<sup>1</sup>, Yaya Cui<sup>2</sup>, Catherine Espinoza<sup>2</sup>, Dongqin Chen<sup>2</sup>, Gary Stacey<sup>2</sup> and **Peter Balint-Kurti**<sup>\*1,4</sup>

<sup>1</sup>North Carolina State University, Raleigh NC,; <sup>2</sup>University of Missouri, Columbia, MO <sup>3</sup>Federal University of Paraná, UFPR, Brazil; <sup>4</sup>US Dept. of Agriculture, Agricultural Research Service, Raleigh, NC, \*Presenting author

https://sites.google.com/a/ncsu.edu/maize-disease/home

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

The objectives of the project are the following

• I. Develop robust assays to measure the microbe-associated molecular pattern (MAMP) response and disease resistance in sorghum.

• II. Screen a set of diverse sorghum germplasm for variation in the MAMP response and disease resistance.

• III. Identify genes differentially regulated during the MAMP response in high- and low-responding sorghum genotypes.

• IV. Assess the effect of MTI on disease progression in sorghum

• V. Identify loci associated with variation in disease resistance and the MAMP response. Examine

possible correlations between variation in the MAMP response and in disease resistance

#### **PROJECT SUMMARY**

Plants recognize certain conserved microbial molecules (microbe-associated molecular patterns or MAMPs) and mount a basal defense response called MAMP-triggered immunity (MTI) that limits subsequent colonization. In many cases, the basal defense response is believed to be responsible for non-host resistance: the phenomenon whereby most plants are resistant to most microbial pathogens. Furthermore, there is some evidence that the MAMP response may be involved with quantitative disease resistance, resistance which although partial, tends to be extremely durable. While much is known about the MAMP response in model species, this is not the case for crop plants. Furthermore, naturally-occurring variation in the MAMP response within a species and its relationship to quantitative disease resistance is not well understood. This project builds on our work investigating the genetics controlling the Arabidopsis and soybean MAMP response and on characterizing maize quantitative disease resistance and seeks to:

- Develop robust assays to measure the MAMP response and disease resistance in sorghum.
- Screen a set of diverse sorghum germplasm for variation in the MAMP response and disease resistance.
- Identify genes differentially regulated during the MAMP response in high and low responding sorghum genotypes.
- Assess the effect of MTI on disease progression in sorghum
- Identify loci associated with variation in disease resistance and the MAMP response.

We are making extensive use of available resources for sorghum genomics and quantitative genetics; such as, the genome sequence and various sorghum mapping populations. Over the first year of the project we have made significant progress:

- We obtained four mapping populations and associated genotypic data from collaborators, increased seed and performed a disease trial using the pathogen *Bipolaris sorghicola*, causal agent of target leaf spot.
- We developed a protocol for assessing the response of sorghum to MAMPs by detecting the production of reactive oxygen species (ROS).
- We assessed a set of parental lines for the MAMP response, and for resistance to one bacterial pathogen (*Herbaspirillum rubrisubalbicans*, causal agent of mottled stripe disease) and to three fungal pathogens (*Bipolaris sorghicola*, *Exserohilum turc*icum and *Colletotrichum sublineolum*-causal agents, respectively, of target leaf spot, northern leaf blight and sorghum anthracnose). We identified significant line-to-line variation for each of these traits
- We identified quantitative trait loci (QTL) on chromosomes 4, 6 and 7 associated with variation in response to the MAMP flg22 (an epitope of flagellin) and QTL on chromosomes 6 and 7 for variation in the response to the MAMP chitin. The QTLs on chromosome 6 for the flg22 and chitin responses are at the same position in the genome.
- Using data from a single environment, we tentatively identified one QTL on chromosome 7 associated with variation in target leaf spot resistance. This QTL overlaps with a QTL associated with the flg22 response.
- We identified QTL on chromosomes 7 and 10 associated with resistance to *H. rubrisubalbicans*.
- We have shown that eliciting the MAMP response increases resistance to *H. rubrisubalbicans* in sorghum.

We are currently working on several sub-projects:

- We are assessing an 800-line sorghum association population for response to flg22 and for resistance to Target leaf spot.
- We have cloned the closest homolog in sorghum to OsFls2, the rice flg22 receptor. We are assessing its function to determine whether it is a Flg22 receptor as well
- We have made proteinaceous and non-proteinaceous extracts from *B. sorghicola* and we are assessing their effectiveness as elicitors of the MAMP response.
- We are characterizing the MAMP response at the transcriptional level. We are comparing the response elicited in several different lines and by several different MAMPs.

Under separate funding we have undertaken a similar project in maize. We anticipate that the results of our sorghum and maize projects will mutually inform the other. We will present the latest data from both these projects in our poster at the PI meeting

#### Funding statement.

This work is funded by DOE grant #DE-SC0014116 "Physiological and Molecular-Genetic Characterization of Basal Resistance in Sorghum"

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

**Thomas E. Juenger**,<sup>1,\*</sup> (tjuenger@austin.utexas.edu) Katherine Behrman,<sup>1</sup> Jeremy Schmutz,<sup>2</sup> Tanja Wolke,<sup>3</sup> Felix Fritschi,<sup>4</sup> Alina Zare,<sup>4</sup> Laura Bartley,<sup>5</sup> Julie Jastrow,<sup>6</sup> Roser Matamala,<sup>6</sup> Sean Watson,<sup>7</sup> Denise Costich<sup>8</sup>

<sup>1</sup>University of Texas, Austin;<sup>2</sup> HudsonAlpha Institute for Biotechnology, Huntsville, Alabama; <sup>3</sup>DOE Joint Genome Institute, Walnut Creek, California; <sup>4</sup>University of Missouri, Columbia; <sup>5</sup>University of Oklahoma, Norman; <sup>6</sup>Argonne National Lab, Argonne, Illinois; <sup>7</sup>Ladybird Johnson Wildflower Center, Austin, Texas; <sup>8</sup>International Maize and Wheat Improvement Center

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

Project Goals: Our collaborative project is focused on understanding switchgrass genetic diversity and adaptation across continental scale environmental gradients. Our goal is to improve the sustainability of switchgrass as a biofuel by gaining insight into the interaction of switchgrass diversity with its associated microbiome and environmental conditions. Our approach will involve 1) the collection and characterization of new switchgrass germplasm from across the species range, 2) the development of a genetic association mapping panel and extensive common gardens to evaluate switchgrass performance, 3) a detailed characterization of the switchgrass microbiome, 4) studies of the impact of switchgrass stands on ecosystem processes, and 5) extensive multiscale modeling to define conditions of sustainability and identify key tradeoffs between genetic diversity, productivity, and ecosystem services. Ultimately, these studies will identify critical plant-microbe-soil traits that may be manipulated, through breeding or agronomic management, to improve the sustainability of biofuel feedstocks.

Less carbon-intensive energy sources are needed to reduce greenhouse gas emissions and their predicted role in climate change. There is growing interest in the potential of biofuels for meeting this need. A critical question is whether large-scale biofuel production can be sustainable over the time scales needed to mitigate our carbon debt from fossil fuel consumption. The carbon balance and ultimately the sustainability of biofuel feedstock production is the result of complex climate-coupled interactions between carbon fixation, sequestration, and release through combustion. The long-term productivity of biofuels depends on the genetic and environmental factors limiting plant growth. These factors are often related to soil resources which involve complex interactions at the plant-microbe-soil interface impacting their availability and cycling.

Our collaborative project will address sustainable switchgrass (*Panicum virgatum*) production through a detailed characterization of plant growth and performance in both individual spaced and stand plantings. The project represents an unprecedented field-based experimental system for a bioenergy grass. We bring together diverse skill sets from plant and microbial genetics and genomics, physiology and ecosystem modeling. An underlying theme of the research is the use of locally adapted plant material to explore plant function, to understand the mechanistic basis of
environmental interactions, and to discover the plant genes important for adaptation and sustainability in the face of climate change. To this end, we have been collecting new genetic diversity in switchgrass from natural populations across the species range in North America. These new genotypes are being characterized through genome resequencing and clonally propagated for inclusion in experimental gardens at 14 field locations. This material will provide a detailed population genomic characterization of switchgrass along with resources for association mapping and genomic selection for future breeding programs.

Our plant-microbiome project will fully characterize the microbial communities associated with switchgrass at our planting locations using genomic tools. Specifically, we will sample field plantings of switchgrass for leaf and root microbial communities with 16S iTAG and metegenomic sequencing in collaboration with the DOE Joint Genome Center. Analyses will center on quantifying the relative importance of switchgrass host genotype, the planting environment, or their interaction on microbial community composition. Ultimately, these data will be linked with plant trait information to evaluate drivers of plant-microbiome interactions and their impacts on ecosystem processes.

Our ecosystem processes research will focus on carbon cycle responses at the ecosystem level using stand plantings of switchgrass diversity. We will couple plant growth and physiology measurements with measurements of key carbon pools and fluxes that integrate carbon dynamics across different temporal and spatial scales. Finally, our modeling will define conditions of a sustainable biofuel system and identify key tradeoffs between genetic diversity, productivity, and ecosystem services. Here, we highlight the background resources for our project, our progress establishing our new experiments, results from 2016 experiments, and plans for the coming field season.

This research is supported by funding through the Office of Biological & Environmental Research within the Department of Energy Office of Science and through the DOE Joint Genome Institute Community Science Program.

# Genomic dissection of anthracnose resistant response in sorghum [Sorghum bicolor (L.)]

**Hugo E. Cuevas**<sup>1, \*</sup>(<u>hugo.cuevas@ars.usda.gov</u>), Louis K. Prom<sup>2</sup>, Joseph E. Knoll<sup>3</sup>, Wilfred Vermerris<sup>4</sup>

<sup>1</sup>USDA-ARS Tropical Agriculture Research Station, Mayaguez, PR; <sup>2</sup>USDA-ARS Crop Germplasm Research Unit, College Station, TX; <sup>3</sup>USDA-ARS Crop Genetics and Breeding Research Unit, Tifton, GA; <sup>4</sup>University of Florida, Department of Microbiology & Cell Science, Gainesville, FL

# **Project Goals:**

The goal of this project is to use a genomics-based approach to identify anthracnose resistance loci from diverse sorghum germplasm, to establish against which pathotypes these loci protect, and to determine the disease resistance mechanism of at least one of these genes. This information will provide plant breeders a tool kit that can be used to maximize levels of resistance in different areas of production.

We developed three sets of recombinant inbred lines (RILs) derived from the cross of three resistant sources (SC112-14, QL3 and IS18760) to a commonly high susceptible line PI609251. A high density linkage map was constructed for the RIL population derived from the cross of SC112-14 x PI609251 using 1,671 single nucleotide polymorphism sites (SNPs). We later used this map to select a reduced sample of individuals with a high number of recombinant breakpoints which were evaluated under greenhouse conditions against 10 pathotypes from Texas, Georgia, Arkansas, and Puerto Rico. In parallel, the three RILs were evaluated under field conditions at Gainesville, FL, Tifton, GA, and College Station, TX. In addition, we identified two QTL associated with the resistant response in sorghum line Bk7 (chromosome 7 and 9; Felderhoff *et al.* 2016,G3), which are now being analyzed in more detail to identify the underlying resistance genes.

Anthracnose resistance responses were observed for the RILs derived from SC112-14 at Gainesville, FL, Tifton, GA, and College Station, TX. However, the RILs derived from QL3 and IS18760 showed a resistance response only against pathotypes from College Station, TX. The linkage mapping analysis of SC112-14 identified one locus at distal region of chromosome 5 which is controlling the resistance response against pathotypes from the three locations (Fig. 1A). Remarkably, some RILs showed variable resistance genes. Greenhouse evaluation against multiple particular pathotypes validated this locus and delimited its location to a 250 Kb genomic region (Fig. 1B). The observed difference among greenhouse and field evaluations indicated that both analysis are complementary and necessary to dissect the anthracnose resistant locus. We also determined that resistant locus of in chromosome 9 present in sorghum line Bk7 enclosed 12 candidate genes that were validated to be expressed in leaves.

We will be evaluating a larger number of segregating progenies to reduce this genomic region to candidate genes for functional analysis. Simultaneously, we are constructing high density genetic maps for RILs derived from QL3 and IS18760 in order to map and study their resistance. The 12 candidate genes from Bk7 inheritance study are now being targeted for down-regulation using virus-induced gene silencing.

### Funding statement.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0014171.

**Figure 1**. Anthracnose resistant locus present in line SC112-14. **A)** Linkage map analysis based on field evaluations at Florida, Georgia and Texas; **B)** Linkage map analysis based on greenhouse evaluation against particular pathotypes from Texas, Georgia, Arkansas and Puerto Rico.



# Systems analysis in *Cellvibrio japonicus* resolves predicted redundancy of β-glucosidases and determines essential physiological functions

Cassandra E. Nelson,<sup>1</sup>\* (cassien@unbc.edu), Artur Rogowski,<sup>2</sup> Carl Morland,<sup>2</sup> Joshua A. Wilhide,<sup>3</sup> Harry J. Gilbert,<sup>2</sup> and **Jeffrey G. Gardner**<sup>1</sup>

<sup>1</sup>University of Maryland - Baltimore County, Baltimore, MD; <sup>2</sup>Newcastle University, Newcastle upon Tyne, UK; and <sup>3</sup>Molecular Characterization and Analysis Complex, UMBC, Baltimore, MD

# Project Goals: This project aims to facilitate the establishment a fundamental systems-level model of lignocellulose deconstruction by saprophytic soil bacteria. Over the course of this project we will identify and characterize novel enzymes that have the potential to accelerate the advancement of renewable fuel and chemical technologies.

Degradation of polysaccharides forms an essential arc in the carbon cycle, provides a percentage of our daily caloric intake, and is a major driver in the renewable chemical industry. Microorganisms proficient at degrading insoluble polysaccharides possess large numbers of carbohydrate active enzymes, many of which have been categorized as functionally redundant. Here we present data that suggests that carbohydrate active enzymes that have overlapping enzymatic activities can have unique, non-overlapping biological functions in the cell. Our comprehensive study to understand cellodextrin utilization in the soil saprophyte Cellvibrio *japonicus* found that only one of four predicted  $\beta$ -glucosidases is required in a physiological context. Gene deletion analysis indicated that only the *cel3B* gene product is essential for efficient cellodextrin utilization in C. japonicus and is constitutively expressed at high levels. Interestingly, expression of individual β-glucosidases in *Escherichia coli* K-12 enabled this noncellulolytic bacterium to be fully capable of using cellobiose as a sole carbon source. Furthermore, enzyme kinetic studies indicated that the Cel3A enzyme is significantly more active than the Cel3B enzyme on the oligosaccharides but not disaccharides. Our approach for parsing related carbohydrate active enzymes to determine actual physiological roles in the cell can be applied to other polysaccharide-degradation systems.

# **Publications**

- 1. Nelson C.E. and Gardner J.G. 2015. In-frame deletions allow functional characterization of complex cellulose degradation phenotypes in *Cellvibrio japonicus*. *Applied and Environmental Microbiology*. 81(17): 5968-75.
- 2. Gardner JG. 2016. Polysaccharide degradation systems of the saprophytic bacterium *Cellvibrio japonicus. World Journal of Microbiology and Biotechnology*. 32(7):121-32.
- 3. Nelson CE, Beri NR, and Gardner JG. 2016. Custom fabrication of biomass containment devices using 3-D printing enables bacterial growth analyses with complex insoluble substrates. *Journal of Microbiological Methods*. 130:136-43.

This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0014183.

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

Kranthi Varala<sup>1</sup>\* (kranthi.varala@nyu.edu), Rodrigo Gutiérrez,<sup>2</sup> Dennis Stevenson,<sup>3</sup> Rob DeSalle,<sup>4</sup> W. Richard McCombie,<sup>5</sup> Jean-Michel Ané,<sup>6</sup> Heidi Kaeppler<sup>6</sup> and **Gloria Coruzzi**<sup>1</sup>

<sup>1</sup>New York University, New York; <sup>2</sup>Pontifica Universidad Católica de Chile, Santiago, Chile; <sup>3</sup>New York Botanical Garden, New York; <sup>4</sup>American Museum of Natural History, New York; <sup>5</sup>Cold Spring Harbor Laboratory, New York and <sup>6</sup>University of Wisconsin, Madison

# http://evonet.org

Project Goals: This DOE BER sustainability project aims to identify the key genes and gene regulatory networks that enable "extreme survivor" plants to adapt and grow in marginal, extremely nitrogen (N) -poor soils in the arid Chilean Andes. These extreme survivor species cover the main branches in flowering plants and include 7 grass species of particular interest for biofuels. We focus on 24 "extreme survivor" Chilean species that are relevant to biofuels and compare their genomes to 24 Californian "sister" species that live in a N replete arid environment. Exploiting a novel phylogenomic pipeline we have developed, we use a "paired species" sampling strategy to identify the genes that support the evolutionary divergence of the extreme survivors in the deserts of the Chilean Andes from their sister species in California. The genes thus identified will help to discover the mechanisms underlying physiological and developmental processes that allow efficient assimilation of nitrogen in nitrogen-poor, dry soils. We will test selected genes in the model grass Brachypodium for their ability to confer enhanced growth on low N and/or dry environments. The genes and network modules so uncovered can potentially be translated to biofuel crops to greatly increase biomass and nitrogen use efficiency in marginal, low-fertility soils.

This collaborative project exploits the genomes of "extreme survivor" plants adapted to thrive in marginal, extremely Nitrogen (N) poor soils in the arid Chilean Andes. It uses a previously validated phylogenomic pipeline we developed called BigPlant [1], and a "paired species" sampling strategy, to identify the genes that distinguish these "extreme survivors" in Chile from their related species adapted to similarly dry regions in California (CA) that are not constrained by N. These "extreme survivor" species and paired species samplings represent multiple independent origins of the traits (adaptation to low N and drought), and therefore offer diverse genomic backgrounds within which the survival traits repeatedly arose. We apply our *BigPlant* phylogenomic approach [1,2] to the study of "marginal survivor" strategies as follows:

*Aim 1.* Sample transcriptomes of 24 Chilean "extreme survivors" & paired CA species (NYU, NYBG, Chile). *Progress:* Completed RNA-seq analysis of all 24 Chilean species and collected 11/24 sister species from CA & their RNA-Seq library preparations are in progress. *Aim 2.* Phylogenomic analysis of 48 "paired species" to identify genes that repeatedly support nodes that distinguish the extreme survivors in Chile from their sister species in CA (AMNH, NYU). *Progress:* We enhanced the *BigPlant* phylogenomic pipeline to build phylogeny from 32

paired species with ~20-30K orthologs per species. This phylogenomic analysis of 70+ species translates to ~1.7 million orthologs and ~2 million informative characters, building the largest phylogenomic tree ever constructed to date. This phylogenomic analysis identified candidate genes that provide support for the evolution of low-N adaptation in the Chilean species.

\*NEW\* We compared 32 Chilean species we sequenced to their closest related species whose transcriptome is already available. In this analysis (*EvoNet v1*), we captured 19 evolutionarily independent origins of low-N/Drought adaptation (see Figure 1). Our *BigPlant* phylogenomic



Figure 1. EvoNet v1: Extreme survivor species in Chile (Red) and paired "sister" species (black). Preliminary phylogenetic analysis of 32 pairs of "extreme survivors" to available sister species identified 1,157 genes that provide recurrent (>3 independent origins) support for the appearance of low-N/Drought adaptations. Our EvoNet analysis captured 19 independent origins of low-N adaptation.

# provide recurrent support (> 3 independent origins) for the divergence of low-N adapted Chilean species from their available sister species. This set of genes is enriched in the processes Nitrogen compound metabolism and Photosynthesis (GO term analysis FDR <0.01). Crucially, we identified 19 transcription factors that provide recurring support for low-N adaptation and are functionally validating them through transgenic lines in the model monocot *Brachypodium*.

analysis pipeline [1] identified 1,157 genes that

*Aim 3.* Combine phylogenomics (protein sequence) and gene networks (gene expression) to identify genes and network modules associated with adaptations to marginal, low-N soils (NYU, Chile). *Progress:* We exploit a comparative analysis of gene regulatory networks by developing a new module *PhyloExpress* that extends the *BigPlant* phylogenomic pipeline to include gene expression data.

*Aim 4.* Functionally validate top-ranked candidate master genes for low-N adaptation in Arabidopsis and Brachypodium (NYU, Chile, U Wisconsin). *Progress:* We (Wisconsin) are transforming Brachypodium with the most promising candidate from our *EvoNet v1* analysis (Fig 1).

# References

1. Lee E et. al., A functional phylogenomics view of the seed plants. PLoS Genet 7(12):e1002411.

2. Delaux et. Al., Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. PloS Genet 10(7):e1004487.

The EvoNet grant is supported by the Genomics Science program within the Office of Biological and Environment Research in the DOE Office of Science.

# Does mycorrhizal symbiosis determine the climate niche for *Populus* as a bioenergy feedstock?

Kabir Peay,<sup>1\*</sup> (kpeay@stanford.edu)

<sup>1</sup>Stanford University, Stanford, CA

http://stanford.edu/~kpeay/

**Project Goals:** This research project uses a native North American tree with strong biofuel potential – *Populus* – to answer fundamental questions about the role of climate, soil environment, and mycorrhizal interactions in determining growth and competition in plant communities. The experiments we are carrying out will provide important basic insights into the way beneficial interactions shape the natural world, but will also have a direct impact on predicting the suitability of particular sites for bioenergy projects in light of climate change, and predicting how above- and below-ground carbon allocation might change with climate.

While microbes have long been viewed as agents of disease, recent explorations of the microbiome have led biologists to recognize that beneficial microbes play an equally vital role in maintaining the health of plants and animals. Perhaps the most ubiquitous form of beneficial interaction in terrestrial ecosystems occurs between fungi and plant roots. These fungus-root, or "mycorrhizal", symbioses involve the reciprocal exchange of plant sugars for soil nutrients obtained by the fungus, such as nitrogen and phosphorous, which are critical for plant growth. It is known that mycorrhizal associations are widespread, occurring in over 90% of plants, and diverse, with single plants associating with over 100 species of fungi. Despite this, the ecological factors that control the distribution and abundance of mycorrhizal symbioses are still poorly known, making it challenging to predict how mycorrhizal symbioses may change in future climate conditions or how these fungal communities might be manipulated to improve agriculture or forestry.

This research project uses a native North American tree with strong biofuel potential – *Populus* – to answer fundamental questions about the role of climate, soil environment, and mycorrhizal interactions in determining growth and competition in plant communities. *Populus* provides a unique opportunity for ecologically relevant experiments because of its widespread distribution across North America and its natural variability in mycorrhizal association types. The first component of this project will use a global forest database to map and model the distribution of different mycorrhizal associations with respect to climate. After establishing this baseline laboratory experiments will then be used to measure the precise ways in which beneficial interactions with mycorrhizal fungi shape the environmental niche dimensions of *Populus*, and how variation in mycorrhizal associations determine the balance of competition with co-occurring tree species across its natural range. Finally, we will use we will use isotope labeling to track changes in carbon flow from *Populus* to mycorrhizal fungi and other soil microbes in different environmental conditions. These experiments will provide important basic insights into the way beneficial interactions shape the natural world, but will also have a direct impact on predicting the suitability of particular sites for bioenergy projects in light of climate change, and predicting how above- and below-ground carbon allocation might change with climate.

This project is funded by an Early Career Award from the Office of Biological and Environmental Research in the DOE Office of Science.

### Genetics and Genomics of Pathogen Resistance in Switchgrass

**Serge Edmé**<sup>1\*</sup> (serge.edme@ars.usda.gov), Gautam Sarath<sup>1</sup>, Nathan Palmer<sup>1</sup>, Rob Mitchell<sup>1</sup>, Satyanarayana Tatineni<sup>1</sup>, and Gary Yuen<sup>2</sup>

# <sup>1</sup>USDA-ARS, Lincoln, NE; <sup>2</sup>University of Nebraska, Lincoln

Project Goals: This project wants to establish switchgrass as a key component of the bioenergy industry in the US by developing cultivars with the right suite of genes for high biomass, ethanol yield, and with good quality profiles and durable resistance to diseases. The specific objectives are:

(1) Understand the genetic and genomic bases of pathogen response in regionally-adapted upland and lowland switchgrass breeding populations with contrasting disease symptoms under field conditions,

(2) Dissect the molecular underpinnings of the broad resistance or tolerance to pathogens in 'Kanlow' vs 'Summer', and

(3) Discover the molecular differences that permit systemic viral infections in some switchgrass plants, but not in other genetically-related plants.

Switchgrass (*Panicum virgatum*) is susceptible to a number of fungal (*Puccinia emaculata* and *Uromyces graminicola*) and viral (*Panicum mosaic virus*) pathogens, making yield stability and biofuel output under biotic stress important selection criteria for biofuel feedstock improvement. This project leverages the differential reactions to challenges from the rust and viral pathogens of a composite tetraploid switchgrass population derived from crossing Kanlow (lowland) with Summer (upland) ecotypes, a population already under improvement for bioenergy-related traits. Three successive generations (parents and offspring) will be measured phenotypically for quantitative traits (biomass yield, rust, and cell wall components) and genomically with molecular SNP (single-nucleotide polymorphisms) markers derived from DArT-seq (see Figure). Prediction models will be developed from these assessments based on quantitative (BLUP, animal model, breeding values, and index selection) and molecular genetics (QTL mapping, genomic selection) methodologies and combined to deploy the best aggregate "phenotype-genotype".

The USDA-ARS and UNL labs in Lincoln, NE, are highly experienced with breeding perennial grasses like switchgrass since the 1930s and have developed and published robust NIRS (near-infrared spectroscopy) calibration equations for quality (SCW) traits. Breeding methodologies are being continually refined and optimized to maximize gains from selection. Selection is practiced in two stages, first with selection of the best 25-30% of tested families from which the top 10% of individuals are chosen as parents of the next generation. Recombinations are carried out by open-pollination which will be augmented with poly- and biparental crosses to capitalize on all types of gene actions (ie additive and dominance, more specifically) and maximize genetic gains. A pedigree will be established to follow the flow of genes from parents to progeny across generations and to increase the efficiency of prediction models by including the pedigree and molecular relationship matrices. The two rust pathogens are prominent in field trials established at the UNL Agricultural Research and Development Center (located near Mead, NE) and intense

natural disease pressures prevail yearly to allow discrimination of tolerant or resistant from susceptible genotypes and segregation of phenotypes/genotypes with all combinations of high/low yield and high/low lignin. Ratings will be carried out on a 1-5 scale and taken twice during the crop cycle. This discrimination will facilitate QTL mapping in a derived population by crossing susceptible with tolerant parents in order to define the regions governing rust, biomass yield, and SCW quality traits. QTL mapping combined with genomic selection will ensure identification of genes with major and minor effects and establish the foundation for better refinement of those genes.

The functional transcriptional networks, underpinning the broad rust and viral disease resistance or tolerance in the Kanlow population, will be studied. RNAseq data will be collected to interrogate the transcriptional changes underlying defense responses in switchgrass, based on the overall and specific differences in gene expression in Kanlow and Summer. These studies will be augmented by engineering viral genomes with fluorescent markers to visualize viral movement and pathogenesis in plants, developing antibodies to coat proteins to detect virus in field-grown samples, and by using recombinantly-expressed viral proteins to identify interacting plant proteins that are crucial for the spread of infection.



This research project is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0016108.

# Quantitative stable isotope probing with <sup>15</sup>N in soil microbial communities

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

<sup>1</sup> Division of Plant and Soil Sciences, West Virginia University, Morgantown, West Virginia

<sup>2</sup> Center of Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona

<sup>3</sup> Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona

<sup>4</sup> Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California

<sup>5</sup> Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington

**Project Goals:** Connecting the composition of microbial communities with biogeochemical process rates has the potential to improve our understanding of, and ability to model, ecosystem function. This project aimed to extend the promising new technique of quantitative stable isotope probing (qSIP) to the study of nitrogen cycling in soil. This work dovetails with our our larger goal to characterize *in situ* rates of biogeochemically significant microbial activity at both the community scale and for specific taxa. This information will be leveraged to establish whether there is a "phylogenetic imprint" on soil carbon and nitrogen cycling processes that can facilitate better incorporation of microbial data into process-scale modeling efforts.

Anthropogenic disruption of the nitrogen (N) cycle is having cascading effects on ecosystems around the global as the cycling of carbon (C) and N are closely linked. In addition to regulating primary production, N availability interacts with other variables to determine rates of soil organic matter decomposition. However, the contributions of microbial taxa to N cycling processes such as immobilization are not well understood. Here we used qSIP with <sup>15</sup>N to measure taxon-specific nitrogen uptake by microorganisms in soil. To assess the effects of C availability of N assimilation we considered control and glucose amended soils. Our results indicate that qSIP is sufficiently sensitive to quantitatively differentiate N assimilation among prokaryotic taxa. At the community level, <sup>15</sup>N assimilation was enhanced by glucose addition, a pattern mirrored by traditional measurements of nitrogen immobilization. The qSIP analysis revealed strong phylogenetic organization in N assimilation with broad phylogenetic groups exhibiting distinct patters of N uptake. For instance, most phylotypes within Acidobacteria (74%), Actinobacteria (93%), Verrucomicrobia (83%), and Proteobacteria (82%) assimilated more N in the presence of glucose. Conversely all taxa within Firmicutes and Crenarchaeota as well as the majority of Bacteriodetes (84%) had greater N assimilation in the absence of added carbon. These patterns suggest that phylogenetic groups of prokaryotes have distinct and coherent patterns of N uptake that reflect their ecological strategies. The distinct activities of phylogenetic groups provide a basis for understanding how phylogenetic microbial community composition influences N immobilization in soil.

*This research was supported by the Office of Biological and Environmental Research in the DOE Office of Science* 

### **Studying Microbial Stress Responses in Soil Ecosystems**

Paul Dijkstra<sup>\*1,2</sup> (paul.dijkstra@nau.edu), Benjamin J. Koch<sup>1</sup>, Jennifer Pett-Ridge<sup>3</sup>, Steve Blazewicz<sup>3</sup>, Xavier Mayali<sup>3</sup>, Kirsten Hofmockel<sup>4</sup>, Egbert Schwartz<sup>1,2</sup>, Ember M. Morrissey<sup>5</sup>, Galya Orr<sup>4</sup>, Petr Baldrian<sup>6</sup>, **Bruce A. Hungate**<sup>1,2</sup>.

<sup>1</sup> Center of Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona

<sup>2</sup> Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona

<sup>3</sup> Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California

<sup>4</sup> Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington

<sup>5</sup> Division of Plant and Soil Sciences, West Virginia University, Morgan Town, West Virginia <sup>6</sup> Institute of Microbiology of the CAS, Prague, Czech Republic

*Project goal*: Using metatranscriptomics, characterize the community response to temperature and resource limitation. This is part of the broader project goal of studying how growth, death, and C use efficiency of individual microbial taxa and whole communities is affected by temperature and changes in substrate availability and regulates whole community and ecosystem responses to changes in temperature.

*Introduction:* The overall project goal is to relate changes of growth, death and C use efficiency of individual microbial taxa in response to temperature to whole community and ecosystem responses. As part of this overall goal, we will study the response of gene-expression to temperature stress and resource limitation.

*In vitro* pure culture studies have demonstrated that, when substrate availability falls below or temperature rises above a critical level, microbial cells respond with alterations in gene-expression. These responses have been observed for different taxa, and the responsible genes, sigma factors, transcription factors and regulators, are distributed widely across the bacterial domain. We propose to study the effects of temperature stress and resource limitation in real, complex and diverse soil ecosystems.

*Results*: Preliminary work was done comparing microbial communities in a litter layer and mineral soil, a marsh sediment, and a marine community (Wemheuer et al 2015). Our analysis focused on sigma factors, transcription factors and regulators that control gene-expression.



Our findings show that

Fig. 1. Relative transcript abundance of sigma factors and main transcription factor and regulator families for a soil (Paul Dijkstra and Petr Baldrian, unpublished results), marsh (Paul Dijkstra and Galya Orr, unpublished results), and marine ecosystem (Wemheuer et al., 2015). Note the low abundance of transcripts for  $\sigma^{S}$  in soil, while no  $\sigma^{S}$  transcripts were observed in marsh and marine ecosystems.

Fig. 2. Transcription factors and regulators in mineral soil and litter layer of a Bohemian Norway spruce forest (Žifčáková et al., 2016) for Acidobacteria and alpha-Proteobacteria (Paul Dijkstra and Petr Baldrian, unpublished results). A shared set of transcription factors and regulators were sorted from high to low abundance in Acidobacteria (soil).

1- Many of the known sigma factors, transcription factors and regulators, responsible for active growth and responses to stress are observed in metatranscriptomes (Fig. 1). For example, sigma D ( $\sigma^{D}$ ), the sigma factor responsible for gene-expression during active <sup>fur</sup> growth, is found in all environmental samples, and individual taxa within each community (Fig. 1, 2).



2- The relative abundance of sigma factors, transcription factors and regulators differs between a mineral soil, marsh sediment and marine ecosystem (Fig. 1). For example, sigma H ( $\sigma^{H}$ ), the sigma factor controlling gene-expression under heat stress, was much higher in the marine than in soil or marsh ecosystems. Similar differences are observed for transcription factors and regulators (Fig. 1B). The low abundance of sigma S ( $\sigma^{S}$ ), the general stress sigma factor and involved in the starvation response, contradicts the paradigm that microbes in soil are C-starved.

3- Within the same community, individual microbial taxa can differ in the relative abundance of sigma factors, transcription factors and regulators. For example, Acidobacteria in mineral soil and marsh sediment have high levels of sigma E ( $\sigma^{E}$ ), an indication of extracellular or envelop stress. At the same time, different microbial taxa respond similarly to (long-term) changes in their environment. For example, Acidobacteria and alpha-proteobacteria decrease the relative abundance of NifA, but increase the abundance of CarD and CspA.

### Conclusion

Many of the regulators of gene-expression can be recognized in metatranscriptomes and are a rich source of information to understand responses of individual taxa and entire communities to environmental change. Research thus far has been focused on comparing ecosystems which may be characterized by parallel changes in gene abundance and geneexpression. Changes in gene-expression, as a response to altered environment, are likely most important over the short term, while changes in gene abundance is more important in the longterm. This project aims to determine the dynamics of gene-abundance versus gene-expression, and its effect on growth and death and C use efficiency of individual microbial taxa and entire community activities. The data will also be used to study the regulons, that is the genes that are controlled by the sigma factors, transcription factors and regulators, for entire communities and individual taxa. Finally, this dataset will be used to describe the ecological responses to temperature and C availability. For example, we will study the transcript abundances of ecologically relevant processes such as motility, chemotaxis, dormancy and microbe-microbe interactions.

### **Communal Metabolism of Methane and the Rare Earth Element Switch**

Ludmila Chistoserdova<sup>1</sup>\*, Sascha Krause<sup>1</sup>, Zheng Yu<sup>1</sup>, David A.C. Beck<sup>1,3</sup>, and Mary E. Lidstrom<sup>1,2</sup>

Departments of <sup>1</sup>Chemical Engineering and <sup>2</sup>Microbiology and <sup>3</sup>eScience Institute, University of Washington, Seattle, USA; <u>milachis@uw.edu</u>

Project Goals: This project addresses the structure and function of microbial communities active in methane consumption, using lake sediment as a model. Through manipulations of synthetic communities and systems biology approaches, we are striving to understand the molecular mechanisms that form a basis for specific interspecies interactions in microbial oxidation of methane. In this phase of the project our goals are 1) Evaluate behavior of multispecies model synthetic communities comprised of major functional guilds defined in prior research, and assess performance of these communities; 2) Determine the metabolic networks governing microbial consortia, through identification of specific enzymes/ pathways/ factors involved in interspecies interactions; 3) Apply machine learning for predictive modeling of community function.

Metabolism of methane is an important part of biogeochemical cycling of carbon. Methane is also a major contributor to climate change. A specialized group of microbes (the methanotrophs) that consume methane, gaining both energy and carbon from this chemically inert compound, represent a natural filter preventing an even faster accumulation of methane in the atmosphere. While methanotrophy has been studied for the past hundred years as a metabolic feature of individual pure cultures, a concept of communal function in methanotrophy has been gaining momentum. However, the mechanistic details are still missing of how and why the methanotrophs share their hard-earned carbon with other species, and whether and what they gain in return. This current project, initiated in August 2016, builds upon results from prior funding by the DOE (DE-SC-0010556).

We are using Lake Washington sediment community as a model. We manipulated complex natural communities using methane as the sole source of carbon, to determine species persisting in methane-consuming communities (the top-down approach). We also built synthetic communities of pure cultures of methanotrophs and non-methanotrophs and tested their behavior under a variety of environmental conditions (the bottom-up approach). We sequenced multiple (meta)genomes and (meta)transcriptomes to gain insights into the genomic potentials and gene expression patterns in relevant microbes.

Through microcosm manipulation, using methane as the sole source of carbon, followed by metagenomic analysis, we established key species active in methane consumption as the bacteria of the family *Methylococcaceae*. We further established the primary and most abundant satellite types, the non-methanotrophic methylotrophic bacteria of the family *Methylophilaceae*. Two other persistent but less abundant types were identified as members of Burkholderiales and Flavobacteriales. Through manipulation of synthetic communities, followed by transcriptomic analysis, we identified at least one metabolic node at which community cross-talk takes place, the methanol oxidation step that involves alternative methanol dehydrogenase enzymes, one requiring calcium as a cofactor, another requiring rare earth elements (REE), one of the first demonstrations of a biological function for this group of metals. Enzyme choice, in turn, appears to be determined by a number of environmental factors, such as oxygen and methane partial pressures, as well as sources of nitrogen.

We conclude that methanol must be the major carbon compound that the methanotrophs share with other community members, and that carbon flow is regulated by the REE switch, presenting an unexpected and unprecedented example for the important role of REEs in complex biological systems. Overall, our data shed new light on social lives of microbes involved in metabolism of methane in natural habitats and highlight some of the metabolic links among the community partners.

### **Publications**

Chistoserdova L (2017) Systems approaches to studying methylotrophs and methylotroph communities. Curr Iss Mol Biol. In Press.

Krause SM, Johnson T, Samadhi Karunaratne Y, Fu Y, Beck DA, Chistoserdova L, Lidstrom ME (2016) Lanthanide-dependent cross-feeding of methane-derived carbon is linked by microbial community interactions. Proc Natl Acad Sci U S A. Dec 27. pii: 201619871. doi: 10.1073/pnas.1619871114.

Yu Z, Krause SM, Beck DA, Chistoserdova L (2016) A synthetic ecology perspective: how well does behavior of model organisms in the laboratory predict microbial activities in natural habitats? Front Microbiol. 15;7:946. doi: 10.3389/fmicb.2016.00946.

Chistoserdova L (2016) Lanthanides: new life metals? World J Microbiol Biotechnol. 32(8):138. doi: 10.1007/s11274-016-2088-2.

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

# Directing traffic in the rhizosphere: how phage and fauna shape the flow and fate of root carbon through microbial pathways

Katerina Estera<sup>1\*</sup> (rinaest@gmail.com), Anne Kakouridis<sup>1</sup>, **Javier Ceja Navarro**<sup>2</sup>, Henrik Krehenwinkel<sup>1</sup>, Nhu Nguyen<sup>3</sup>, Evan Starr<sup>1</sup>, Steve Blazewicz<sup>7</sup>, **Eoin Brodie**<sup>2</sup>, **Trent Northen**<sup>2</sup>, **Neo Martinez**<sup>4</sup>, **Zhili He**<sup>5</sup>, **Jizhong Zhou**<sup>5</sup>, **Mary Lipton**<sup>6</sup>, **Rosemary Gillespie**<sup>1</sup>, **Jennifer Pett-Ridge**<sup>7</sup>, and **Mary Firestone**<sup>1</sup>

<sup>1</sup>University of California, Berkeley; <sup>2</sup>Lawrence Berkeley National Laboratory, <sup>3</sup>University of Hawai'i, Manoa; <sup>4</sup>University of Arizona, Tucson; <sup>5</sup>University of Oklahoma, Norman; <sup>6</sup>Pacific Northwest National Laboratory; <sup>7</sup>Lawrence Livermore National Laboratory

Project Goals: Our project is designed to explore the complex interactions controlling carbon (C) flow in the rhizosphere -- addressing two overarching topics: 1) How multitrophic interactions control soil C dynamics, and 2) How changing precipitation regimes alter these interactions, and thus impact flow and fate of soil C. We will study interactions among key groups of soil organisms: 1) arbuscular mycorrhizal fungi (AMF), bacteria, archaea, and roots; 2) phage and their microbial hosts; and 3) rhizosphere fauna (protoand metazoan) and their prey. Primary goals of this work are to expand the knowledge of food web dynamics in the rhizosphere, how these multi-trophic interactions play a role in terrestrial C cycling, and to investigate how drought alters these interactions and the fate of soil C. The resulting data and information will substantially expand our knowledge of microbial ecology, food web interactions, and terrestrial C cycling.

Rhizosphere soil immediately surrounding plant roots is a zone of abundant biological activity and an area of intense C cycling. Organisms that reside in the rhizosphere, including bacteria, fungi, viruses, and fauna, actively interact with each other as they utilize, transform, and transfer C from root exudates and root debris. These food web interactions enable C that has been fixed via photosynthesis to return to the atmosphere or remain in soil for varying periods of time. While food web interactions are generally well recognized, little is known about how rhizosphere multi-trophic interactions impact C transformation and persistence. Additionally, it is likely that future alterations in precipitation regimes will differentially impact the participants and interactions in multitrophic, root-C based food webs. Our research investigates two primary questions: 1) How do the complex interactions among bacteria, fungi, phage, and fauna mediate and control C flow and fate in the rhizosphere? 2) How do changing precipitation patterns alter these interactions?

We hypothesize that the major phage and faunal grazers of bacteria and fungi redirect a substantial portion of root-derived C towards mineralization. This means that a high abundance of faunal grazers could lead to more rapid mineralization. We expect drought to reduce the function and abundance of protozoa, nematodes, and phage, but it may have less impact on the abundance and function of arthropod fauna. To test these hypotheses, we have constructed 16 "trenched", rainout plots in a Mediterranean grassland located in the Hopland Research and Extension Center (HREC) in Hopland, California. These plots contain monoculture stands of *Avena barbata*, a naturalized slender oat found throughout California. Precipitation will be manipulated such that half of the plots receive a 50% reduction of the 50-year rainfall average.

The other half of the plots will receive the full average rainfall amount. We will use <sup>13</sup>C labeled carbon dioxide (CO<sub>2</sub>) to trace the pathway of CO<sub>2</sub> as it is fixed by the plant, and delivered belowground by the roots in the form of exudates and fresh root, and consumed by the various residents of the rhizosphere. Soil will be sampled at multiple time points to track the location and persistence of the recently fixed C in the soil. DNA and RNA will be extracted from rhizosphere soil for stable isotope probing (SIP) enabled -omic techniques. Results will help elucidate how C travels through the rhizosphere food web.

Until recently, our analyses of the faunal component of food webs has been limited by laborintensive methods and a lack of molecular tools. Next generation sequencing approaches are just now becoming available to study the community assembly of higher eukaryote taxa. Based on ever-growing reference sequence collections of mitochondrial and nuclear DNA sequences, accurate predictions of species composition and richness are now feasible from mixed environmental samples. To identify components of the soil fauna, we are developing a metabarcoding approach based on mitochondrial *cytochrome oxidase I* (COI) and the nuclear ribosomal *DNA* markers 18S and 28S. We use a two-step PCR approach to simultaneously sequence all markers on by MiSeq. Our initial work on arthropod communities suggests an average recovery of 97% of the species in a mixed sample. We are currently optimizing sequence datasets with multiple gene fragments (e.g. COI) that allows for prediction of species composition and relative abundance in environmental samples. Combining these gene-based approaches with stable isotopes will allow us to follow C from roots to faunal food sources to faunal predators, and ultimately to CO<sub>2</sub> or heavy fractions of soil.

AMF are ubiquitous soil organisms with critical roles in ecosystems, notably as plant root symbionts. The development of molecular methods has made it possible to identify AMF taxa in plant roots and soil instead of relying on morphological characteristics of spores. Illumina sequencing offers new opportunities to investigate the molecular diversity and community ecology of AMF. We use two primer pairs spanning regions of increasing variability in the AMF nuclear ribosomal DNA (rDNA): WANDA/AML2, which targets a DNA sequence in the small subunit (SSU) and gITS7/ITS4, which targets a DNA sequence in the internal transcribed spacers (5.8S, ITS2). As different primer pairs may be biased towards certain groups of AMF, this use of primer pairs with complementary strengths leads to better resolution and inclusion during molecular analyses, as well as more robust information for understanding AMF community ecology.

We hypothesize that a myriad of cross-domain interactions provides pathways and controls for root C flow into the soil. The interactions of these organisms form complex ecological networks. We will be using random matrix theory (RMT) approaches to define the structures and connections of the root-C based food webs.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Numbers DE-SC0010570 and DE-SC0016247 to UC Berkeley, University of Oklahoma, and University of Arizona. Work at LLNL was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344.

### Mapping the pathways of root carbon flow into and through soil microbial food webs

Evan Starr<sup>1</sup>\* (starr.evan@berkeley.edu), Shengjing Shi<sup>1</sup>, Alexander Probst<sup>1</sup>, Donald Herman<sup>1,2</sup>, Steve Blazewicz<sup>3</sup>, Brian Thomas<sup>1</sup>, Jillian Banfield<sup>1,2</sup>, and **Mary Firestone<sup>1,2</sup>** 

<sup>1</sup>University of California, Berkeley, California; <sup>2</sup>Nuclear and Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, California; and <sup>3</sup>Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California

Project goals: The flow of carbon (C) from roots into soil is controlled by a complex array of interactions. Our project addresses how multi-trophic interactions mediate the flow and fate of root C into soil and how changing precipitation regimes alter these interactions. This newly initiated research project uses stable isotope probing (SIP) and genome-resolved metagenomics to identify and characterize the participants in root-C-based food webs and to understand the ecological interactions in the rhizosphere that ultimately control the fate of C entering soil. We are tracking <sup>13</sup>C-labeled C moving from roots into root-exudate and debris consumers and through the members of the soil food web supported by these primary consumers. Genome-resolved metagenomic analyses of the SIP-isolated DNA then allows us to better understand the functional characteristics of rhizosphere C-transformers and illuminate the carbon basis of these interdomain interactions in soil.

The zone of soil influenced by roots, the rhizosphere, is of great importance to plant and soil health and carbon cycling. While there have been numerous explorations of root-bacterial and root-fungal interactions, there has been little work on how complex biotic, multi-domain interactions mediate and control C flow in the rhizosphere. We hypothesize that the complex interactions among bacteria, archaea, fungi, bacteriophages, and fauna are primary controllers of the flow and fate of root C. We propose to identify the players and ultimately assess the quantitative importance of the multiple pathways of C-flow in the rhizosphere using stable isotopes (<sup>13</sup>C) coupled with metagenomic analyses. We expect that metagenomic data will allow us to better understand interdomain interactions as well as bacteriophage-host interactions in soil. For the work reported here, we grew Avena fatua, common wild oat, in <sup>13</sup>CO<sub>2</sub> (99 atom%) and collected starting soil, rhizosphere soil, and bulk soil (soil not associated with roots) at weeks 0, 6, and 9. The DNA extracted from these samples was then prepared using density-gradient centrifugation, yielding DNA samples with a range of <sup>13</sup>C label: unlabeled, partially labeled, and heavily labeled DNA. Samples for each time point were prepared using this procedure and the "purified" DNA was then sequenced (HiSeq 3000); the reads were assembled, and binned to yield genomes.

Using relevant marker genes and whole genomes, we investigated microbes living in the rhizosphere: those that grew and incorporated the <sup>13</sup>C into their DNA and those that did not. The microbes living in rhizosphere that do not incorporate root carbon into their DNA are more like the bulk soil community, suggesting that investigations of rhizosphere ecology that do not employ stable isotope probing may underestimate the degree of difference between the rhizosphere community and the bulk community. We were able to retrieve many genome bins from our soil samples that were at least 70% complete. Using whole genome information, we were able to functionally characterize the rhizosphere microbial community and better

understand the selective pressure that the root exerts. Using hidden Markov models and homology based searches we investigated genes involved in energy metabolism and carbon and nitrogen cycling in our recovered microbial genomes.

In addition to microbial functions, assembled metagenomic data from soil allow us to explore other members of the microbial world. This is especially true for the study of bacteriophages, for which there are no universal primers. Through genome resolved metagenomics, we could assemble and identify bacteriophage genomes; some bacteriophages were identified in the heavily labelled fraction indicating they must have parasitized microbes growing on root derived carbon. Using bioinformatic analyses of CRISPRs as well as tRNAs, and homologous matches, we were able to link a number of bacteriophages to their host bacteria with confidence. This suggests that not only are bacteriophages shaping bacterial communities in the rhizosphere but that bacteriophages may be critical in determining the flow and fate of C entering soil as root exudates and debris. In addition to metagenomic identification of bacteriophages, we were also able to identify labelled bacterivore nematode 18S rRNA genes indicating they are consuming labeled bacterial cells. The quantitative importance of these top-down control processes is a major focus of this ongoing project.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Numbers DE-SC0010570 and DE-SC0016247 to MF at UC Berkeley and DE-SC10010566 to JB at UC Berkeley. ES is supported by an NSF GRFP. Work at LLNL was performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344. Roots stimulate expression of decomposition transcripts in the soil microbiome

Erin Nuccio<sup>1\*</sup> (nuccio1@llnl.gov), Ulas Karaoz,<sup>2</sup> Eoin Brodie,<sup>2,5</sup> Jizhong Zhou,<sup>3</sup> Susannah Tringe,<sup>4</sup> Rex Malmstrom,<sup>4</sup> Tanja Woyke,<sup>4</sup> Mary Firestone,<sup>5</sup> and Jennifer Pett-Ridge<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory, Livermore, USA, <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, USA, <sup>3</sup>University of Oklahoma, Norman, USA, <sup>4</sup>Joint Genome Institute, Walnut Creek, USA, <sup>5</sup>University of California, Berkeley, USA

Project Goals: Our project (*Mapping soil carbon from cradle to grave: drafting a molecular blueprint for C transformation from roots to stabilized soil organic C*) focuses on a fundamental understanding of C cycling in soil as mediated by soil microorganisms and their interactions with plants. We are measuring how organic C decomposition is altered when soil microbial communities interact with living roots, and want to better understand how interactions between soil minerals and microorganisms affect C stabilization processes, particularly in the rhizosphere. Through our research we seek to provide a mechanistic understanding of the conversion of root-derived C to stabilized soil C, clarify the impacts of microbial activities on soil C sequestration, and substantially expand our understanding of molecular regulation of terrestrial C cycling.

The soil surrounding roots, the rhizosphere, has long been recognized as a zone of great functional importance to plants and is a hotspot of belowground carbon cycling in terrestrial systems. Plants transfer atmospheric CO<sub>2</sub> to belowground soil C pools, while microbes are the primary mediators of C transformation and mineralization in soils. The rhizosphere environment alters the microbial breakdown of plant tissues and root litter, and can accelerate the decomposition of detrital plant biomass, a process commonly termed as "priming". However, the molecular mechanisms underlying rhizosphere C cycling are poorly understood, and the hydrolytic and lignolytic proteins mediating the decomposition of root litter in soil are largely unknown

We hypothesized that root exudates stimulate the expression of enzymes that are involved in decomposition of macromolecular C compounds. To assess how enzyme-mediated decomposition differs in the rhizosphere relative to the surrounding bulk soil, we analyzed community gene expression (metatranscriptomes) and single cell genomes of rhizosphere and bulk soil associated with wild oat (Avena fatua) over time (3, 6, 12, and 22 days). To isolate roots of a defined age in a mature plant, we used microcosms with a transparent experimental sidecar to track roots as they grew. Half of the microcosms were amended with dried A. fatua root litter, which was added to the experimental sidecar immediately prior to the start of the experiment. After harvesting rhizosphere and bulk soils, RNA was extracted, and ribosomal RNA was depleted to enrich for mRNA. In total, we sequenced 48 soil metatranscriptomes, which contained approximately 40 million high-quality, paired-end mRNA reads per library. Transcripts were mapped to 96 metagenomic genome bins, 35 single-cell genomes, and 39 isolate genomes; all derived from soil collected at the same location in Hopland, CA, USA. Differential expression analysis showed significant changes in gene expression between rhizosphere and bulk soils at all time points.

Our results indicate that a large number of C decomposition transcripts were more highly expressed in the rhizosphere compared to bulk soil. Of the Carbohydrate Active Enzyme (CAZyme) transcripts that significantly differed between rhizosphere and bulk soil, 96% of were significantly elevated in the rhizosphere. These included transcripts for cellulose and hemicellulose degradation genes, including beta-glucanases, beta-glucosidases, and xylanases. Gene transcripts potentially involved in decomposing microbial necromass were also elevated in the rhizosphere (e.g., chitinases, lytic murein transglycosylases, peptidoglycan/xylan/chitin deacetylases). While we found that many of our reference genomes had the genetic capacity to decompose plant polymers (ca. 33%) and all of these organisms had detectable gene expression, we measured significant transcription of plant decomposition genes in only a small subset of these organisms (ca. 4%). Three genome bins derived from the soil metagenome were particularly active in the rhizosphere compared to bulk soil, and altered their expression of CAZymes depending on the age of the root. Janthinobacterium (Oxalobacteriaceae) had higher expression of cellulases and xylanases near young roots (3-6 days old), while Rhizobacter (Burkholderiales) and an unclassified Streptomycetaceae (Actinobacteria) had higher expression of cellulases and xylanases in more mature roots (12-22 days old).

This work identifies potential molecular mechanisms that underpin the 'priming effect' in rhizosphere soil. We found that transcription of genes involved in decomposition was stimulated in the rhizosphere at all time points, which supports the hypothesis that roots stimulate enzymes for the decomposition of macromolecular carbon in soil. The expression of decomposition genes was dynamic and changed as the root grew, indicating that decomposition is undertaken by a series of different organisms as the root grows. However, while many rhizosphere taxa appear to have the capacity for lignocellulose degradation, only a limited group is actively filling this niche at a given time. This suggests that rhizosphere priming is influenced by the stage of root growth, and that the organisms catalyzing this decomposition may be more limited than previously expected based on genomic or metagenomic surveys.

This research is based upon work supported by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Numbers DE-SC0010570 and DE-SC0016247 to UC Berkeley and the University of Oklahoma. Part of this work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

# A Simple Pyrocosm for Controlled, Replicated Studies of Post-Fire Soil Microbial Communities

Akiko Carver,<sup>1\*</sup> (aacarver@berkeley.edu) and Thomas D Bruns<sup>1</sup>

<sup>1</sup>University of California, Berkeley

# Project Goals: Our goal is to produce an experimental system to study post-fire microbial communities that enables control and replication of key conditions.

Post-fire soil is a chemically distinct environment that is structured by depth in a predictable pattern due to the thermal properties of soil and the heat gradient that is produced by fire. The upper layers are rich in pyrolyzed organic material (PyrOM). These partially burned organic compounds include forms that are highly recalcitrant to degradation and effectively sequester carbon in the soil for decades [1, 2]. Volatilized waxes and lipids condense in the soil at shallow depths where the peak temperatures reach around 220°C. This creates a hydrophobic zone that prevents water from percolating and facilitates surface runoff and erosion, which can reduce site productivity for decades [3-6].

Pyrophilous fungi are a specialized guild of post-fire saprobes that fruit exclusively on burned soil and appear in an ordered sequence following fire [7, 8]. These fungi are readily isolated into axenic culture and are known to grow rapidly in sterile or burnt soil [9, 10], but little is know about their function in post-fire soils. Recently it was found that some members of this guild dominate soil in the first months after forest fire (Bruns, unpublished results). We predict that pyrophilous fungi and other post-fire microbes are likely to interact with the unique post-fire soil chemistry in ways that impact carbon storage and long-term site productivity.

To test our predictions we developed a simple system the replicates the post-fire soil environment in controlled, repeatable ways and allows us to manipulate the key variables of heat, water, and inoculum. To that end we have developed a "pyrocosm," consisting of a metal bucket filled with 6.5 litters of unsterilized, dry forest soil, and topped with an organic/litter layer similar to intact forest soil. The bucket is buried in the ground to surface depth. A small fire is built on the surface of the soil, and heat at various depths is monitored with thermocouples for 12 hours. After soil has cooled to ambient soil temperature, it is wetted, covered, and incubated on site for three weeks.

By varying fuel loads and using amplicon metagenomics of the internal transcribed spacer region we have learned the following: 1) There is a lag time of several hours after the fire goes out before peak temperatures are achieved at a given depth. 2) Peak temperatures are predicted by the log10 of soil depth. 3) Controlling mass of coarse fuels yields repeatable temperature profiles, while flash fuels have almost no measurable effects. 4) *Pyronema spp.* are stimulated in uninoculated forest soil, and dominate the pyrocosm within one to two weeks.

The results are important for several reasons. First, the repeatability of the temperature profiles and the predicable depth gradient mimics the expected profile in forest fires [11]. Second, the stimulation and dominance of *Pyronema spp*. mimics closely what we see in real forest fire events. Thus, we have been able to recreate conditions that are biologically similar enough to forest fires that we can stimulate the major fungal dominant species seen in real fire events. Our next steps will be to verify that a soil chemical profile is also created that matches the profile expected of a real fire, and to test the way select soil microbes interact with it.

### References

- 1. Santin, C., S.H. Doerr, C.M. Preston, and G. Gonzalez-Rodriguez, Pyrogenic organic matter production from wildfires: a missing sink in the global carbon cycle. Global Change Biology, 2015. 21(4): p. 1621-1633.
- Lehmann, J., J. Skjemstad, S. Sohi, J. Carter, M. Barson, P. Falloon, K. Coleman, P. Woodbury, and E. Krull, Australian climate-carbon cycle feedback reduced by soil black carbon. Nature Geoscience, 2008. 1(12): p. 832-835.
- 3. Certini, G., Effects of fire on properties of forest soils: a review. Oecologia, 2005. 143(1): p. 1-10.
- de Blas, E., M. Rodriguez-Alleres, and G. Almendros, Speciation of lipid and humic fractions in soils under pine and eucalyptus forest in northwest Spain and its effect on water repellency. Geoderma, 2010. 155(3-4): p. 242-248.
- 5. Garcia-Chevesich, P., R. Pizarro, C.L. Stropki, P.R. de Arellano, P.F. Ffolliott, L.F. DeBano, D.G. Neary, and D.C. Slack, Formation of Post-fire Water-Repellent Layers in Monterrey Pine (Pinus radiata D. DON) Plantations in South-Central Chile. Journal of Soil Science and Plant Nutrition, 2010. 10(4): p. 399-406.
- Mainwaring, K., I.L. Hallin, P. Douglas, S.H. Doerr, and C.P. Morley, The role of naturally occurring organic compounds in causing soil water repellency. European Journal of Soil Science, 2013. 64(5): p. 667-680.
- 7. Seaver, F.T., Studies in Pyrophilous Fungi: I. The Occurrence and Cultivation of Pyronema. Mycologia, 1909. 1(4): p. 131-139.
- 8. Petersen, P.M., Danish Fireplace Fungi. Dansk Botanisk Arkiv, 1970. 27(3): p. 1-97.
- 9. Elabyad, M.S.H. and J. Webster, <u>Studies on Pyrophilous Discomycetes 1</u>. Comparative Physiological Studies. Transactions of the British Mycological Society, 1968. 51: p. 353-367.
- 10. Elabyad, M.S.H. and J. Webster, Studies on Pyrophilous Discomycetes 2. Competition. . Transactions of the British Mycological Society, 1968. 51: p. 369-&.
- 11. Massman, W.J., Modeling soil heating and moisture transport under extreme conditions: Forest fires and slash pile burns. Water Resources Research, 2012. 48.

Initial work was funded by NSF Dimensions of Biodiversity grant 1046115 to T.D. Bruns and JW Taylor. Current work is funded by Department of Energy, Systems Biology Enabled Research on the Roles of Microbial Communities in Carbon Cycle Processes program, grant DE-SC0016365 to T.D. Bruns, Thea Whitman, Matthew Traxler, and Igor Grigoriev.

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

Thea Whitman,<sup>1,\*</sup>(twhitman@wisc.edu), Matthew Traxler,<sup>2</sup> Igor Grigoriev,<sup>3</sup> and **Thomas D Bruns**<sup>2</sup>

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

# **Project Goals:**

The frequency of large, high severity wild fires is increasing in the western US and in regions around the world due to long-term fire suppression strategies and climate change [1]. These fires have direct, negative effects on soil carbon stocks through combustion, but they have indirect and potentially positive effects on soil carbon stocks through the production of pyrolyzed organic matter (PyOM) [2]. We are dissecting the effects of microbes on post-fire soil carbon dynamics by using a systems biology approach that couples small experimental "pyrocosms", highly controlled production of <sup>13</sup>C-labeled pyrolyzed substrates, genomics, transcriptomics, stable isotope techniques, and mass spectrometry.

Post-fire soil systems are important to understand, because they have significant direct and indirect effects on global carbon storage. For example, fires result in a large amount of carbon that remains resident on the site as dead and partially pyrolyzed (*i.e.*, burnt under low oxygen) material that has long residence times and constitutes a major pool of C in fire-prone ecosystems [3],[4]. In addition, fire-induced hydrophobic soil layers, caused by condensation of pyrolyzed waxes and lipids, increase post-fire erosion and lead to long-term productivity losses [5]. Soil microbes are likely to be involved with the degradation of all of these compounds, yet little is currently known about the organisms or metabolic processes involved. We are using the following objectives and hypotheses to address our goal of understanding how the post-fire microbial community affects the fates of pyrolyzed carbon and soil carbon in post-fire soil environments:

# Objectives:

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

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

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

# Hypotheses:

1. Specific microbes will colonize post-fire and PyOM-amended soils in a predictable sequence.

2.1. The earliest microbial colonizers will primarily target easily-mineralizable carbon sources.

2.2. Early microbial colonizers will be able to at least partially degrade the hydrophobic layer produced from fire.

2.3. The second stage microbial colonizers will target partially pyrolyzed lignocellulose and non-water-soluble PyOM.

3.1. PyOM additions will stimulate a subset of pyrophilous microbes independent of heat.

3.2. PyOM additions will result in changes in SOC mineralization rates.

We propose a physicochemical gradient model that evokes the steep heat gradient produced in soils by forest fires to predict the soil chemical environment and patterns of microbial recolonization. This model predicts that the post-fire soil environment is likely to be much simpler and more experimentally tractable than undisturbed soil systems. Preliminary work shows that the fungal community and at least a part of the bacterial community fit this model and are indeed simplified following fire. We further simplify this system by moving it into experimental "pyrocosms" that allow us to control and replicate the physiochemical gradient, and by using a "charcoalator" to create highly reproducible, <sup>13</sup>C-labeled forms of pyrolyzed carbon. In addition, we study the response of individual dominant organisms to these environments as well as the response of intact communities, and by deconstructing the community in this way, we expect to maximize the power of genomics and simplify the problem of interpreting metatranscriptomes and metabolomes. The combination of these relatively simple communities in the context of this heat structured, chemically complex, yet manipulable environment will give us unprecedented ability to dissect carbon storage in this pivotal system.

# References

- 1. Westerling, A.L., H.G. Hidalgo, D.R. Cayan, and T.W. Swetnam, Warming and earlier spring increase western US forest wildfire activity. Science, 2006. 313(5789): p. 940-943.
- 2. Gonzalez-Perez, J.A., F.J. Gonzalez-Vila, G. Almendros, and H. Knicker, The effect of fire on soil organic matter a review. Environment International, 2004. 30(6): p. 855-870.
- 3. Santin, C., S.H. Doerr, C.M. Preston, and G. Gonzalez-Rodriguez, Pyrogenic organic matter production from wildfires: a missing sink in the global carbon cycle. Global Change Biology, 2015. 21(4): p. 1621-1633.
- 4. Lehmann, J., J. Skjemstad, S. Sohi, J. Carter, M. Barson, P. Falloon, K. Coleman, P. Woodbury, and E. Krull, Australian climate-carbon cycle feedback reduced by soil black carbon. Nature Geoscience, 2008. 1(12): p. 832-835.
- 5. DeBano, L.F., The role of fire and soil heating on water repellency in wildland environments: a review. Journal of Hydrology, 2000. 231: p. 195-206.

# Funding Statement

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

### Systems Biology to Improve Camelina Seed and Oil Quality Traits

Chaofu Lu<sup>1\*</sup> (clu@montana.edu) and John Browse<sup>2</sup>

<sup>1</sup>Montana State University, Bozeman; <sup>2</sup>Washington State University, Pullman

Project Goals: Camelina has great potential to become a sustainable high energy-yielding source of biofuel in the US. This project aims to address two critical needs for realizing this potential: to increase seed size and oil content for improved seedling establishment and oil yield, and to optimize oil quality for satisfactory fuel properties. Specifically, quantitative trait loci (QTLs) and molecular markers associated with these important traits will be identified using high-density genome maps and repeated field trials in Montana and Washington states. Modern genomics and biotechnological approaches will be employed to uncover novel molecular mechanisms (including genes and gene networks regulated by microRNAs and transcription factors) regulating fatty acid modification, oil accumulation and seed size in Camelina.

Camelina (*Camelina sativa*) is a low-input, non-food oilseed plant that has great potential to become a sustainable high energy-yielding source of biofuel in the US. This project aims to address two critical needs for realizing this potential: 1) to increase seed size and oil content for improved seedling establishment and oil yield, and 2) to optimize oil quality for satisfactory fuel properties.

The available genomic resources (including genome, transcriptome and microRNA sequences) of Camelina and its close relative Arabidopsis and our efficient biotechnological research tools will ensure our success to achieve the following specific objectives:

- To identify quantitative trait loci (QTLs) and molecular markers associated with seed size, oil content and other important agronomic traits. Association studies of a collection of over 250 Camelina accessions and molecular mapping using recombinant inbred line (RIL) populations will be conducted. Agronomic traits will be evaluated in repeated field trials in Montana and Washington states.
- 2. To discover novel molecular mechanisms (including gene networks regulated by microRNAs and transcription factors) regulating fatty acid modification and seed size in Camelina. MicroRNAs affecting seed traits will be obtained by a high-throughput screen in transgenic Camelina for seed-specific overexpression. Bioinformatics analyses, genetics

and biochemical experiments will be conducted to decipher the mechanisms through which miRNAs affect seed development and oil accumulation.

**3.** To test functions of known genes, as well as newly identified candidates from 1&2 for their effects on improving seed size, oil content and fatty acid composition. Novel desaturases and transcription factors that have shown promising preliminary results will be vigorously tested for their roles in improving fatty acid composition and increasing seed size and oil content. These efforts are facilitated by our efficient biotechnological tools such as an Agrobacterium-mediated transformation protocol and the CRISPR/Cas9 method that allow for efficient gain/loss-of-function testing of beneficial genes in a semi-high-throughput fashion. Improved traits will be validated in field conditions following USDA APHIS guidelines.

This project will result in several natural and engineered Camelina lines with advanced traits like large seed and high content of high-quality oil. Increased oil content in larger Camelina seeds will greatly boost harvestable oil yield and improve the ability to establish seedlings particularly under unfavorable field conditions. Camelina oils containing ultrahigh oleic (>80%) and enhanced palmitoleic acids will greatly improve fuel properties such as enhanced oxidative stability and cold flow. Because novel approaches will be explored to engineering beneficial traits in Camelina, this program will also contribute to understanding fundamental mechanisms regulating oilseed metabolism and physiology. Specifically, identification of genes and molecular markers associated with seed size, oil content and other important agronomic traits will facilitate the development of the next generation Camelina varieties. Novel gene regulatory networks uncovered by studying microRNAs will greatly advance the understanding of plant lipid metabolism, and provide new tools for improving oil qualities of many oilseed crops. In addition, this project will provide training opportunities for several young scientists such as graduate students and postdocs in plant genomics and biotechnology.

This research project is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE- SC0016400.

# Trait-based approaches for linking metagenomic data with microbial carbon cycling under drought conditions

**Steven D. Allison (allisons@uci.edu)**,<sup>1</sup>\* Jennifer B. H. Martiny<sup>1</sup>, Adam C. Martiny,<sup>1</sup> Renaud Berlemont,<sup>1,2</sup> Kathleen K. Treseder,<sup>1</sup> Michael L. Goulden,<sup>1</sup> and Eoin L. Brodie<sup>3</sup>

<sup>1</sup>University of California, Irvine, CA; <sup>2</sup>California State University, Long Beach, CA; and <sup>3</sup>Lawrence Berkeley National Laboratory, Berkeley, CA

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

Project Goals: A key challenge in microbial ecology is to predict the functioning of microbial communities under changing environmental conditions. Metagenomics and other high-throughput molecular approaches can help address this challenge by revealing the functional potential of microbial communities. We coupled metagenomics with models and experimental manipulations to address microbial responses to drought in a California grassland ecosystem along with the consequences for carbon cycling (Fig. 1). We developed an approach for extracting trait information from metagenomic data and asked: 1) What is the phylogenetic structure of drought response traits? 2) What is the relationship between these traits and those involved in carbohydrate degradation? 3) How do both classes of traits vary seasonally and with precipitation manipulation? 4) How resilient are these traits in the face of perturbation?

We found that drought response traits are phylogenetically conserved at an equivalent of 5-8% ribosomal RNA gene sequence dissimilarity. Experimental drought treatment selected for the genetic potential to degrade starch, xylan, and mixed polysaccharides, suggesting a link between drought response and carbon cycling traits. In addition, microbial communities exposed to experimental drought showed a reduced potential to degrade plant biomass. Particularly among bacteria, seasonal drought had a larger impact on microbial composition, abundance, and carbohydrate-degrading genes compared to experimental drought. Bacterial communities were also more resilient to drought perturbation than fungal communities, which showed legacies of drought perturbation for up to three years (Fig. 2). Altogether, these findings imply that microbial communities exhibit trait diversity that facilitates resilience but with substantial time lags and consequences for carbon turnover. This information is being used to inform new traitbased models that address the challenge of predicting microbial functioning under precipitation change.

# Publication

1. Martiny, J. B. H., C. Weihe, Y. Lu, R. Berlemont, E. L. Brodie, M. L. Goulden, K. K. Treseder, and S. D. Allison. 2016. Microbial legacies alter decomposition in response to simulated global change. *ISME Journal* in press. doi:10.1038/ismej.2016.122

This research was funded by the Genomic Science Program, the (now retired) Program in Ecosystem Research (PER), and direct support to Lawrence Berkeley National Laboratory. The Genomic Science Program is part of the Biological System Science Division (BSSD) and PER was part of the Climate and Environmental Sciences Division (CESD) of Biological and Environmental Research (BER) in the US Department of Energy Office of Science.



Fig. 1. Drought manipulation



Fig. 2. NMDS ordination of fungal communities four months after transplant

# Comparative genomics and functional characterization of assimilatory sulfate reduction in methanogenic and methanotrophic archaea

Hang Yu<sup>1\*</sup> (hyu@caltech.edu), Dwi Susanti<sup>2</sup>, Shawn McGlynn<sup>3</sup>, Connor Skennerton<sup>1</sup>, Patricia Tavormina<sup>1</sup>, Biswarup Mukhopadhyay<sup>2</sup>, and **Victoria Orphan**<sup>1</sup>

<sup>1</sup>Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California, 91125, USA; <sup>2</sup>Department of Biochemistry, Virginia Tech, Blacksburg, VA 24061, USA; <sup>3</sup>Tokyo Institute of Technology, Earth-Life Science Institute, Tokyo, 152-8550, JAPAN

Project Goals: Biological methane production or consumption may be constrained by the availability of sulfur in the environment. Comparative genomics could be used to characterize metabolic pathways and select for key genes of interest for further experimental studies. We applied this approach in our project below to understand how methane-metabolizing archaea meet their obligate requirement for sulfur in its ecological niche.

Methanogenic and methanotrophic archaea are critical in the global carbon cycle, and their growth is dependent on sulfur. While sulfide is the most common sulfur source, other sulfur compounds such as sulfite, thiosulfate and elemental sulfur can be present in their environments, but the capability and mechanism of assimilation are less well-understood. Here we explored new genomes of anaerobic methanotrophic archaea (ANME) by comparing to genomes of their relatives, and found a pathway for sulfate reduction that was overlooked previously. Phylogenetic analyses suggest that while intermediate sulfur species usage maybe more widespread, the ability to activate sulfate is more restricted. Multiple homologs of sulfite reductases could be found in these archaeal genomes, and we examined their transcriptional responses to different sulfur species. The result is consistent between ANME and a cultured relative in the Methanosarcinales, Methanococcoides burtonii. Further investigations on Group II Fsr, a sulfite reductase found in all methane seep environments surveyed, revealed a novel substrate of this enzyme as supported by protein homology models and heterologous expression studies. This expanded sulfur-utilizing ability may help ANME, or methane-metabolizing archaea in general, to broaden their environmental niche and thrive in conditions regardless of the oxidation state of sulfur.

This work was supported by the Department of Energy BER program (DE-SC0016469) and funding by the Gordon and Betty Moore Foundation through Grant GBMF3306 (to VO).

### Development of Resources and Tools to Improve Oil Content and Quality in Pennycress

Ana Paula Alonso<sup>1</sup>\* (alonso.19@osu.edu), Erich Grotewold,<sup>1</sup> and Ajay Shah<sup>1</sup>

<sup>1</sup>The Ohio State University, Columbus, OH

Project Goals: The main objective of this research is to apply a biodesign strategy for improving oil content in a promising alternative source of jet-fuel, pennycress. To advance towards this goal, we are: 1) Investigating pennycress natural variation to identify candidate genes and biomarkers associated with oil accumulation and fatty acid composition; 2) Identifying targets to improve oil content and composition, and 3) Establishing metabolic engineering targets and develop community resources.

Bioenergy crops, which have potential for jet fuel production and do not compete with food crops, are urgently needed. Our strategy to address this fundamental challenge is to develop pennycress (*Thlaspi arvense*), a member of the Brassicaceae, as a bioenergy crop, taking advantage of its ability to produce seed oil that is ideally suited as a renewable source of biodiesel and aviation fuel<sup>1</sup>. Moreover, pennycress performs well on marginal land, has a short maturity time and grows off-season, serving as a winter cover crop, and complementing the production of commodity food crops. Essential pennycress molecular and genetic resources can be rapidly developed by leveraging the fully sequenced genome and research tools of its close relative *Arabidopsis thaliana*. Furthermore, a draft genome<sup>2</sup> for *Thlaspi arvense*, as well as transcriptome<sup>3</sup> and metabolome<sup>4</sup> information have been made recently available. However, for this plant to become an economically viable and sustainable source of jet fuel, molecular and genetic resources need to be developed, and integrated with multi-variable techno-economical analyses to guide strategies for increase oil production through breeding and/or genetic manipulation. These are the gaps that this project intends to fill.

Specifically, we are: 1) Investigating pennycress natural variation to identify candidate genes and biomarkers associated with oil accumulation and fatty acid composition. For this purpose, we are determining variation in genome-wide gene expression (derived from RNA-Seq) and intracellular metabolites (derived from metabolomics) in seeds from pennycress natural accessions. 2) Identifying targets to improve oil content and composition. Towards this goal, we are generating a flux map of carbon partitioning in developing pennycress embryos, and will overlay metabolic maps with levels of transcripts and intracellular compounds to identify metabolic bottlenecks in oil accumulation. Finally, we will use <sup>13</sup>C-based metabolic flux analysis to validate bottlenecks in two accessions with contrasting oil contents. 3) Establishing metabolic engineering targets and develop community resources. For this purpose, we are analyzing techno-economics of pennycress based agronomic and supply systems that will provide targets for future efforts aimed at increasing seed oil production through rational metabolic engineering and breeding. We will also develop a public seed collection of pennycress mutants and transgenic lines, facilitating community synergy and accelerating research towards the established production goals.

Taken together, the knowledge and resources generated through this interdisciplinary project will facilitate rational breeding and metabolic engineering of pennycress and related alternative bioenergy crops.

### References

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

2. Dorn, K.M., Fankhauser, J.D., Wyse, D.L. & Marks, M.D. A draft genome of field pennycress (*Thlaspi* arvense) provides tools for the domestication of a new winter biofuel crop. *DNA Res* 22, 121-31 (2015).

3. Dorn, K.M., Fankhauser, J.D., Wyse, D.L. & Marks, M.D. De novo assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *Plant J.* **75**, 1028-1038 (2013).

4. Tsogtbaatar, E., Cocuron, J.C., Sonera, M.C. & Alonso, A.P. Metabolite fingerprinting of pennycress (*Thlaspi arvense* L.) embryos to assess active pathways during oil synthesis. *J Exp Bot.* **66**, 4267-4277 (2015).

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0016490.

# Get Your Model Out of the Clouds! Ground-Truthing Assumptions About how the Earth's Tiniest Engines Drive the Carbon Cycle

Grace Pold<sup>1\*</sup> (apold@umass.edu), Seeta Sistla,<sup>2</sup> Serita D. Frey,<sup>3</sup> Kevin Geyer,<sup>3</sup> Stuart Grandy,<sup>3</sup> and Kristen M. DeAngelis<sup>4</sup>

<sup>1</sup>University of Massachusetts, Amherst; <sup>2</sup>Hampshire College, Amherst, MA; <sup>3</sup>University of New Hampshire, Durham

Project Goal: Our long-term goal is to better predict the fate of soil C by developing mathematical and ecosystem models of SOM decomposition that incorporate a genetic and physiological understanding of CUE. Specifically, we propose to define the genomic basis of soil microbial CUE in the context of changing environments, and use this new knowledge to generate more realistic models of SOM decomposition.

Soil microbes are central players in Earth's carbon cycle, yet the diverse physiologies displayed by these organisms are poorly integrated into models of the carbon cycle. Such negligence exists on the one hand due to the ignorance about the role soil microbes play, and on the other hand due to the infeasibility of integrating the entirety of microbial physiology into models. To overcome these shortcomings, we are completing extensive physiology and comparative genomics of bacterial and fungal isolates to identify scalable genomic markers of key steps in microbial processing of soil organic matter. This data will be used to modify the structure and parameterization of soil carbon models working on a number of temporal and spatial scales in order to evaluate how soil carbon stocks may change with climate warming. Preliminary research shows that the efficiency with which bacteria convert soil carbon to additional cell biomass differs between temperatures in a "species"-specific manner. Further work to identify the specific responses of bacteria and fungi to other environmental stressors is expected to greatly expand on the limited existing data currently used to parameterize soil carbon models, and therefore improve their ability to project soil carbon stocks in decades to come.

This work is supported by the University of Massachusetts Graduate School, the University of Massachusetts Graduate Program in Organismic and Evolutionary Biology, and the Office of Biological and Environmental Research in the DOE Office of Science.

# Fungal Responses To Elevated Temperature And Soil Nitrogen Availability

Shana Whitney<sup>1</sup>, Kevin Geyer<sup>1</sup>, Eric W. Morrison<sup>1</sup>, Kristen M. DeAngelis<sup>2</sup>\* (deangelis@microbio.umass.edu), and **Serita Frey**<sup>1</sup>

<sup>1</sup>University of New Hampshire, Durham; <sup>2</sup>University of Massachusetts, Amherst

Project Goals: Our long-term goal is to better predict the fate of soil C by developing mathematical and ecosystem models of SOM decomposition that incorporate a genetic and physiological understanding of CUE. Specifically, we propose to define the genomic basis of soil microbial CUE in the context of changing environments, and use this new knowledge to generate more realistic models of SOM decomposition.

The soil microbial community controls decomposition of organic residues which constitute a large portion of soil organic matter. Microbial growth is impacted by global changes such as warming and soil nitrogen (N) availability. Carbon use efficiency (CUE) is an important parameter that influences soil C dynamics by partitioning organic matter between soil C and CO<sub>2</sub> pools. This research focuses on the growth of different fungal species' exposed to varying temperatures and N availabilities, while quantifying respiration (CO<sub>2</sub> flux) and microbial growth. To assess individual fungal isolates, we constructed a sterilized artificial soil medium to mimic a sandy loam soil by mixing 70% sand, 20% silt, and 10% clay. Several fungal species of the phyla *Ascomycota* and *Basidiomycota* were individually grown in this media at different temperatures (15 and 25°C) and N levels. Soil respiration was measured over the incubation period. Fungal biomass was estimated by chloroform fumigation extraction and qPCR of the fungal ITS region. Our results indicate that fungi were able to grow effectively and reproducibly in the artificial soil medium, demonstrating that using an artificial soil is an effective method for assessing individual species responses. Temperature and N availability had a positive affect on C mineralization and biomass. CUE varied among fungal species and, in general, declined with temperature.

This work is supported by an REU supplement to our National Science Foundation LTREB grant, and the Office of Biological and Environmental Research in the DOE Office of Science.

# Microbial Growth and Metabolism in Soil – Refining the Interpretation of Carbon Use Efficiency

Kevin Geyer,<sup>1</sup> Kristen M. DeAngelis<sup>2</sup>\* (email@address.gov), and Serita Frey<sup>2</sup>

<sup>1</sup>University of New Hampshire, Durham; <sup>2</sup>University of Massachusetts, Amherst

Project Goals: Our long-term goal is to better predict the fate of soil C by developing mathematical and ecosystem models of SOM decomposition that incorporate a genetic and physiological understanding of CUE. Specifically, we propose to define the genomic basis of soil microbial CUE in the context of changing environments, and use this new knowledge to generate more realistic models of SOM decomposition.

Carbon use efficiency (CUE) describes a critical step in the terrestrial carbon cycle where microorganisms partition organic carbon (C) between stabilized organic forms and CO<sub>2</sub>. Application of this concept, however, begins with accurate measurements of CUE. Both traditional and developing approaches still depend on numerous assumptions that render them difficult to interpret and potentially incompatible with one another. Here we explore the soil processes inherent to traditional (e.g., substrate-based, biomass-based) and emerging (e.g., growth rate-based, calorimetry) CUE techniques in order to better understand the information they provide. Soil from the Harvard Forest Long Term Ecological Research (LTER) site in Massachusetts, USA, was amended with both <sup>13</sup>C-glucose and <sup>18</sup>O-water and monitored over 72 h for changes in dissolved organic carbon (DOC), respiration (R), microbial biomass (MB), DNA synthesis, and heat flux (O). Four different CUE estimates were calculated: 1) ( $\Delta DOC -$ R)/ $\Delta$ DOC (substrate-based), 2)  $\Delta^{13}$ C-MB/( $\Delta^{13}$ C-MB + R) (biomass-based), 3)  $\Delta^{18}$ O-DNA/( $\Delta^{18}$ O-DNA + R) (growth rate-based), 4) Q/R (energy-based). Our results indicate that microbial growth (estimated by both  $^{13}$ C and  $^{18}$ O techniques) was delayed for ~40 h after amendment even though DOC had declined to pre-amendment levels within ~48 h. Respiration and heat flux also peaked after 40 h. Although these soils have a relatively high organic C content (5% C), respired CO<sub>2</sub> was greater than 88% glucose-derived throughout the experiment. All estimates of microbial growth (Spearman's  $\rho > 0.83$ , p<0.01) and efficiency (Spearman's  $\rho > 0.65$ , p<0.05) were positively correlated, but strong differences in the magnitude of CUE suggest incomplete C accounting. This work increases the transparency of CUE techniques for researchers looking to choose the most appropriate measure for their scale of inquiry or to use CUE estimates in modeling applications.

This work is supported the New Hampshire Agricultural Experiment Station (NHAES), and the Office of Biological and Environmental Research in the DOE Office of Science.
# Mapping photoautotrophic metabolism of engineered cyanobacteria to identify reactions that limit production of renewable chemicals

Yi Ern Cheah<sup>1</sup>\* (<u>vi.ern.cheah@vanderbilt.edu</u>), Yao Xu<sup>2</sup>, Carl H. Johnson<sup>2</sup>, Jamey D. Young<sup>1,3</sup>

<sup>1</sup> Chemical & Biomolecular Engineering, Vanderbilt University, Nashville, TN; <sup>2</sup> Biological Sciences, Vanderbilt University, Nashville, TN; <sup>3</sup> Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN

## http://www.vanderbilt.edu/younglab

**Project Goals**: This project aims to study the metabolism of engineered cyanobacteria to identify reactions that limit carbon flux to synthetic pathways. The long term goal is to develop technologies that can be used to enhance the performance of industrially pertinent photosynthetic microorganisms.

Cyanobacteria are emerging as ideal biocatalysts for the synthesis of renewable fuels and chemicals from sunlight and CO<sub>2</sub>. Despite current advances in cyanobacterial production systems, the efficiencies needed for large scale commercialization have yet to be achieved. In addition to the limited tools available for streamlining strain performance, redirecting carbon flux from central metabolism to product producing pathways remains a challenge. Mapping metabolism of these engineered systems can help identify metabolic reactions that limit carbon flux to desired product(s) and pinpoint 'wasteful' by-product pathways that pull carbon flux away from central metabolism.

Our group focusses on developing novel approaches that use isotopically nonstationary <sup>13</sup>C-MFA to quantitatively assess *in vivo* metabolic phenotypes of photoautotrophs [1-3]. We have previously applied this approach to characterize photoautotrophic metabolism of the model terrestrial plant *Arabidopsis thaliana* [4] and the model cyanobacteria *Synechocystis* PCC6803 [5]. More recently, we used this approach to map carbon fluxes and identify metabolic bottlenecks in an isobutyraldehyde (IBA) producing mutant of the cyanobacteria *Synechococcus elongatus* PCC7942 (strain SA590). By overexpressing the genes associated with the metabolic bottleneck, we successfully generated mutants that showed significant improvement in IBA productivity over SA590 [6].

This presentation summarizes the lessons we learned and demonstrates the potential of using <sup>13</sup>C-MFA in tandem with rational metabolic engineering methodologies to enhance the performance of industrially relevant photosynthetic microorganisms.

#### References

[1] Jazmin, L. J., O'Grady, J. P., Ma, F., Allen, D. K., *et al.*, Isotopically Nonstationary MFA (INST-MFA) of Autotrophic Metabolism, in: Dieuaide-Noubhani, M., Alonso, A. P. (Eds.), *Plant Metabolic Flux Analysis: Methods and Protocols*, Humana Press, Totowa, NJ 2014, pp. 181-210.

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

[3] Adebiyi, A. O., Jazmin, L. J., Young, J. D., C-13 flux analysis of cyanobacterial metabolism. *Photosynthesis Research* 2015, *126*, 19-32.

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

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

[6] Jazmin L.J., Xu Y., Cheah Y., Adebiyi A. A., Johnson C.H., Young J.D., Isotopically nonstationary <sup>13</sup>C flux analysis of cyanobacterial isobutyraldehyde production." (*Under revision*).

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

# Pooled Assembly, Genotyping and Scoring of Complex Synthetic Genomic Libraries

\*Robert Egbert<sup>1,2</sup> (rgegbert@lbl.gov), Eric Yu<sup>2</sup>, Dylan McCormick<sup>3</sup>, Ben Adler<sup>2</sup>, **Adam Arkin**<sup>1,2</sup>

<sup>1</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>2</sup>Department of Bioengineering, University of California Berkeley; <sup>3</sup>Department of Molecular and Cell Biology, University of California Berkeley.

Precision engineering of complex behaviors in living systems is complicated by our limited understanding of the contextual determinants of gene expression and our limited ability to manipulate the genetic code in individual cells at scale. Iterative design, build and test cycles thus often sample only a small fraction of the functional parameter space. Further, the relative ease of optimizing gene networks on plasmids for bacterial expression often is not predictive of system performance when chromosomally expressed for deployment in complex environments from bioreactor to soil or gut microbial communities. We have developed a comprehensive engineering platform for pooled assembly, genotyping and single-variant fitness scoring of barcoded genomic libraries coupled with retrieval of individual genotypes from the library via CRISPRi. Mapping individual genotypes to fitness drives predictions of function across the sequence landscape, balancing the costs and benefits of expression.

To validate the genome engineering platform we generated a genomic library of over two million barcoded variants of the violacein biosynthetic pathway *vioABCDE* from *Chromobacterium violaceum* in *E. coli*, sampling from a genotype space of over 260,000 combinations. We employed  $\lambda$ -Red recombination to serially integrate five DNA fragment libraries - here, each a gene with a degenerate ribosome binding site - fused to a selection-enrichment cassette (SEC) with homology for the target genomic locus. Successive "inchworm" integration stages replaced the SEC from the previous stage, cycling among SECs, each consisting of one of two fluorescence markers and one of three antibiotic resistance markers. We pooled transformants from each integration stage using fluorescence activated cell sorting to enrich for the expected fluorescence phenotype and screen out spontaneous resistance mutants and off-target integration events manifest as cells with errant fluorescent phenotypes.

We genotyped the pooled library by associating an individual barcodes with the ribosome binding site sequence of each *vio* gene by deep sequencing amplicon fusion libraries generated by emulsion PCR<sup>1</sup>. Time-series measurements of barcode

abundance allow us to generate fitness scores for each barcoded variant under pathway induction conditions and in resource competition with violacein-sensitive *Bacillus subtilis* to identify variants that optimally balance the benefit of violacein production as an antimicrobial against the cost of gene expression. We are analyzing the genotypes and fitness scores to comprehensively map the expression space and predict sequence to function relationships.

To isolate individual genotypes from the pooled library we developed a barcode-specific CRISPR interference (CRISPRi) technique to use with fluorescence-activated cell sorting. We used the random barcode embedded in the SEC of the final integration stage to repress GFP expression with the RNA-encoded guide sequence specific repressor dCas9<sup>2</sup>. We validated the isolation of individual genotypes from a library of eight clones with a wide range of GFP expression levels by sorting the library transformed with CRISPRi plasmids that encode dCas9 with a guide RNA that target individual barcodes. We have also demonstrated the retrieval of low-abundance cell types from the pooled genomic library by transforming the library with CRISPRi plasmids encoding guides that target rare barcodes and sorting for CRISPRi-responsive cells.

We believe our approach to high-throughput construction, analysis and manipulation of pooled genomic libraries through comprehensively sampling of the parameter space of engineered gene networks will aid the prediction of sequence to function relationships and enable complex engineered phenotypes that are unreachable with existing methods. We anticipate this genomic assembly, screening and isolation platform will advance synthetic biology efforts to optimize large biosynthetic gene clusters for expression of natural products and engineering of other complex cellular behaviors encoded by multiple genes.

This work is supported by the Genome Science program within the Office of Biological and Environmental Research (Project grant number DE-SC008812, Funding Opportunity Announcement DE-FOA-0000640)

## References

- 1. Zeitoun, R.I. *et al.* 2015. Multiplexed tracking of combinatorial genomic mutations in engineered cell populations. *Nature Biotechnol.* **33**, 6, 631–637.
- 2. Qi, L. S. *et al.* 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183.

#### A novel design strategy for industrially relevant, unnatural modular megasynthases

William C. Grau<sup>1,2\*</sup> (william.grau@colorado.edu), Marcelo C. Bassalo<sup>2,3</sup>, and Ryan T. Gill<sup>2</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, University of Colorado Boulder, Boulder, CO 80303. <sup>2</sup>Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO 80303. <sup>3</sup>Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, CO 80303.

#### **Project Goal:**

The project aims to develop a strategy for designing unnatural polyketide synthases (PKSs) that produce industrially relevant molecules in *E. coli*. As descendants of fatty-acid synthases (FASs), PKSs are known for synthesizing natural products with complex structures from simple building blocks such as malonyl-CoA. Many examples of natural PKSs and a few examples of engineered PKSs have been successfully expressed in *E. coli* but no universal approach to engineering PKSs has been elucidated. This project attempts to leverage new computational tools and the power of synthetic biology to solve this problem and thus expand the set of molecules produced at industrially relevant scale by *E. coli*.

#### Abstract:

One of the most significant challenges facing industrial biotechnology as a field is selecting the proper target molecules for strain development. The easiest molecules to produce in industrially relevant strains, such as ethanol in *S. cerevisiae* or fermentation products in *E. coli*, also tend to be the most competitive. This makes it incredibly difficult to balance the cost of strain development with the low-value of the target molecule, and has forced industrial biotechnology into high-value molecules. Generally, these high-value molecules are structurally complex and more synthetic steps from a commodity chemical, increasing the difficulty of strain development. Thus, technologies that simplify the process of strain development for more complex, valuable molecules are in high demand.

To this end, we have developed a new approach to designing modular megasynthases such as Type I Polyketide Synthases (PKSs) or Non-ribosomal Peptide Synthases (NRPSs). This design approach starts with a computational pipeline that searches publically available bacterial genomes for design rules specific to PKSs and NRPSs. This pipeline outputs amino acid sequences for linking the modular catalytic domains. These amino acid linker sequences are then combined with the requisite catalytic domains to create a complete amino acid sequence for an unnatural PKS/NRPS that should synthesize the target molecule. Codon harmonization is applied to this amino acid sequence to generate a nucleotide sequence that is then incorporated into the *E. coli* genome using a CRISPR/Cas9-based technique reported by our lab.<sup>1</sup> This gene is assembled via TAR cloning in *S. cerevisiae*. Here we report the first complete implementation of this design strategy to develop an *E. coli* strain that produces the fragrance ingredient deltahexalactone via an unnatural Type I PKS. This strategy is universally applicable to all Type I PKSs and NRPSs and theoretically can be used to expand the biosynthetic capabilities of *E. coli* to increasingly complex molecules.



Figure 1. Outline of approach for the design of unnatural Type I PKS producing delta-hexalactone in E. coli.

<sup>1</sup>Bassalo, M.C., Garst, A.D., Halweg-Edwards, A.L., Grau, W.C., Domaille, D.W., Mutalik, V.K., Arkin, A.P., Gill, R.T.. Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. ACS Synth. Biol. 2016. doi:10.1021/acssynbio.5b00187

#### **Funding Statement:**

**Grant title:** A Platform for Genome-scale Design, Redesign, and Optimization of Bacterial Systems; Project grant number (DE-SC008812) and FWP number (ERWER44) for NREL.

# Combinatorial engineering of 3-hydroxypropionate production from hemicellulose hydrolysate

Rongming Liu<sup>1</sup>\* (<u>Rongming.liu@colorado.edu</u>), Liya Liang<sup>1</sup>, Andrew Garst<sup>2</sup>, Alaksh Choudhurya<sup>1</sup>, Violeta Sànchez i Noguéc<sup>3</sup>, Gregg T. Beckhamc<sup>3</sup>, and **Ryan T. Gill<sup>1</sup>** 

<sup>1</sup>Renewable and Sustainable Energy Institute, University of Colorado Boulder, Boulder, Colorado; <sup>2</sup>Muse Biotechnology Inc., Boulder, Colorado; and <sup>3</sup>National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colorado

Project Goals: Engineering of strains for industrial production requires the targeted improvement of multiple complex traits ranging from pathway flux to tolerance to mixed sugar utilization. Here, we report the use of an iterative CRISPR EnAbled Trackable genome Engineering (iCREATE) method for generating targeted genomic modifications at high efficiency along with high throughput phenotypic screening and growth strategies to rapidly engineer multiple traits in *Escherichia coli*.

Advances in DNA synthesis and sequencing have motivated increasingly complex efforts for programming cells on laboratory timescales. Realization of such efforts requires dramatically improved strategies for the precise and efficient editing of genomes with methods that can be performed at throughputs compatible with the latest sequencing technologies. CRISPR EnAbled Trackable genome Engineering (CREATE) couples the high efficiency CRISPR editing with the massively multiplexed rational design offered by parallel oligomer synthesis. This technology enables a single researcher to generate hundreds of thousands of designer variants in a few days and to map each of these variants to a selected phenotype using the designed barcode. To meet complex lab and industrial environments, an iterative CREATE (iCREATE) strategy was constructed and tested in *E. coli* for enhancement of its sugar mixture utilization rate and tolerance of typical hydrolysate inhibitors. Optimal combinations identified were also tested for 3-hydroxypropionate (3HP) production.

After pretreatment of lignocellulose, glucose and xylose are the main carbon sources in the hydrolysate. In *E. coli*, glucose completely inhibits the uptake of xylose, thus limiting the conversion of sugars to product molecules in fermentation of cellulosic biomass. To solve this problem, deletion of *ptsHI* genes was applied in *E. coli* BG. As a result, the strain BGgx (BG,  $\Delta ptsHI$ ) can utilize glucose and xylose simultaneously. However, the glucose consumption rate was significantly lower than that of xylose during the culture. To obtain an efficient sugar mixture utilization strain, RBS Library of *galP* and *glk* gene was constructed in BGgx by iCREATE. After plasmid and genomic sequencing of the top 5 variants, the results showed that BGgxk<sub>4</sub>P<sub>1</sub> and BGgxk<sub>5</sub>P<sub>1</sub> were the positive variants during first-round screening, with cell growth rates 36% higher than BG in 50-ml bioreactor tubes. A second round of iCREATE was then used for BGgxk<sub>4</sub>P<sub>1</sub> and BGgxk<sub>5</sub>P<sub>1</sub> after gRNA plasmid curing. Due to 80% editing efficiency with 410 of CFU/ $\mu$ l, another 3-fold more colonies than the design size were tested in 96-well plates. The results showed that BGgxk<sub>4</sub>P<sub>4</sub> was the best strain in the second round of iCREATE after sequencing of the plasmids and targeting genes of the top 5 variants.

During the pretreatment of corn stover or corn stalk, side-reaction products (furfural, 5hydroxymethylfurfural, formate, acetate, and soluble lignin products) are formed. Here, we used the iCREATE method to construct a 27-gene library about 40,000 mutations for enhancement of hydrolysate tolerance in BGgxk<sub>4</sub>P<sub>4</sub>. These targeted genes encompass the global high level regulators that regulate the central pathway of metabolism, the transcription factors that play important roles in genome level transcription, and enzymes that function in NAD(P)H metabolism and the aldehyde reduction system. After another two round iCREATE, the best producing quadruple mutant strain BGHP<sub>ht</sub> was tested under high furfural and high acetate hydrolysate fermentation, demonstrating a 6.3-7 fold increase in productivity relative to the parent strain.

This work is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

## Developing a predictive method for tunable control over gene expression based on CRISPR interference technology.

Katia Tarasava,<sup>1,2\*</sup> (katia.tarasava@colorado.edu) Jonathan Lavington,<sup>3</sup> Manual Lladser,<sup>3</sup> and **Ryan Gill**<sup>2,4\*</sup>

<sup>1</sup> Department of Materials Science and Engineering, University of Colorado Boulder, Boulder, Colorado;
 <sup>2</sup> Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, Colorado;
 <sup>3</sup> Department of Applied Mathematics, University of Colorado Boulder, Boulder, Colorado;
 <sup>4</sup>Renewable and Sustainable Energy Institute, NREL/University of Colorado, Boulder.

http://www.gillgroup.org/

Project Goals: This project focuses on the development of new methods for controlling cellular gene expression in a multiplex, tunable and predictable manner. The goal is to develop a tool that enables to obtain a specified level of gene expression based on transcriptional control. This method would allow to fine-tune the level of transcription and to investigate the dynamic range of cellular functions. The advantage of this system is that it does not require making chromosomal modifications, as the regulation occurs at the level of transcription. Moreover, it offers a temporal control over gene expression, making it suitable for studying dynamic cellular responses. Such a system can be used, for instance, for metabolic flux optimization of industrial strains, as well as for investigating the mechanism of complex traits that rely upon the coordinated action of combinations of genes and regulators. Applications for these technologies lie within the general areas of sustainable fuels and chemicals production.

Precise control over gene expression is essential for advancing metabolic engineering, as well as general understanding of the global context of cellular regulatory networks. Engineering and optimizing metabolic pathways requires being able to fine-tune expression of multiple genes simultaneously in a precise manner, which is restricted by time-consuming traditional strain engineering methods and a limited number of dose-response promoters. These limitations can be overcome by using CRISPR-based gene repression and activation [1]. In addition to offering multiplex gene regulation, CRISPR interference allows to control the degree of gene expression at the level of transcription. Transcriptional control can be achieved by adjusting the strength of interaction between the guide RNA and target gene through incorporating mismatches into the guide RNA sequence. The number, type and position of mismatches can affect the strength of gRNA binding and consequently, target gene expression. We seek to describe the relationship between gRNA sequence and its effective strength by using an adjustable parameter Markov model. The model can be trained on real expression data and be used as a tool for predicting and precisely controlling the level of transcription for a given gene. The application of this method has the potential to greatly simplify and reduce the cost of strain engineering, as well as provide insight into fundamental properties of metabolic networks.

		gRNA sequence	Repression
f(x)	Original sequence:	GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT	100%
	Mismatched sequences:	GTGAGCAAGAGCGAGGAGCTGTTCACCGGGGT	80%
	Distance from PAM	GAGAGCAAGGGCGAGGAGCTGTTCACCGGGGT	50%
	Type of mismatch	GCGAGCAAGGGCGAGGAGCTGTTCACCGGGGT	65%
	Number of mismatches	GCGATCAAGGGCGAGGAGCTGTTCACCGGGGT	30%
Model	Contiguity of mismatches	G <b>CA</b> AGCAAGGGCGAGGAGCTGTTCACCGGGGT	15%

Figure 1: Example of the parameters that go into a machine-learning algorithm to train the model.

#### References

[1] Qi, Lei S., et al. "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression." Cell 152.5 (2013): 1173-1183.

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

## Enhanced CRISPR-based trackable protein engineering using modeling

Eun Joong Oh<sup>1,2\*</sup> (eun.oh@colorado.edu), Alaksh Choudhury<sup>1,2</sup>, Liya Liang<sup>1,2</sup>, Rongming Liu<sup>1,2</sup> and **Ryan T. Gill**<sup>1,2</sup>

<sup>1</sup>University of Colorado Boulder, Boulder; <sup>2</sup>Renewable and Sustainable Energy Institute, Boulder, Colorado

## http://www.gillgroup.org

Project Goals: CRISPR EnAbled Trackable genome Engineering (CREATE) is a strategy that combines CRISPR/CAS9 editing with multiplexed oligo synthesis, enabling mapping of mutations to traits of interest. In this project, we attempted to apply CREATE to protein engineering. This research project addresses two important issues in developing optimal CRISPR-based trackable protein engineering: efficient cutting of the target site by modulating CAS9 promoters and homologous recombination by controlling homology arm. Modeling aided optimization would provide a better understanding of CREATE system and overcome bottlenecks in implementing CREATE for protein engineering. The long-term goal of the project is to develop optimal protein engineering system in various microbial cell factories including bacteria and yeast.

Humans have been used microbes to produce fermented foods and beverages for a long time. Recently, microbes have been harnessed to produce value-added products as cell factories. Many research groups are investigating strategies to use microbial cell factories for sustainable and economical protein production such as enzymes and antibodies. Because the microbes have extensive regulation between metabolic enzymes, protein engineering by error-prone PCR and rational design sometimes leads to disruptive cellular metabolism and regulatory mechanism. In addition, our knowledge is limited to uncover the underlying system controlling metabolic homeostasis (1). Advances in synthetic biology technologies such as CRISPR/Cas9 system enabled researchers to overcome bottlenecks in microbial genome editing (2). CREATE is an advanced approach based on CRISPR/CAS9 editing and barcode tracking, and it enables multiplex editing and mapping at the genome scale (3).

Currently, we are attempting to apply this CREATE tool to protein engineering. To overcome the bottlenecks in improving editing efficiency of the CREATE system, additional optimization of CRISPR/CAS9 editing and cassette design is necessary. First, modulating CAS9 promoters can control CRISPR toxicity in microorganisms. Modeling aided optimization might improve CAS9 cutting efficiency, resulting in enhanced throughput of the CREATE technology. Second, cassette design including the homology arm (HA) length and the distance between PAM-codon/target site effects the editing efficiency. The validation of cassette design based on the factors is also essential for high-throughput and precise protein engineering. The results can be used as the conditions for cassette design automation.

Our long-term goal is to develop the optimized system for protein engineering including antibodies and enzymes. For example, CREATE-based single chain antibody mutants might provide high-affinity antibodies with trackable high-throughput mapping of desired phenotypes. We could evaluate the

contribution of each mutation to the improved affinity. Also, CREATE-based enzyme evolution might provide a better understanding of protein fitness landscapes than random mutagenesis in optimizing protein function. We envision that the CREATE technology will enhance understanding of complex biological networks in protein engineering and further improve biofuel production in microbes.

#### References

- 1. Nielsen J, Keasling JD. 2016. Engineering Cellular Metabolism. Cell 164:1185–1197.
- 2. **Sander JD**, **Joung JK**. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol **32**:347–355.
- Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, Liang L, Wang Z, Zeitoun R, Alexander WG, and Gill RT. 2016. Genome-wide mapping of mutations at singlenucleotide resolution for protein, metabolic and genome engineering. Nat Biotechnol DOI 10.1038/nbt.3718.

**Grant title**: A Platform for Genome-scale Design, Redesign, and Optimization of Bacterial Systems; Project grant number (DE-SC008812)

## Extending functional genome annotations using high-throughput CRISPRi

Harneet S. Rishi<sup>1,2\*</sup> (HSRishi@lbl.gov), Esteban Toro,<sup>3</sup> Honglei Liu,<sup>4</sup> Xiaowo Wang,<sup>4,5</sup> Lei S. Qi,<sup>6,7</sup> and **Adam P. Arkin**<sup>2,3</sup>

<sup>1</sup>Biophysics Graduate Program, University of California Berkeley, Berkeley, CA; <sup>2</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>3</sup>Department of Bioengineering, University of California Berkeley, Berkeley, CA; <sup>4</sup>Bioinformatics Division, Center for Synthetic and Systems Biology, Tsinghua National Laboratory for Information Science and Technology, Beijing, China; <sup>5</sup>Department of Automation, Tsinghua University, Beijing, China; <sup>6</sup>Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA; and <sup>7</sup>California Institute for Quantitative Biomedical Research, San Francisco, CA

## http://genomicscience.energy.gov/biosystemsdesign/2012awards.shtml

Project Goals: Synthetic biology aims to leverage the engineering principles of modularity, standardization, and reliability with the design-build-test-learn cycle to rapidly engineer novel biological functions. One of the key hurdles in adopting this strategy is that the innate complexity of biological systems (ex: unmapped interaction networks, context dependence, temporal & spatial population variations) makes it difficult to understand first principles, which in turn makes it difficult to predictably build scalable systems. High-throughput technologies to quantitatively characterize sequence-function-phenotype landscapes can help overcome such barriers. As such, we aim to develop CRISPR interference (CRISPRi) as a platform for pooled bacterial functional genomics with the goal of investigating combinatorial genetic interactions in high-throughput.

The advent of next-generation sequencing has led to an explosion of genome sequences, and now reverse-genetics efforts are becoming increasingly vital to finding the phenotype and, by extension, function associated with a given gene of interest. As a first pass, computational workflows can predict the location of genes and regulatory features in a genome, and automated annotation pipelines can associate hypothetical functions with a given gene based on sequence homology. However, such *in silico* predictions can sometimes be erroneous – necessitating the experimental validation of predicted gene functions. Large mutant collections, which often comprise of gene disruptions in the form of knockouts or insertions, can complement computational prediction pipelines by affording the ability to experimentally interrogate many genetic perturbations in a native context.

The compactness, modularity, and largely species-independent functionality (transcriptional regulation, genome editing, imaging) of the CRISPR system make it an attractive tool for genome-wide screens. Here, we apply the catalytically inactive dCas9 to conduct high-throughput transcriptional and regulatory studies in *E. coli*. Using an Agilent OLS library of 32992 unique sgRNAs, we targeted 4500 genes, 5400 promoters, 640 transcription factor

binding sites (TFBSs), and 106 small RNAs (sRNAs) in the *E. coli* genome. These genomic targets cover a wide range of functionalities such as metabolism, stress response, transport, and cell division. By combining CRISPRi with next-generation sequencing, we were able to interrogate the fitness effect of transcriptional knockdown for each of the aforementioned genomic features in a single-pot experiment both aerobically and anaerobically. Our fitness results agreed well with current knockout databases, and our ability to induce transcriptional knockdown at any point during an experiment has allowed us to explore target essentiality under different conditions with great ease. We demonstrate this by showing that although *nrdA* and *nrdB* are essential under aerobic conditions – and are annotated as such in databases – they are dispensable anaerobically. We also show that CRISPRi can recapitulate genomic features with redundant functions, thus demonstrating our ability to interrogate combinatorial interactions. Additionally, we explore the efficacy of non-genic CRISPRi on a genome-wide scale and create cofitness profiles linking together promoters, TFBSs, and genes in transcriptional units.

We also leverage the CRISPRi library with flow cytometry to obtain non-growth phenotypes such as morphology in high-throughput. We first show that CRISPRi can generate filamentous cellular phenotypes and next present a flow-seq methodology that allows us to enrich for filamentous mutants in our library. Finally we show that our single-pot flow-seq results agree well with single-genotype microscopy studies and provide novel phenotypes for several essential genes and genes of unknown function.

Overall, HT-CRISPRi enables single-pot, precise measurements of fitness for a large set of genomic features and will prove useful in genomic studies of model and non-model organisms. The extension of the platform to perform genetic interactions should also provide a comprehensive testbed for exploring epistatic landscapes and revealing the basis of complex traits.

This work is supported by the Genome Science program within the Office of Biological and Environmental Research (Project grant number DE-SC008812, Funding Opportunity Announcement DE-FOA-0000640).

#### Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli

Marcelo C. Bassalo<sup>1,2\*</sup> (marcelo.bassalo@colorado.edu), Andrew D Garst<sup>2</sup>, Andrea L. Halweg-Edwards<sup>2</sup>, William C. Grau<sup>2,3</sup>, Dylan W. Domaille<sup>2</sup>, Vivek K. Mutalik<sup>4,5</sup>, Adam P. Arkin<sup>4,5</sup>, and **Ryan T. Gill<sup>2</sup>** 

<sup>1</sup>Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, CO 80303. <sup>2</sup>Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO 80303. <sup>3</sup>Department of Chemistry and Biochemistry, University of Colorado Boulder, Boulder, CO 80303. <sup>4</sup>Lawrence Berkeley National Laboratory, Physical Bioscience Division, Berkeley, CA 94720. <sup>5</sup>Department of Bioengineering, University of California Berkeley, Berkeley, CA 94720.

#### **Project Goal:**

The project aimed to develop a strategy that allows integration of heterologous constructs ranging in size from single genes to entire metabolic pathways at efficiencies high enough (>50%) to remove any requirement for further selection. With such strategies available, rationally designed pathways and genes can be rapidly tested in platform strains in an efficient and stable manner. Multiplex editing technologies could then be applied on top of such integrated constructs, establishing the basis for an efficient *in vivo* optimization strategy. The combination of these technologies can allow rapid engineering of strains for a broad range of biotechnology applications.

#### Abstract:

Methods for importing heterologous genes into genetically tractable hosts are among the most desired tools of synthetic biology. Many metabolic engineering applications and functional genomics studies require constructing and implementing heterologous functions from synthetic DNA. Chromosomal integration of such constructs provides better long-term stability and reduced cell-to-cell variability in copy number and expression levels, making such strains more suitable for industrial-scale processes or downstream engineering.

Genome integration methods has been available for decades in the model bacteria *Escherichia coli*, relying mostly on site-specific recombinases or the red system from the lambda phage. While gene-size constructs (~1kb) work with relative well efficiency in these systems, integration of multi-gene constructs (such as entire metabolic pathways) is a cumbersome process that requires multiple steps, selection and counter-selection strategies, and suffer from low efficiencies. By combining the genome editing tool CRISPR-Cas9 with lambda red-assisted recombination, we describe a strategy that allows highly efficient, single step integration of large pathways in *Escherichia coli*.

This strategy allows high efficiency integration in a broad range of homology arm sizes and genomic positions, with efficiencies ranging from 70 to 100% in 7 distinct loci. Using this strategy, we integrated a 10 kb construct encoding 5 genes required for production of the biofuel isobutanol in a single step, rapidly implementing production in the host strain. Further efforts from our group have demonstrated efficient integration of constructs up to 20 kb in size.

The ability to efficiently integrate entire metabolic pathways in a rapid and markerless manner will facilitate testing and engineering of novel pathways using the *E. col*i genome as a stable testing

platform. Moreover, increasingly sophisticated multiplex editing technologies could be applied on top of such integrated constructs, providing the basis for an efficient *in vivo* optimization platform.



Figure 1: Overall integration strategy combining CRISPR-Cas9 and lambda red recombination for high efficiency integration of genes and pathways.

#### **Publications:**

<sup>1</sup>Bassalo, M.C., Garst, A.D., Halweg-Edwards, A.L., Grau, W.C., Domaille, D.W., Mutalik, V.K., Arkin, A.P., Gill, R.T.. Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. ACS Synth. Biol. 2016. doi:10.1021/acssynbio.5b00187

<sup>2</sup>Garst, A.D., Bassalo, M.C., Pines, G., Lynch, S.A., Halweg-Edwards, A.L., Liu, R., Liang, L., Wang, Z., Zeitoun, R., Alexander, W.G., Gill, R.T. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat. Biotechnology 2016. doi:10.1038/nbt.3718

#### **Funding Statement:**

**Grant title:** A Platform for Genome-scale Design, Redesign, and Optimization of Bacterial Systems; Project grant number (DE-SC008812) and FWP number (ERWER44) for NREL.

## Genome-scale design and engineering approach towards optimizing ethylene production in *E. coli*

Aparna Nagarajan<sup>1,2,3</sup>, Sean Lynch<sup>1,2,3</sup>, Bradley Prythero<sup>1,2</sup>, Carrie Eckert<sup>1</sup>, Jianping Yu<sup>1</sup>, Pin-Ching Maness<sup>1</sup>, and **Ryan Gill**<sup>2,3</sup>\* (<u>rtg@colorado.edu</u>)

<sup>1</sup>Biosciences Center, National Renewable Energy Laboratory, Golden, CO; <sup>2</sup>Department of Chemical and Biological Engineering, University of Colorado, Boulder; <sup>3</sup>Renewable and Sustainable Energy Institute, NREL/University of Colorado, Boulder

## **Project Goal:**

This project aims to apply rational design and state-of-the-art synthetic and systems biology tools to optimize *E. coli* for sustained production of biofuels. Chassis biofuel strains, optimized for production based on predictive design and systems biology knowledge, will serve as the framework for high throughput genome re-design. Using targeted genome-scale and multiplex genome-engineering technologies, strains with improved production and thermal stability will be selected for further optimization. Herein, we will focus on the construction of an *E. coli* prototype chassis strain for the production of ethylene and the development of a selection strategy for gene-to-trait mapping at single nucleotide resolution to identify key factors for optimizing ethylene production. Moreover, we are developing high-throughput strategies for selection of strains with increased biofuel/precursor levels.

## Abstract

Ethylene is the most highly utilized organic compound for the production of plastics and chemicals, and its catalytic polymerization to alkane fuels has been demonstrated. At present, global ethylene production involves steam cracking of a fossil-based feedstock, representing the largest CO<sub>2</sub>-emitting process in the chemical industry. Biological ethylene production has the potential to provide a sustainable alternative while mitigating CO<sub>2</sub> emission. The expression of a single gene found in some bacteria and fungi, ethylene-forming enzyme (*efe*), can catalyze ethylene formation (1, 2). Construction of the first generation chassis strain is based on *E. coli* MG1655 as the host and the *efe* gene from *Pseudomonas syringae* (*Ps*). EFE has been postulated to catalyze ethylene production according to the equation (3):

 $3 \alpha$ -ketoglutarate + Arginine +  $3O_2 \rightarrow 2C_2H_4$  + Succinate +  $7CO_2 + 3H_2O$  + guanidine + P5C\*

The two key substrates  $\alpha$ -ketoglutarate (AKG) and arginine are tightly controlled by an intricate regulatory network that coordinates carbon and nitrogen metabolism (Figure 1). We conducted genetic modifications to rewire central carbon metabolic flux and improved ethylene production by 2.3-fold (4). This chassis strain will serve as the framework to guide genome-scale redesign and optimization to further boost ethylene production using CRISPR enabled trackable genome engineering (5). Succinate is a byproduct of the EFE reaction. We generated a succinate auxotroph in *E. coli* and showed that it must rely on an active heterologous EFE pathway yielding succinate to afford growth. EFE is not stable above 30 °C. We thereby screen for thermal stable EFE by expressing an *efe* mutant library in the succinate auxotroph and select for growth at 37 °C. We identified key mutations of *efe* mapped to semi-conserved residues in EFE homologues with its outcomes unraveling the catalytic mechanism of EFE. Work is also ongoing to construct high-throughput sensors to screen for AKG and ethylene, *in situ*. As such, current work from our groups at the National Renewable Energy Laboratory and the University of

Colorado at Boulder seeks to improve ethylene production by combining traditional metabolic engineering strategies with synthetic biology-enabled evolutionary approaches involving the high-throughput construction of genome-scale libraries. Coupled with novel screens and selections, these methods will identify strains with increased production of key intermediates and/or ethylene.

\*P5C: L- $\Delta^1$ -pyrroline-5-carboxylate



Figure 1. Putative metabolic scheme for ethylene production in *E. coli*. EFE: ethylene-forming enzyme. P5C:  $L-\Delta^{1}$ pyrroline-5-carboxylate.

## References

- 1. Fukuda, H., T. Ogawa, M. Tazaki, K. Nagahama, T. Fujii, S. Tanase, and Y. Morino. 1992. Two reactions are simultaneously catalyzed by a single enzyme, the arginine-dependent simultaneous formation of two products, ethylene and succinate, from 2-oxoglutarate by and enzyme from *Pseudomonas syringae*. Biochem. Biophys, Res. Comm. 188: 483-489.
- 2. Eckert, C., W. Xu, W. Xiong, S. Lynch, J. Ungerer, L. Tao, R. Gill, P. C. Maness, and J. Yu. 2014. Ethylene-forming enzyme and bioethylene production. Biotech. Biofuels 7:33-43.
- Fukuda, H., T. Ogawa, M. Tazaki, K. Nagahama, T. Fujii, S. Tanase, and Y. Morino.. 1992. Molecular cloning in Escherichia coli, expression and nucleotide sequence of the gene for the ethylene-forming enzyme of *Pseudomonas syringae* pv. *phaseolicola* PK2. Biochem. Biophys. Res. Comm. 826-832.
- 4. Lynch, S., C. Eckert, J. Yu, R. Gill, and P. C. Maness. 2016. Overcoming substrate limitations for improved production of ethylene in *E. coli*. Biotech. Biofuels 9:3, DOI 10.1186/s13068-015-0413-x
- Garst, A. D., Bassalo, M. C., Pines, G., Lynch, S. A., Halweg-Edwards, A. L., Liu, R., Liang, L., Wang, Z., Zeitoun, R., Alexander, W. G., Gill, R. T. 2016. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat. Biotechnol, DOI 10.1038/nbt.3718

## **Funding Statement**

**Grant title:** A Platform for Genome-scale Design, Redesign, and Optimization of Bacterial Systems; Project grant number (DE-SC008812) and FWP number (ERWER44) for NREL.

#### Title: Enabling Metaproteomics Research

Ernesto S Nakayasu<sup>1</sup>, Meagan C Burnett<sup>1</sup>, Grant M Fujimoto<sup>1</sup>, Joon-Yong Lee<sup>1</sup>, **Samuel H Payne<sup>1\*</sup> (samuel.payne@pnnl.gov)** 

1. Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA

Project Goals: This project is focused on improving algorithms and methods for mass spectrometry data analysis of metaproteomics data. Recent advances in mass spectrometry and biological separations have dramatically increased the depth of proteomic discovery. Unfortunately, traditional computational workflows are in many cases preventing researchers from realizing these benefits for microbial communities. We propose to create a new generation of computational workflows to overcome the sensitivity limitations inherent in status quo data processing schemes.

To advance our ability to annotate tandem mass spectrometry data from microbial communities, our project has been developing algorithms to match spectra from metaproteomics experiments to a library of annotated spectra. This year we made significant improvements in both the breadth of coverage for proteomics data of environmental bacteria and also the efficiency of algorithms for peptide annotation. We have expanded our proteomic coverage to 50 new organisms, focusing on organisms found in the soil and the terrestrial/aquatic interface. The availability of proteomics data for many previously unstudied organisms in these niches helps to elucidate protein functional regulation, e.g. by studying conserved post-translational modifications and conserved gene expression networks across orthologs. With the expansion of proteomics libraries growing to hundreds of species and millions of spectra, annotation algorithms face a major hurdle in computational efficiency. To overcome this, we created **a new algorithm called "FLASH"** to dramatically improve the speed of annotation. FLASH is a hybrid library assisted annotation algorithm, meaning that it leverages annotated libraries but is not limited by them. This flexibility is necessary for the diverse biomes that are sampled in metaproteomics data.

This project was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Early Career Research Program.

## Combining Deuterium-Labeling and Neutron Scattering to Gain Molecular-Level Insights Relevant to Biomass Deconstruction

Hugh O'Neill<sup>1,2\*</sup> (oneillhm@ornl.gov), Sai Venkatesh Pingali,<sup>1</sup> Riddhi Shah,<sup>1,2</sup> Barbara R. Evans,<sup>1</sup> Volker Urban,<sup>1</sup> Kevin Weiss,<sup>1</sup> Daisuke Sawada,<sup>1</sup> Loukas Petridis,<sup>1</sup> Jeremy Smith,<sup>1</sup> Paul Langan,<sup>1</sup> and **Brian Davison**<sup>1,2</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville

http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if costeffective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.

Neutrons have no charge, are highly penetrating, and do not cause radiation damage, allowing measurements of biomass structure under conditions that are relevant to thermochemical and enzymatic deconstruction. Neutrons interact with nuclei making it possible to observe the lighter elements such as hydrogen (H) and deuterium (D) and distinguish these light elements next to heavy ones. Furthermore, the neutron scattering cross-sections of H and D are very different making it possible to selectively highlight different components within a complex system. The scattering length densities of biomolecules such as lipids, proteins and DNA are inherently different which allows structural studies of complex systems or combinations of these molecules by varying the  $H_2O/D_2O$  ratio of the solvent. However, the only way to distinguish between components of a system, in which the scattering length densities are all similar, such as with proteins and carbohydrate polymers, is through the use of D-labeling techniques.

We have developed an approach to produce deuterated bacterial cellulose as a model material to investigate different aspects of biomass structure and dynamics that are relevant to biomass deconstruction [1,2]. At the atomic scale, using neutron fiber diffraction it as possible to gain insight in the structural rearrangements that occur in crystalline structure of cellulose microfibrils during conversion from cellulose I to cellulose II. At the mesoscale, small-angle neutron scattering can be used to investigate conformational changes in cellulase enzymes when bound to deuterated bacterial cellulose providing molecular-level details about the inter-domain interactions that cannot be obtained by other means. Model lignocellulose materials composed of individual matrix copolymers (e.g., hemicellulose and lignin) and deuterated bacterial cellulose interactions and the structural rearrangements that occur during pretreatment. The

aforementioned mesoscale studies take advantage of the contrast between deuterated cellulose and protiated polymers or proteins to distinguish changes in the individual components. The dynamical properties of water when bound to cellulose can be probed using quasi-elastic neutron scattering to further enhance our understanding of water's role and how it partitions in the cellulose supramolecular structure, potentially leading to more efficient pretreatment approaches. The presentation will provide specific details, from the examples above, of how the combination of biodeuteration with neutron scattering can enhance our knowledge of the underlying processes that change biomass morphology during different pretreatment regimes for biofuels production.

## References

- 1. Bali, G., M. B. Foston, H. M. O'Neill, B. R. Evans, J. H. He and A. J. Ragauskas (2013). "The effect of deuteration on the structure of bacterial cellulose." *Carbohydrate Research* **374**: 82-88.
- He, J. H., S. V. Pingali, S. P. S. Chundawat, A. Pack, A. D. Jones, P. Langan, B. H. Davison, V. Urban, B. Evans and H. O'Neill (2014). "Controlled incorporation of deuterium into bacterial cellulose." *Cellulose* 21(2): 927-936.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-000R22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

## Conformations of Low-Molecular Weight Lignin in Water: Insights on the Chemical Character of Lignin

Loukas Petridis<sup>1</sup>\* (petridisl@ornl.gov), Sai Venkatesh Pingali,<sup>1</sup> Hugh M. O'Neill,<sup>1</sup> Yunqiao Pu,<sup>1</sup> Volker Urban,<sup>1</sup> Paul Langan,<sup>1</sup> Riddhi Shah,<sup>1,2</sup> Samarthya Bhagia,<sup>2</sup> Arthur J. Ragauskas,<sup>1,2</sup> Barbara R. Evans,<sup>1</sup> Jeremy C. Smith,<sup>1,2</sup> and **Brian H. Davison**<sup>1</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>University of Tennessee, Knoxville

http://cmb.ornl.gov/research/bioenergy/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if costeffective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that are necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.



Representative structures of seven lignin polymers of degree of polymerization N at two temperatures.

Low-molecular weight lignin binds to the cellulose during thermochemical pretreatment of biomass for biofuel production, preventing the efficient hydrolysis of the cellulose to sugars. The binding properties of lignin are strongly influenced by the conformations it adopts. Here we use molecular dynamics simulations in aqueous solution to investigate the dependence on chain length and temperature of the shape of lignin polymers. Lignin is found to adopt collapsed conformations in water at temperatures of 300 K and 500 K. However, at 300 K, a discontinuous transition is found in the shape of the polymer as a function of chain length. Below a critical

degree of polymerization,  $N_c = 15$ , the polymer adopts less spherical conformations than above  $N_c$ . The transition disappears at high temperatures (500 K) at which only spherical shapes are adopted. An implication relevant to cellulosic biofuel production is that lignin will self-aggregate even at high pretreatment temperatures. The above picture is consistent with small-angle neutron scattering experiments showing the building blocks of extracted lignin being oligomeric, with N~5-12 and adopting aspherical, rod-like conformations.

The study sheds light on the physico-chemical character of lignin. In general, the conformational properties of polymers in dilute solution are influenced by a balance of interactions between monomers and with the solvent. Block copolymers composed of both hydrophilic and hydrophobic monomers self-assemble in a manner that minimizes unfavorable interfacial areas (hydrophilic-hydrophobic and water-hydrophobic) while maximizing favorable interfaces (hydrophilic-hydrophilic, hydrophobic-hydrophobic and water-hydrophilic), subject to chain connectivity constraints. These polymers are able to do so because their domains consist of multiple monomers that are large enough in number that they can phase separate. However, this phase separation is not possible for "poly-amphiphiles", complex polymers that contain both polar and non-polar groups in their monomers, such as lignin. Lignin monomers themselves contain both hydrophobic due to their dominant aromatic rings, they also contain polar hydroxyl groups that interact favorably with water. The present simulations show that the balance of interactions between lignin and water leads to a chain-length dependent shape of the collapsed polymer in water instead of the microphase separation found in block copolymers.

#### References

- 1. Petridis, L.; Smith, J. C., Conformations of Low-Molecular-Weight Lignin Polymers in Water. *ChemSusChem* **2016**, *9* (3), 289-295.
- 2. Ratnaweera, D. R.; Saha, D.; Pingali, S. V.; Labbe, N.; Naskar, A. K.; Dadmun, M., The impact of lignin source on its self-assembly in solution. *RSC Adv.* **2015**, *5* (82), 67258-67266.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

# Deuteration Effects on Switchgrass Structure and Metabolism: Lignin Deposition Changes in Cell Walls of Deuterated Switchgrass

Samarthya Bhagia<sup>1</sup>\*(sbhagia@utk.edu), Xianzhi Meng,<sup>1</sup> John R. Dunlap,<sup>1</sup> Barbara Evans,<sup>2</sup> Garima Bali,<sup>3</sup> Jihua Chen,<sup>2</sup> Kimberly Shawn Reeves,<sup>2</sup> Hoi Chun Ho,<sup>2</sup> Arthur J. Ragauskas,<sup>1,2</sup> and **Brian H. Davison**<sup>2</sup>

<sup>1</sup>University of Tennessee, Knoxville; <sup>2</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>Georgia Institute of Technology, Atlanta, Georgia

http://cmb.ornl.gov/research/neutron-scattering/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if costeffective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that are necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.



Differences in TEM image of protiated and deuterated switchgrass. Images (A) and (B) show protiated and deuterated switchgrass cell walls, respectively. Arrow heads show regions of higher lignification. Long arrow shows lignin-like deposit in image of deuterated switchgrass. Short arrow shows convoluted cell walls of deuterated switchgrass with reduced lignification compared to a cell wall with normal lignification on the left-side in the same image (B). Scale bar is  $0.5 \,\mu\text{m}$ .

Neutron scattering enables the study of structural and dynamic properties of lignocellulosic biomass at multiple length scales in a non-destructive manner. High levels of deuterium substitution are desired to enable contrast variation techniques which reduce background. However, plants grown in concentrations of D<sub>2</sub>O greater than 30% typically exhibit slower growth, stunting, inhibition of root elongation, and delay or abolishment of flowering and seed

set. To evaluate utility of deuterated lignocellulosic biomass for structural studies, any changes in lignocellulosic cell wall composition and molecular structure need to be determined. In contrast to the response of other plants including C<sub>3</sub> annual grasses *Triticum*, *Secale* and *Lolium*, the bioenergy crop switchgrass *Panicum virgatum*, a C<sub>4</sub> drought-resistant perennial prairie grass, has been shown to adapt to growth in 50% D<sub>2</sub>O. Switchgrass grown hydroponically from tiller cuttings and transferred to 50% D<sub>2</sub>O growth solution after rooting maintains growth rates, gross morphology, and cellular appearance, with 35% deuterium incorporation.<sup>1</sup> Deuterated switchgrass plants continue to grow for at least two years, grow new roots as well as tillers, and even produce reproductive tillers with normal appearing flower spikes.

However, enzymatic hydrolysis and microscopy identified ultrastructural changes in properties of deuterated switchgrass biomass with direct relevance to understanding recalcitrance and response to abiotic stressors.<sup>2</sup> In this study, enzymatic hydrolysis of bacterial cellulose, and cellulose and holocellulose isolated from switchgrass showed the expected kinetic isotope effect, with deuteration lowering glucose yields by 17, 18 and 4% of theoretical yield, respectively. However, the opposite trend was found for deuterated switchgrass, in which glucose yield was 5% higher than that obtained with protiated switchgrass at lower enzyme loading. These data indicated that alterations to lignin might be responsible for this novel inverse isotope effect. Lignin content of deuterated switchgrass was about 2% higher than that of protiated switchgrass, while nuclear magnetic resonance (NMR) spectroscopy indicated no significant compositional or structural differences in lignin, and Simons' staining found comparable cellulose surface area. However, extensive confocal fluorescence microscopy (CFM) and transmission electron microscopy (TEM) imaging showed that deuterated switchgrass had abnormal lignin distribution in some of its cell walls and many of them were collapsed possibly due to reduced rigidity, which would render them easier to deconstruct by cellulases (Fig. 1). The changes in morphology resembled those of drought stressed plants, consistent with abiotic stress due to growth in deuterated media. These protio/deutero investigations clearly illustrate the key role of component distribution in the multilamellar cell wall architecture on biomass recalcitrance.

## References

- Evans, B. R., Bali, G., Foston, M., Ragauskas, A. J., O'Neill, H., Shah, R., McGaughey, J., Reeves, D., Rempe, C. S., Davison, B. H. (2015) Planta 242, 215 – 222, "Production of deuterated switchgrass by hydroponic cultivation."
- 2. Bhagia, S., Meng, X., Evans, B. R., Chen, J., Reeves, K. S., Ho, H. C., Davison, b. H. and Ragauskas, A. J. "Novel Inverse Isotope Effect for Cellulase Hydrolysis of Deuterated Switchgrass" (submitted).

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

#### Real-time Elucidation of Structure and Morphology of Native and Mutant Poplar During Dilute Acid and Alkali Pretreatments Using Neutron Scattering

Sai Venkatesh Pingali<sup>1</sup>\* (pingalis@ornl.gov), Riddhi Shah,<sup>1,2</sup> Shawn Mansfield,<sup>3</sup> Volker Urban,<sup>1</sup> Paul Langan,<sup>1</sup> Hugh O'Neill<sup>1,2</sup> and **Brian H. Davison**,<sup>1,2</sup>

<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>University of Tennessee, Knoxville; <sup>3</sup>University of British Columbia, Vancouver, Canada

http://cmb.ornl.gov/research/neutron-scattering/lignocellulose-dynamics

Project Goals: Lignocellulosic biomass comprises the vast majority of biomass on Earth and has the potential to play a major role in generation of renewable biofuels if costeffective conversion can be achieved. Largely composed of plant cell walls, it is a complex biological composite material that is recalcitrant to the structural deconstruction and enzymatic hydrolysis into sugars that is necessary for fermentation to bioethanol. The Scientific Focus Area in Biofuels is developing "Dynamic Visualization of Lignocellulose Degradation by Integration of Neutron Scattering Imaging and Computer Simulation" for multiple-length scale, real-time imaging of biomass during pretreatment and enzymatic hydrolysis. This is providing fundamental information about the structure and deconstruction of plant cell walls that is needed to drive improvements in the conversion of renewable lignocellulosic biomass to biofuels.



Dilute acid reactions of native and mutant poplar wood were examined by real-time smallangle neutron scattering. Time resolved curves for native (top left) and mutant (bottom left) poplar samples; real-time pressure reaction cell (right top); reaction temperature profiles (right bottom).

Biomass deconstruction by thermochemical pretreatments, a precursor to enzymatic hydrolysis, is intended to enhance enzymatic accessibility to cellulose. However, the complex nature of biomass causes initiation of several processes at different stages of the pretreatment regime. These processes cause structural rearrangements in cellulose and the matrix copolymers (lignin, hemicellulose, etc.,), as well as to mesoscale surface morphology. Knowledge of structural

changes to biomass and the chronological order in which they appear during pretreatment will enable an informed approach to improving biomass conversion efficiencies.<sup>1,2</sup> Our recently published *ex-situ* results indicated that enzymatic hydrolysis of switchgrass produced higher yields despite dilute acid pretreatment causing higher degree of coalescence of neighboring cellulose microfibrils, increased cellulose crystallinity, and higher redistribution of lignin leading to larger aggregates.<sup>2</sup> Yet current conventional wisdom would suggest that such structural rearrangements should lead to reduced enzymatic yields. A plausible reason for such observation could be in our inability to observe structural changes occurring in-real time.

Consequently, we developed a pressure reaction cell to monitor morphological changes in biomass using small-angle neutron scattering (SANS) during thermochemical pretreatment in real-time.<sup>3,4</sup> This approach takes advantage of the non-destructive and high penetration properties of neutrons to perform *in-situ* studies. Most importantly, we have recently extended the capabilities of the Bio-SANS instrument at the High Flux Isotope Reactor at Oak Ridge National Laboratory by installation of an additional detector array; this upgrade makes it possible to capture neutrons scattered at higher angles simultaneously to the neutrons captured at smaller scattering angles by the main detector.<sup>4</sup> As a result, data collection times are dramatically reduced and it is possible to obtain structural information at shorter length scales than was previously possible. Here, we report on real-time SANS studies of native poplar and a transgenic variant that is deficient in lignin synthesis during dilute acid and alkali pretreatments. We observe significant differences in lignin aggregation patterns in the native and mutant poplar comparing the two pretreatment regimes. On the other hand, the scattering signature assigned to the cellulose microfibrils remains unchanged in dilute acid and alkali pretreatment. A quantitative analysis of the structural changes in the native and mutant plants will be presented. Using this approach, it is possible to obtain molecular level insights into structural rearrangements of biomass polymers during pretreatment to better understand the consequences of genetic mutations on the overall digestibility of biomass.

## Reference

- 1. S.V. Pingali et al., Breakdown of cell wall nanostructure of dilute acid pretreated switchgrass. *Biomacromolecules* **2010**, 11, 2329-2335.
- 2. S.V. Pingali et al., Understanding multiscale structural changes during dilute acid pretreatment of switchgrass and poplar. *ACS Sustainable Chemistry & Engineering* **2016**, (accepted) DOI:10.1021/acssuschemeng.6b01803.
- 3. S.V. Pingali et al., Morphological changes in the cellulose and lignin components of biomass occur at different stages during steam pretreatment. *Cellulose* **2014**, 21, 873-878.
- 4. The development of cell and detector upgrade were partially and fully funded by the Oak Ridge National Laboratory's Center for Structural Molecular Biology (CSMB), which is supported by the Office of Biological and Environmental Research, using facilities supported by the U.S. Department of Energy, managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. This program is supported by the Office of Biological and Environmental Research in the DOE Office of Science.

# Species-specific evolution of membrane-bound receptors mediating host-symbiont specificity in the genus *Salix*

Kate Stuart,<sup>1</sup> Timothy B. Yates, <sup>2</sup> Jin Zhang,<sup>1</sup> Stephen P. DiFazio, <sup>3</sup> Lawrence Smart, <sup>4</sup> and Wellington Muchero<sup>1</sup>\* (mucherow@ornl.gov)

<sup>1</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>2</sup>Bredesen Center for Interdisciplinary Research, University of Tennessee, Knoxville; Tennessee; <sup>3</sup>Biology Department, West Virginia University, Morgantown, West Virginia; <sup>4</sup>School of Integrative Plant Science, Cornell University, Ithaca, New York

Project Goals: The goal of this project is to identify and characterize species-specific Pattern Recognition Receptors (PRRs) that are involved in host immune suppression to allow successful colonization by microbial symbionts. Using *Salix*, a dominant species in the warming arctic region and a widely used biofuels feedstocks, we leverage strong species-specific colonization by endophytic microbes to create hybrid populations segregating for colonization efficiency. Complementary characterization of endosphere microbial communities, mapping genomic structural polymorphisms involving PRRs across segregating hybrid *Salix* populations and correlations of these features will enable assignment of species-specific PRRs to their microbial targets. Output from this project will inform efforts to rationally engineer plant-microbe interactions between previous unassociated host plants and microbial partners.

**Abstract:** Innate host defense poses a fundamental challenge in the utilization of beneficial microbes for improving plant productivity and carbon sequestration for applications in biofuels feedstock production and global warming mitigation. As an evolutionary adaptive mechanism to recruit symbiotic microbes, plants evolved membrane-bound Pattern Recognition Receptors (PRRs) that, upon recognition of microbe-associated molecular patterns (MAMPs) of beneficial microbes, will trigger molecular signals to suppress activation of the host defense machinery. In this study, we sought to establish species divergence in genomic composition of these PRRs as well as root endophytic microbial communities in two *Salix* species, *S. purpurea* and *S. suchowensis*. To assess putative species-level divergence, we sampled two genotypes to represent each species, P63 and P295 (*S. suchowensis*) and 94006 and 94001 (*S. purpurea*) from a replicated field trial in Morgantown, West Virginia. Differentiation in genomic PRR composition and their expression under field conditions was assessed using long-read genomic DNA and transcriptome sequencing. Microbial community divergence was established using 16S phylotyping. Results of host-species divergence in these two characteristics will be presented.

**Funding:** This project is funded by the Department of Energy's Office of Science Early Career Research Program in the Biological and Environmental Research Program Office.

## Chemical Analysis of Carbon and Nitrogen Cycling Through the Extracellular Matrix Produced During the Formation of a Multi-Species Community

Matt Marshall<sup>1</sup>\* (matthew.marshall@pnnl.gov), Ryan Renslow,<sup>1</sup> Abigail Tucker,<sup>1</sup> Mary Lipton,<sup>1</sup> Aaron T. Wright,<sup>1</sup> Jim Fredrickson,<sup>1</sup>, and Lee Ann McCue<sup>1</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, WA

http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The goal of the Metabolic and Spatial Interactions in Communities (MOSAIC) Foundational Scientific Focus Area is to understand the fundamental mechanisms by which microbial metabolic interactions and spatial organization impact carbon, nitrogen, and energy dynamics in microbial communities. Our studies focus on the coupling of carbon and nitrogen cycles in microbial communities, the role of environmental variables in governing the rates of these cycles, and the impact of environmental perturbations on microbial community dynamics. We employ tractable model consortia whose member genome sequences have been defined, advanced omics measurements, functional imaging, taxonomic profiling, and modeling to elucidate interaction mechanisms within complex microbial communities. Our research supports the DOE goals to achieve a predictive understanding of Earth's integrated biogeochemical processes.

Microbial autotroph-heterotroph interactions influence biogeochemical cycles on a global scale, but the molecular mechanisms underlying microbial community interactions and functional processes for which microorganisms cycle carbon, nitrogen and energy are largely unknown. We hypothesize that the extracellular polymeric substance (EPS) matrix is an exchangeable resource that can serve as a complex and dynamic pool of biologically-available carbon, nitrogen, and other nutrients during nutrient limitation or other types of environmental stress. We have investigated the EPS matrix using a tractable, autotroph-heterotroph consortia. These stable consortia cultures, containing one autotrophic cyanobacteria and multiple heterotrophic bacteria, have been previously isolated and temporal community member dynamics have been established.

To test our hypothesis, we developed an extraction protocol to isolate distinct operational fractions of EPS from the communities, and performed a temporal study of the EPS matrix to determine changes in the distribution of biological molecules within the EPS fractions during community maturation. The chemical components of the purified EPS fractions were evaluated using attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy analysis to determine the relative abundances of lipids, protein, sugars, and nucleic acids in EPS fractions. Principle component analysis was used to resolve temporal- and EPS fraction- dependent changes in these biological molecules that were indicative of carbon and nitrogen cycling in the EPS matrix. Temporal changes were observed in the distributions of lipids observed in EPS, while the relative abundances of sugars remained stable over the study and possibly suggested that the extracellular carbon in nascent polysaccharides was being quickly consumed by

heterotrophic community members. Interestingly, we observed low abundances of proteins and nucleic acids in all EPS fractions. The low abundance of extracellular proteins suggested that they were quickly degraded and consumed by the community, indicating that nitrogen cycling (as nitrate) through the community was also, in part, dependent on autotrophic metabolism. To identify the community members that are secreting EPS degrading enzymes, we are using activity-based probes that are selective towards functionally active glycosyl hydrolase and serine and cysteine protease activities, in conjunction with mass spectrometry-based proteomics, to directly identify and quantify the functionally-active extracellular EPS degrading enzymes. This work enhances our understanding for how a microbial community cycles carbon, nitrogen and energy through a community through the EPS matrix and establishes a foundation for understanding carbon and nitrogen cycles through other autotroph-heterotroph systems and microbial communities found in complex physical systems such as soils.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

## High-Resolution Spatial Analysis Reveals How Nitrogen Source Governs Carbon Partitioning Between Members in a Phototrophic Consortium

Christopher R. Anderton<sup>1</sup>\* (christopher.anderton@pnnl.gov), Ryan S. Renslow,<sup>1</sup> Jennifer M. Mobberley,<sup>1</sup> William B. Chrisler,<sup>1</sup> Jessica K. Cole,<sup>1</sup> Amy A. Boaro,<sup>1</sup> Jamie Nuñez,<sup>1</sup> Yasemin Yesiltepe,<sup>1</sup> Beau Morton,<sup>1</sup> Alexandra Cory,<sup>1</sup> Hayley Cardamone,<sup>1</sup> Mary S. Lipton,<sup>1</sup> James Moran,<sup>1</sup> Stephen R. Lindemann,<sup>2</sup> Jim Fredrickson,<sup>1</sup> and Lee Ann McCue<sup>1</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, WA; <sup>2</sup>Purdue University, West Lafayette, IN

## http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The goal of the Metabolic and Spatial Interactions in Communities (MOSAIC) Foundational Scientific Focus Area is to understand the fundamental mechanisms by which microbial metabolic interactions and spatial organization impact carbon, nitrogen, and energy dynamics in microbial communities. Our studies focus on the coupling of carbon and nitrogen cycles in microbial communities, the role of environmental variables in governing the rates of these cycles, and the impact of environmental perturbations on microbial community dynamics. We employ tractable model consortia whose member genome sequences have been defined, advanced omics measurements, functional imaging, taxonomic profiling, and modeling to elucidate interaction mechanisms within complex microbial communities. Our research supports the DOE goals to achieve a predictive understanding of Earth's integrated biogeochemical processes.

Interactions between microbial photoautotrophs and associated heterotrophic organisms are ubiquitous in nature and exert significant impacts on global biogeochemical cycling. Hence, elucidating the molecular mechanisms by which environmental conditions impact community interactions in phototrophic-heterotrophic consortia is critical to predicting how they will respond to environmental change.<sup>1,2</sup> Here, we used a model consortium containing one photoautotrophic cyanobacterium (Phormidium sp. OSCR) and 18 associated heterotrophic species, for which a species-resolved metagenome reconstruction is available,<sup>3</sup> to understand fundamental carbon and nitrogen coupling in microbial communities. We hypothesized that the genome-predicted inability of most heterotrophic consortium members to directly acquire NO<sub>3</sub> from the medium would result in widespread nitrogen limitation for these species. Altering the available nitrogen source  $(NO_3^- vs. NO_3^- + NH_4^+)$  drastically altered community composition and dynamics, though not in ways easily predictable from members' ability to assimilate  $NO_3^{-1}$ . Examination of the consortia, using a multimodal mass spectrometry-based stable isotope probing approach, revealed that although  $NH_4^+$  was acquired approximately twice as rapidly as  $NO_3^-$ , there was no significant difference in community  $H^{13}CO_3^-$  incorporation. To resolve whether C/N partitioning among members was altered, we used high-lateral resolution mass spectrometry imaging (NanoSIMS), in conjunction with our previously developed image processing pipeline<sup>4</sup>. These data showed significant differences in the rate at which heterotrophs acquired cyanobacterially-fixed carbon, where carbon was much more rapidly transferred to heterotrophs when  $NO_3^-$  was the sole nitrogen source than when  $NH_4^+$  was also available. Relating these results to that from proteomic analysis of the cyanobacteria indicates altered iron acquisition and pyruvate metabolism based on the available nitrogen source, suggesting that carbon transfer rate may be increased when  $NO_3^-$  is the sole nitrogen source as the cyanobacteria are discarding excess reductant. Fluorescence *in situ* hybridization (FISH) imaging with speciesspecific probes provided insight into the potential role of select heterotrophs within the community. We observed that the most abundant heterotroph (*Aliidiomarina calidilacus*), independent of nitrogen source, was always proximal to the cyanobacterial filaments, whereas the heterotroph (*Algoriphagus marincola*) whose abundance was significantly reduced upon  $NH_4^+$  amendment co-localized with lysed cyanobacterial cells. These results suggest that *A. calidilacus* cells form an epibiotic or parasitic relationship with the cyanobacteria, and that *A. marincola* cells are detritivores that recycle proteinaceous biomass.

#### References

(1) Cole, J.; Hutchison, J.; Renslow, R.; Kim, Y.-M.; Chrisler, W.; Engelmann, H.; Dohnalkova, A.; Hu, D.; Metz, T.; Fredrickson, J.; Lindemann, S. *Frontiers in Microbiology* **2014**, *5*.

(2) Lindemann, S.; Moran, J.; Stegen, J.; Renslow, R.; Hutchison, J.; Cole, J.; Dohnalkova, A.; Tremblay, J.;

Singh, K.; Malfatti, S.; Chen, F.; Tringe, S.; Beyenal, H.; Fredrickson, J. Frontiers in Microbiology 2013, 4.
(3) Nelson, W. C.; Maezato, Y.; Wu, Y.-W.; Romine, M. F.; Lindemann, S. R. Applied and Environmental Microbiology 2016, 82, 255-267.

(4) Renslow, R. S.; Lindemann, S. R.; Cole, J. K.; Zhu, Z.; Anderton, C. R. Biointerphases 2016, 11, 02A322.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. The work conducted by the U.S. Department of Energy Joint Genome Institute (JGI) was supported by the Office of Science of DOE under Contract No. DE-AC02-05CH11231 and Community Sequencing Project 701.

#### **Metabolite Characterization in Complex Microbial Communities**

Ryan S. Renslow<sup>1</sup>\* (ryan.renslow@pnnl.gov), Erin S. Baker,<sup>1</sup> Thomas O. Metz,<sup>1</sup> Aaron T. Wright,<sup>1</sup> and Lee Ann McCue<sup>1</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, WA

http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The goal of the Metabolic and Spatial Interactions in Communities (MOSAIC) Foundational Scientific Focus Area is to understand the fundamental mechanisms by which microbial metabolic interactions and spatial organization impact carbon, nitrogen, and energy dynamics in microbial communities. Our studies focus on the coupling of carbon and nitrogen cycles in microbial communities, the role of environmental variables in governing the rates of these cycles, and the impact of environmental perturbations on microbial community dynamics. We employ tractable model consortia whose member genome sequences have been defined, advanced omics measurements, functional imaging, taxonomic profiling, and modeling to elucidate interaction mechanisms within complex microbial communities. Our research supports the DOE goals to achieve a predictive understanding of Earth's integrated biogeochemical processes.

Microbial communities are present in, and interact with, a wide variety of challenging ecosystems, such as soil. The members of these communities play vital roles in the cycling of carbon and nitrogen and for sustaining ecosystem productivity and overall health; however, the molecular bases for how community members interact with each other and their environment is yet to be fully elucidated. We have developed a suite of integrated molecular measurement capabilities for the comprehensive characterization of metabolites and other soil organic matter and for characterizing the mechanisms by which soil microbial communities interact with this matter. These capabilities include MetFish, and ultra-high throughput ion mobility spectrometry-mass spectrometry (IMS-MS) measurements combined with the associated physical-chemical property of collisional cross section (CCS).

MetFish is a suite of chemoselective probes that target metabolites and other small molecules containing amine, carboxyl, carbonyl, and hydroxyl functional groups, to enable their enrichment from complex and extreme ecosystems, such as soil and hypersaline systems. The MetFish workflow provides sensitive and specific quantification of metabolites, with limits of quantification at low nM levels in sample matrices containing up to 2 M total dissolved salts. We have applied MetFish in analyses of microbial consortia in hypersaline media, as well as soil and fracking fluid. For soil metabolomics, MetFish in combination with a high salt wash provided higher recovery (2-10 fold) of extracellular metabolites in comparison to water and low salt extractions, allowing for more precise quantification of extracellular metabolites while imparting minimal to no effects on microbial intracellular metabolite profiles.

We have also developed an automated solid phase extraction (SPE) coupled with IMS-MS metabolomics platform capable of performing both targeted and global measurements with high reproducibility ( $CV \le 3\%$ ), sensitivity and throughput (>8000 analyses per day). This combination of separation techniques allows in-depth metabolite characterization due to the fact that the SPE distinguishes each molecule based on chemical characteristics, IMS by structure, and MS by mass. To facilitate broad metabolite identification using the SPE-IMS-MS platform, we are creating a reference library of accurate mass and molecular CCS through experimental measurements and theoretical predictions based on chemical structures of the molecules. CCS is the key structural property measured by IMS-MS, and can be calculated accurately *in silico*, which helps overcome longstanding obstacles to metabolite identification in the absence of authentic chemical standards. We employ a custom quantum chemistry-based, supercomputerdriven software engine for calculating CCS in which chemical identifiers (i.e., InChIs) are converted into 2D structures, and then protonation/deprotonation states and adduct sites are predicted. Final geometry optimizations are performed using a density functional theory implemented in the NWChem quantum chemistry software developed at the DOE EMSL user facility. The generation of accurate predicted CCS values can facilitate the broad identification of detected molecules in combination with accurate mass and MS/MS spectra.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area. Additional support for IMS-MS platform and metabolite reference library construction was provided by the PNNL Laboratory Directed Research and Development Program by the Microbiomes in Transition Initiative and the National Institutes of Health, National Institute of Environmental Health Sciences through grant R01ES022190. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

#### Modeling Approaches for Understanding Metabolic Coupling in Microbial Communities

Hyun-Seob Song<sup>1</sup>\* (hyunseob.song@pnnl.gov), Hans C. Bernstein,<sup>1</sup> Christopher S. Henry,<sup>2</sup> Pamela B. Weisenhorn,<sup>2</sup> Ronald C. Taylor,<sup>1</sup> Joon-Yong Lee,<sup>1</sup> Jeremy D. Zucker,<sup>1</sup> James Moran,<sup>1</sup> Jim K. Fredrickson,<sup>1</sup> and Lee Ann McCue<sup>1</sup>

<sup>1</sup>Pacific Northwest National Laboratory, Richland, WA; <sup>2</sup>Argonne National Laboratory, Argonne, IL

http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The goal of the Metabolic and Spatial Interactions in Communities (MOSAIC) Foundational Scientific Focus Area is to understand the fundamental mechanisms by which microbial metabolic interactions and spatial organization impact carbon, nitrogen, and energy dynamics in microbial communities. Our studies focus on the coupling of carbon and nitrogen cycles in microbial communities, the role of environmental variables in governing the rates of these cycles, and the impact of environmental perturbations on microbial community dynamics. We employ tractable model consortia whose member genome sequences have been defined, advanced omics measurements, functional imaging, taxonomic profiling, and modeling to elucidate interaction mechanisms within complex microbial communities. Our research supports the DOE goals to achieve a predictive understanding of Earth's integrated biogeochemical processes.

Metabolic interactions within microbial communities exert a strong impact on nutrient cycling in the environment. Metabolic network models of microbial communities can be used to deepen our understanding of the drivers controlling interspecies metabolic coupling and how these interactions create emergent properties such as resilience and resistance, and to predict the interplay between microbial communities and their environment. In general, genome-scale metabolic network reconstruction is an iterative process in which community-level reconstructions are significantly more complex than single organism models, due to the need to account for interspecies interactions. Current modeling approaches focus on the reconstruction of high-quality individual networks followed by combining these individual networks to facilitate prediction of interspecies interactions and community behavior. However, this approach is challenging when exploring complex environmental communities containing members not characterized in isolation. To address this limitation, we tested a novel approach that leverages community (i.e., mixed culture) data as a critical input for network reconstruction. The community data includes phenotypic growth observations and multi-omics profiles, which provide key information on microbial interactions that are not necessarily obtainable from axenic cultures.

As a case study used to evaluate our method, we considered a photoautotroph-heterotroph binary consortium and reconstructed community metabolic networks based on alternative strategies of gapfilling: individual gapfilling (or pre-gapfilling) vs. community-level gapfilling (or post-gapfilling). We implemented all pipeline steps of network reconstruction and refinement using the DOE Systems Biology Knowledgebase (KBase) platform (www.kbase.us). Post-gapfilled metabolic networks provided predictions of organic carbon and nitrogen exchange that supported the growth of an obligate heterotrophic species, with these predictions being supported by transcriptomic data. The community network model provides a platform to perform flux

coupling analysis, identify metabolically coupled reactions across species, and evaluate associated metabolic costs (e.g., in terms of ATP consumption). Reproducible narratives of model building for both single species and community networks are publicly available in KBase (https://narrative.kbase.us/narrative/ws.13807.obj.1).

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.
## Newly Identified Regulatory Roles for Vitamin B<sub>12</sub> Suggest Coordination of Community Metabolism

Aaron T. Wright,<sup>1</sup> William C. Nelson,<sup>1</sup> Lindsey N. Anderson,<sup>1</sup> Joshua J. Rosnow,<sup>1</sup> Bryan Killinger,<sup>1</sup> and Lee Ann McCue<sup>1</sup>\* (leeann.mccue@pnnl.gov)

<sup>1</sup>Pacific Northwest National Laboratory, Richland, WA

### http://www.pnnl.gov/biology/programs/fsfa/

Project Goals: The goal of the Metabolic and Spatial Interactions in Communities (MOSAIC) Foundational Scientific Focus Area is to understand the fundamental mechanisms by which microbial metabolic interactions and spatial organization impact carbon, nitrogen, and energy dynamics in microbial communities. Our studies focus on the coupling of carbon and nitrogen cycles in microbial communities, the role of environmental variables in governing the rates of these cycles, and the impact of environmental perturbations on microbial community dynamics. We employ tractable model consortia whose member genome sequences have been defined, advanced omics measurements, functional imaging, taxonomic profiling, and modeling to elucidate interaction mechanisms within complex microbial communities. Our research supports the DOE goals to achieve a predictive understanding of Earth's integrated biogeochemical processes.

Individual members of complex microbial communities can interact by exchanging metabolites and signaling compounds. Using genome-informed approaches coupled to proteomics, chemical probing, and other molecular biology studies, we have predicted and tested B vitamin dependencies in a model autotroph-heterotroph community. In particular, our findings suggest requirements for vitamin  $B_{12}$  exchange to support community growth and to regulate microbial metabolism.

Vitamin  $B_{12}$  encompasses a group of closely related corrinoid compounds best known for their role as cofactors of enzymes, and also for playing regulatory roles by binding to riboswitches. Recent studies have also revealed that vitamin  $B_{12}$  is a cofactor of transcriptional regulators. A portion of vitamin  $B_{12}$  is photolabile, thereby when bound to these proteins it provides light-dependent regulation of transcription, controlling processes such as biosynthesis of carotenoids, tetrapyrroles, and photosystems. To identify new roles for  $B_{12}$ , we developed a chemical probe that mimics the natural vitamin, in fact to such an extent as to support microbial growth, transcription, and translation. This probe was deployed in individual bacterial isolates, and within a photoautotroph-heterotroph community. Coupling the probe to proteomics analyses, we identified more than 50  $B_{12}$ -binding proteins including enzymes known to use it as a cofactor, a transcriptional regulator, enzymes in the one carbon pool by folate pathway, and enzymes involved in ubiquinone biosynthesis. Importantly, we also identify a  $B_{12}$ -dependent role for MetE (methionine synthase) in the photoautotroph when the organism limits vitamin salvage. The unexpected discovery of  $B_{12}$  involvement in these processes suggests a pivotal role in the control

of cell growth, potentially leading to coordination of cell behavior in complex multicellular systems. We predict that these roles for  $B_{12}$  may be generalizable in myriad communities, a possibility that we are now exploring.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.

### Multi-Omics Analysis of a Mycorrhizal System

Peter E. Larsen,<sup>1,2\*</sup> (plarsen@anl.gov), Shalaka Shinde,<sup>1</sup> Phil Laible,<sup>1</sup> Sarah Zerbs,<sup>1</sup> Peter Korajczyk,<sup>1</sup> and **Philippe Noirot**<sup>1</sup>

<sup>1</sup>Argonne National Laboratory, Argonne IL; and <sup>2</sup>University of Illinois at Chicago, Chicago IL

**Project Goals:** The goal of the Environmental Sensing and Response SFA is to decipher the molecular dialog between rhizobacteria, plant roots, and ECM fungi that leads to the establishment of symbiotic interactions and beneficial effects for the plant. The mechanisms for environmental sensing and response between rhizosphere community members span multiple types of biological interactions that can be revealed by using an integration of metabolomics, proteomics, and transcriptomics data through computational analysis. Analysis of multi-omics data proposes specific molecular mechanisms of mycorrhizal interactions that can be directly validated through biological experiments.

In terrestrial ecosystems, plants are never solitary entities. Rather, they exist as complex metaorganisms, comprised of plant hosts and communities of interacting soil microorganisms, which include mutualistic fungi and bacteria. Symbiosis between soil bacteria, mycorrhizal fungi, and tree species leads to coordinated resource exchange and enhanced productivity and resiliency in forest ecosystems. These symbiotic associations provide multiple benefits to the host tree, especially under conditions of limiting nutrient availability. In return, the rhizosphere microorganisms acquire photosynthetically-derived carbon from the plant in form of sugars and organic acids. There are significant knowledge gaps in our understanding of how rhizosphere communities establish and maintain symbiotic interactions that are beneficial to the plant host. We address these knowledge gaps using a laboratory model of mycorrhizal symbiosis between P. tremuloides (aspen) and the ectomycorrhizal fungi Laccaria bicolor (Laccaria) and Paxillus involutus (Paxillus), multiomics analysis, and machine learning approaches. Aspen seedlings were grown in sand pots, alone or in mycorrhizal association with either Laccaria or Paxillus. Root transcriptomic, community metabolomic, and root membrane proteomic data were collected, through collaborations with JGI and EMSL. Data was collected for each of the 'omics data types and integrated into a system-scale model of mycorrhizal interactions.

- *Transcriptomics*: In an analysis of aspen root transcriptomic data, about 60% of short reads have multiple possible alignments to *P. trichocarpa* genes (JGI genome, v3) demonstrating the value of statistical approaches like 'Bowstrap' for community transcriptomics. 1376 significantly differentially expressed genes (6%) in response to mycorrhizal interaction were detected by ANOVA.
- *Metabolic Models*: Transcriptomics can be used to predict metabolism using PRMT [1]. 1235 predicted metabolites are present in the metabolic model, of which 55 (4%) are differentially metabolized in response to mycorrhizal interaction.
- *Proteomics*: Application of statistical modeling tool 'BowStrap' [2] yielded significant improvements in our ability to detect significant protein expression from EMSL proteomics data. A total of 5646 proteins were detected as significantly expressed. BowStrap-predicted

protein expression level was substantially different from those predicted by standard alignment method using unique reads only (r = 0.69). There were 1776 observations where BowStrap detected significant protein expressions missed using unique reads alignments, including 358 proteins uniquely detected by BowStrap, and 536 proteins (22% of proteome) detected to be differentially expressed in the membrane fraction as a function of mycorrhizal interaction.

- *Metabolomics*: There were a total of 165 metabolites detected in aspen root, 88 of which (53%) were identified as 'unknown'. Eleven metabolites (7%) were identified as significantly different by ANOVA as a function of mycorrhizal interactions.
- *Linking Omics Data*: The combination of omics data has the potential to generate a system-scale representation of biological interactions [3,4]. Integrated datasets can reveal novel observations that no single omics dataset could uncover. Predicted links between metabolic models and unknown compounds from metabolic data, informed by correlations between metabolomics and proteomic data, potentially identifies 'unknown' metabolites. Linking proteomics data with transcriptomics data identifies potential post-translational regulatory mechanisms that are undetectable by either transcriptomic or proteomic data alone.

While any single omics dataset is valuable, combining multi-omics data into models provides new insights no single omics data can offer. By linking transcriptomics, proteomics, and metabolomics in a single model, it is possible to predict specific interaction mechanism-related biological phenomenon, (e.g., biosynthesis of signaling molecules, post-transcriptional modifications of mRNA, post-translational modification of proteins, and linking metabolic models to observed metabolomics) that illuminate the molecular mechanisms of rhizosphere community interaction and seedling phenotype. Crucially, this data integration helps prioritize protein targets for biological experimental validation.

#### References

1. Peter E Larsen, Collart F, Field D, Meyer F, Keegan KP, Henry CS, McGrath J, Quinn J, Gilbert JA. 2011. Predicted Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. *Microbial Informatics and Experimentation* 2011, 1:4.

2. Larsen, P.E. and F.R. Collart, BowStrap v1.0: Assigning statistical significance to expressed genes using short-read transcriptome data. *BMC Res Notes*, 2012. 5: p. 275.

3. Peter E Larsen, Avinash Sreedasyam, Geetika Trivedi, Shalaka V Shinde, Yang Dai, Leland J Cseke, Frank R Collart. Multi-omics approach identifies molecular mechanisms of plant-fungus mycorrhizal interaction. *Frontiers in Plant Science*, 2016. 6:1061.

4. Peter E Larsen, Avinash Sreedasyam, Geetika Trivedi, Gopi K Podila, Leland J Cseke and Frank R Collart. Using Next Generation Transcriptome Sequencing to Predict an Ectomycorrhizal Metabolome. *BMC Systems Biology*, 2011. 5:70. 48.

*Funding statement.* This contribution originates from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory.

### Functional assignment of ligand specificities for Pseudomonas transport proteins

Peter Korajczyk<sup>1,\*</sup> (pkorajczyk@anl.gov), Sarah Zerbs<sup>1,\*</sup>, Peter Larsen, Shalaka Shinde, Phil Laible, Frank Collart<sup>1</sup>, and **Philippe Noirot**<sup>1</sup>

<sup>1</sup>Biosciences Division, Argonne National Laboratory, Lemont, IL.

**Project Goals**: Symbiosis between soil bacteria, mycorrhizal fungi and tree roots leads to coordinated resource exchange and enhanced productivity and resiliency in forest ecosystems. Although nutrient cycling is a key feature of these interactions, very little is known about the molecular mechanisms underpinning this process or the impact on soil community structure. The Argonne "Environment Sensing and Response" Scientific Focus Area (ESR-SFA) program aims to elucidate plant-microbial interactions between *Populus tremuloides* (Quaking aspen), the ectomycorrhizal fungus *Laccaria biocolor* and the bacterium *Pseudomonas fluorescens*, and to understand how the fungal and bacterial symbionts influence plant capture, partitioning, and allocation of carbon (C) under nutrient limitation stresses.

Recently, a machine learning approach using sequenced Pseudomonad genomes coupled with outputs of metabolic and transportomic computational models revealed that the molecular mechanisms most predictive for ecological role of *Pseudomonas* in the rhizosphere is the ability to sense and manipulate its environment via its transmembrane transport capacity (*i. e.* the transportome) [Larsen et al., 2015]. However, the high number of transporter systems in *Pseudomonas* and the lack of specific information on the transported ligands limits the accuracy of our computational approaches and hinders our ability to experimentally validate specific predictions for mechanisms of plant-microbe interactions.

To address this limitation, we functionally characterized groups of high-affinity nutrient transporters belonging to the ATP Binding Cassette (ABC) and Tripartite ATP-independent Periplasmic (TRAP) transporters. Both systems utilize solute binding proteins (SBPs) to deliver substrates to transmembrane complexes for transport into the cytoplasm. Two sets of SBPs from plant-associated *Pseudomonas* have been characterized *in vitro*: organosulfur compound-binding SBPs and SBPs for simple carbon sources such as monosaccharide and dicarboxylic acids.

A high-throughput screening procedure was used to evaluate ligand binding activity through a fluorescence assay registering increased thermal stabilization. This approach is widely used for ligand identification and can integrate multiple protein targets and large ligand screening libraries. The resulting qualitative rankings of protein-ligand interactions were subsequently validated and quantified by determining affinity constants via Isothermal Titration Calorimetry.

We report protein-ligand interactions for eight SBPs binding organosulfur compounds, including two previously unknown binding activities. These activities expand the range of small molecules recognized by the methionine ABC transporter family and provide additional insight into the transport and metabolic capabilities of Pseudomonads [Zerbs et al., *in press*]. We also report protein-ligand interactions for eight SBPs binding monosaccharides and dicarboxylic acids—comprised of six ABC-derived proteins and two TRAP-derived proteins. Interestingly, several SBPs bound multiple ligands (e.g. L-arabinose and D-galactose, galacturonic and glucuronic acids) with similar affinities in the micro- and submicromolar range.

To investigate whether these binding profiles accurately reflected ligand transport, we deleted genes encoding the SBP and cognate membrane-associated transporter from the genome. Knockout (KO) mutants were grown with various ligands as sole carbon source to determine in vivo effects of transporter removal. Six transporter systems were successfully deleted from the P. fluorescens SBW25 chromosome, using a DNA recombineering approach. Removal of transporters for C4-dicarboxylic acids and uronic acids had no observable effect on growth, suggesting that SBW25 has additional transporters for these compounds. Removal of the transporter associated with glucose-binding SBP had a moderate growth defect, confirming its physiological role in glucose transport in a context where other glucose transporters are operating. Of the two putative arabinose-galatose transporters, one preferentially transports arabinose while the other preferentially transports galactose. A double KO of both transporters showed reduced growth on both sugars relative to either single mutant, indicating that while each transporter has a preference for arabinose or galactose, it also has a secondary role in transporting the other sugar. These results support the functional assignments obtained from biochemical assays, and suggest that transport specificity is also determined by other factors downstream of the SBP. Finally, these functional assignment will allow us to test experimentally the predicted role of arabinose transport in plant growth promotion (see S. Shinde poster).

### Strain-Specific Transportomic Capacity of *Pseudomonas fluorescens* Linked to Plant Growth Promotion in Aspen Seedlings under Nutrient Stress

Shalaka Shinde<sup>1</sup> (sdesai@anl.gov), Jonathan Cumming<sup>2</sup>, Frank Collart<sup>1</sup>, Peter Larsen <sup>1,3</sup>, **Philippe** Noirot<sup>1</sup>

<sup>1</sup>Argonne National Laboratory, Argonne IL; <sup>2</sup>West Virginia University, Morgantown, West Virginia and <sup>3</sup>University of Illinois at Chicago, Chicago IL

**Project Goals:** The goal of the Environmental Sensing and Response SFA is to understand the molecular mechanisms associated with plant-microbe interaction, particularly those interactions that lead to Plant Growth Promotion (PGP) effects in marginal soils. Using a plant-bacteria co-culture, vertical plant laboratory system, we have a collection of Pseudomonad bacteria observed to possess strain- and growth condition-specific PGP effects. By linking laboratory observations with computational modeling, we identify specific molecular mechanisms by with PGP protect plant from abiotic stresses. These molecular mechanisms will provide rich opportunities for rational design of rhizosphere communities to optimize plant biomass grown in marginal lands.

Forest ecosystems are major components of the biosphere and contribute extensive ecosystem services. Trees form a significant storage sink in the global carbon (C) cycle, facilitate water fluxes in the hydrologic cycle, and provide wood and fiber for human consumption. The provision of these benefits depends upon the supply and utilization of resources (carbon dioxide, water, nutrients, and light) to and by the tree. However, photosynthesis and primary productivity are often limited by nutrient availability, which, in turn, affects the ecological roles and economic output of forests. The plant rhizosphere hosts a large and diverse community of microbes whose interactions with roots and soils influence ecosystem productivity. Diverse communities of bacteria colonize plant roots and the rhizosphere. Many of these rhizobacteria are symbionts and provide plant growth promotion (PGP) services, protecting the plant from biotic and abiotic stresses and increasing plant productivity by providing access to nutrients that would otherwise be unavailable to roots. In return, these symbiotic bacteria receive photosynthetically-derived carbon (C), in the form of sugars and organic acids, from plant root exudates. PGP activities have been characterized for a variety of forest tree species and are important in C cycling and sequestration in terrestrial ecosystems. The molecular mechanisms of these PGP activities, however, are less well known. In a previous analysis of Pseudomonas genomes, we found that the bacterial transportome, the aggregate activity of a bacteria's transmembrane transporters, was most predictive for the ecological niche of Pseudomonads in the rhizosphere.

Here, we used *Populus tremuloides* Michx. (trembling aspen) seedlings inoculated with one of three *Pseudomonas fluorescens* strains (Pf0-1, SBW25, and WH6) and one *Pseudomonas protegens* (Pf-5) as a laboratory model to further investigate the relationships between the predicted transportomic capacity of a bacterial strain and its PGP effects. Conditions of low nitrogen (N) or low phosphorus (P) availability and the corresponding replete media conditions were investigated. We measured phenotypic and biochemical parameters of *P. tremuloides* seedlings and correlated *P. fluorescens* strain-specific transportomic capacities with *P. tremuloides* seedling phenotype to predict the strain and nutrient environment-specific transporter functions that lead to PGP activities and the capacity to protect plants against nutrient stress. These predicted transportomic functions fall in three groups: (i) transport of compounds that modulate aspen seedling root architecture, (ii) transport of compounds that help to mobilize

nutrients for aspen roots, and (iii) transporters that enable bacterial acquisition of C sources from seedling root exudates. These predictions point to specific molecular mechanisms of PGP activities that can be directly tested through future, hypothesis-driven biological experiments. For example, the correlation network (Figure) links rootlet formation, an amenable phenotype, with arabinose transport. The characterization of SBW25 mutants deficient in arabinose transport (see P. Korajczyk poster) will allow us to test specifically the predicted interaction with rootlet formation.



**Phenotype and** *Pseudomonas* **transportome correlation network**. This is a graphical representation of the correlation network for aspen phenotypes and *Pseudomonas* transportome. Green rounded rectangles are aspen seedling phenotypes and diamonds are predicted ligands transported by *Pseudomonas* strains. Ligands are colored according to the specific *Pseudomonas* strain with the greatest (by PRTT-score) relative capacity for that ligand's transport: orange for Pf0-1, yellow for Pf-5, red for SBW25, and purple for WH6. Wavy green lines are strong correlations between aspen seedling phenotypic features. Solid straight lines are strong correlations between *Pseudomonas* transportomic capacity and aspen seedling phenotypes colored by culture condition: grey for replete media, blue for low N media, and yellow for low P media

*Funding statement*. This contribution originates from the "Environment Sensing and Response" Scientific Focus Area (SFA) program at Argonne National Laboratory.

### Genetic Systems to Enable Biosystems Design in Rhizosphere Pseudomonas

Marie-Francoise Gros<sup>1</sup>, Sarah Zerbs<sup>1</sup>, Grace Malato<sup>1</sup>, and **Philippe Noirot**<sup>1</sup> (pnoirot@anl.gov)

<sup>1</sup>Biosciences Division, Argonne National Laboratory, Lemont, IL.

**Project Goals**: The goal of the Environmental Sensing and Response SFA is to understand the molecular mechanisms associated with plant-microbe interactions, particularly those interactions that lead to Plant Growth Promotion (PGP). With a focus on deciphering the PGP mechanisms mediated by rhizobacteria, we are developing/adapting genome editing tools based on recombineering and CRISPR-Cas9 systems for the *Pseudomonas fluorescens* group. Our aim is to accelerate Biosystems design by streamlining the construction of multiple mutations in any *P. fluorescens* strain.

Highly sophisticated and powerful genetic approaches exist in Pseudomonads that are based on transposon mutagenesis and conjugational transfer from *E. coli*. However the design and construction of targeted mutations with these approaches require numerous steps that are hardly amenable to high-throughput. In a context where our aim is to interrogate the mechanisms of plant growth promotion in rhizobacteria by systematically testing predictions generated through data integration and modeling and to manipulate these mechanisms to modulate PGP effects, genome editing tools providing a simple and consistent way to design and construct mutations across *Pseudomonas* strains are necessary.

Recombination-mediated genetic engineering (Recombineering) relies on the homologous replacement of host genome sequences by an incoming DNA fragment. Recombineering is efficiently promoted by bacteriophage-encoded recombination functions which accommodate short homologies. It is highly amenable to high-throughput as demonstrated by the development in *E. coli* of Multiplex Automated Genome Engineering (MAGE) by the Church lab. We have developed a recombineering system based on a bacteriophage recombinase from *P. fluorescens* SBW25 genome. This system was used to construct knockouts of transporters in SBW25 (see P. Korajczyk poster) and to integrate unique combinations of fluorescens.

In the CRISPR-Cas9 system, the sequence specificity of the nuclease is programmed by a small RNA. CRISPR-Cas9 has been applied for bacterial genome engineering where it provides a strong selection to recover cells carrying the desired mutations. We are using a catalytically



inactive variant of CRISPR-Cas9 (dCas9) which binds its DNA target and prevents transcription, thus providing a tool to control gene expression. This CRISPRi system was used to downregulate various *Pseudomonas* genes involved in cell division and signaling (see Figure). Future work with CRISPRi will involve systematic down-regulation of genes and pathways for validation of functional hypotheses and model predictions.

**CRISPRi-mediated depletion of FtsZ in Pseudomonas.** Induction of dCas9 in cells results in the depletion of an essential cell division protein and leads to dramatic cell filamentation.

### A Trait Based Dynamic Energy Budget Approach to Explore Emergent Microalgal-Bacterial Dynamics and Productivity

Yiwei Cheng<sup>1</sup>\* (yiweicheng@lbl.gov), Nicholas J. Bouskill<sup>1</sup>, Jennifer Pett-Ridge<sup>2</sup>, Xavier Mayali<sup>2</sup> and Eoin L. Brodie<sup>1</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>2</sup>Lawrence Livermore National Laboratory, Livermore, CA.

Project Goals: The LLNL Biofuels SFA seeks to support robust and sustainable microalgae fuel production through a systems biology understanding of algalbacterial interactions. We hypothesize that by understanding the factors that control cellular physiology and biogeochemical fluxes in and out of algal cells, particularly through the phycosphere, we can advance the efficiency and reliability of algal biofuel production. Our research includes studies of beneficial traits of phycosphere-associated bacteria, systems biology studies of model algae, and genome-enabled metabolic modeling to predict the interspecies exchanges that promote algal growth, lipid production and healthy co-cultures. Our overall goal is to develop a comprehensive understanding of complex microbial communities needed to advance the use of biological properties for practical energy production.

URL: http://bio-sfa.llnl.gov/

Research has been underway for decades to realize the full potential of algal biofuels at the commercial scale; however, open pond algal monocultures are frequently subject to collapse due to a range of factors including microalgal grazing, pathogens and parasite invasions. Recently, it has been proposed that functionally diverse microalgal-bacterial communities can achieve higher biomass and/or lipid yields, and exhibit greater resistance to invasion relative to monocultures. Similar positive diversity-productivity relationships have been observed in a wide range of ecosystem studies, but the purposeful maintenance of a diverse microbiome is less common in managed systems.

In order to explore the relationship between microalgal functional diversity, bacterialalgal interactions, and pond productivity we have developed a trait-based dynamic energy budget model to explore emergent microalgal community structure under various environmental (e.g. light, nutrient availability) conditions. We reduced the complex algal community into functional groups (guilds). Each algal guild is characterized by a distinct combination of physiological traits (e.g. nutrient requirement, substrate affinity, growth rate) constrained by biochemical trade-offs. The trait values (i.e. relating to light, N and P harvesting) follow those from the literature. Trait values relating to N and P harvesting kinetics show allometric scaling as shown in previous studies. Trade-offs between maximum uptake rate and substrate affinity also follow known relationships.

We have explored competition between algal guilds, distinguished by light harvesting kinetic traits, using Monte Carlo simulations under constant light intensity and nutrient

influx conditions. Simulations of reconstructed guilds showed that the dominant guilds to be functionally similar to Diatoms and Haptophytes, consistent with literature observations. A second suite of Monte Carlo simulation was conducted, where N and P harvesting traits distinguished the competing algal guilds. Simulations were conducted across N and P input gradients, and under constant light. Results demonstrate that higher N and P inputs are correlated with higher productivity of the system. Across the nutrient gradients, the Diatom guild consistently exerts dominance due to their larger cell size and higher maximum uptake rates. Further simulations are underway to explore these relationships under variable environmental conditions (diurnal light and temperature profiles, nutrient pulses) and in the presence of phycosphere bacterial associates that alter nutrient flux.

This research was supported by the LLNL Biofuels Scientific Focus Area, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under FWP SCW1039. Work was performed at Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231and at LLNL under the auspices of the U.S. Department of Energy under Contract DE-AC52- 07NA27344.

## Isotope tracing and phylogenetic composition of simplified bacterial communities conferring growth and biomass enhancements to biofuel-producing microalgae

Jeffrey A. Kimbrel<sup>1\*</sup> (kimbrel1@llnl.gov), Ty J. Samo<sup>1</sup>, Jorge Ligeti<sup>1</sup>, **Peter Weber<sup>1</sup>**, Anthony Siccardi<sup>2</sup>, **Todd W. Lane<sup>3</sup>**, and Xavier Mayali<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory, Livermore CA; <sup>2</sup>Texas A&M Agrilife, Corpus Christi, TX; <sup>3</sup>Sandia National Laboratories, Livermore, CA, USA

Project Goals: The LLNL Biofuels SFA seeks to support robust and sustainable microalgae fuel production through a systems biology understanding of algal-bacterial interactions. Our research includes studies of beneficial traits of phycosphere-associated bacteria, systems biology studies of model algae, and genome-enabled metabolic modeling to predict the interspecies exchanges that promote algal growth, lipid production and healthy cocultures. Our overall goal is to develop a comprehensive understanding of the microbial and metabolic factors that control cellular physiology and biogeochemical fluxes in and out of algal cells, particularly through the phycosphere.

URL: http://bio-sfa.llnl.gov/

Mutualistic algal-bacterial interactions can arise when bacteria provide metabolically beneficial substances in exchange for energy-dense algal compounds. We have studied these positive exchanges in closely interacting algal-bacterial cultures, and simplified algal-attached bacterial communities developed via culture enrichments. We monitored the algae for elevated growth and biomass characteristics and quantified C and N exchanges with both bulk and single-cell approaches (using nanoscale secondary ion mass spectrometry (NanoSIMS)). Through successive rounds of culturing, we have established over 100 stable algal-bacterial co-cultures that exhibit increased algal productivity (as determined by chlorophyll fluorescence), and we have isolated several dozen phycosphere-attached bacteria.

We used two model biofuel-producing microalgal strains, Phaeodactylum tricornutum (Pt) and Nannochloropsis salina (Ns), to enrich for growth-promoting bacteria acquired either from the coastal Pacific Ocean or established algal raceway ponds in Texas. After enriching for phycosphere-attached bacteria, microbial communities (characterized by 16S rDNA gene sequencing) exhibited a higher abundance of some bacteria that were rare in the source communities, and lower abundance of bacteria with an exclusively "free-living" lifestyle (those abundant in the culture supernatant but below the limit of detection in the washed algal filtrate). The mechanisms leading to these community composition changes vary, and appear to depend on the source inoculum. Replicates from some sources had nearly identical emergent communities due to a consistent increase in bacteria that were rare in the source communities. However, in emergent communities from other sources, stochastic losses result in increased heterogeneity of community composition among the replicates. Inoculating a source community with either Pt or Ns had similar results, in both cases the host exerted strong selection to shape the microbial community. Compared to the original source communities, enrichments led to significant increases in the *Rhodobacterales*, most notably *Loktanella*, *Ruegeria* and *Labrenzia* genera. Additionally, Rhodobacterales, Sphingobacterales and Alteromonadales genera were

found to be significantly enriched in the phycosphere-attached over the free-living fractions.

In the next phase of this project, we established twelve co-cultures of *Pt* and single bacterial species isolated from the community enrichments. Each isolated bacterium is highly abundant in the phycosphere-attached fraction of the enriched community, suggesting they could play a role in algal population dynamics. To test the hypothesis that these isolates may shape these algal-bacterial symbioses, we quantified the exchanges of C and N between the bacteria and the algal host using nanoscale secondary ion mass spectrometry (NanoSIMS). Indeed, phycosphere attached bacteria lead to a greater incorporation of fixed algal <sup>13</sup>CO<sub>2</sub> products over unattached bacteria. While carbon transfer from algae to bacteria was confirmed, algal utilization of microbial-derived nitrogen via metabolite and/or vitamin release remains ambiguous. Algal health (quantified here as C fixation) is significantly affected by individual bacterial species, and the precise metabolites responsible for these observations are currently being examined. By combining tools such microbial community analysis with advanced isotopic imaging, we are generating a sensitive picture of how these interactions can be exploited to provide reliable and renewable fuels.

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. Additional funding was provided by the Laboratory Directed Research and Development Program at Sandia National Laboratories, which is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

## System-level analysis of metabolic trade-offs and changes during diurnal cycle of *Chlamydomonas* reinhardtii

Ali Navid<sup>1\*</sup> (navid1@llnl.gov), Marc Griesemer<sup>1</sup>, Daniela Strenkert<sup>2</sup>, Sean Gallaher<sup>2</sup>, Stefan Schmollinger<sup>2</sup>, Sabeeha Merchant<sup>2</sup>, Xavier Mayali<sup>1</sup>

<sup>1</sup>Lawrence Livermore National Laboratory, Livermore CA; <sup>2</sup>University of California, Los Angeles, CA

Project Goals: The LLNL Biofuels SFA seeks to support robust and sustainable microalgae fuel production through a systems biology understanding of algal-bacterial interactions. We hypothesize that by understanding the factors that control cellular physiology and biogeochemical fluxes in and out of algal cells, particularly through the phycosphere, we can advance the efficiency and reliability of algal biofuel production. Our research includes studies of probiotic traits of phycosphere-associated bacteria, systems biology studies of model algae, and genome-enabled metabolic modeling to predict the interspecies exchanges that promote algal growth, lipid production and healthy co-cultures. Our overall goal is to develop a comprehensive understanding of complex microbial communities needed to advance the use of biological properties for practical energy production.

### URL: http://bio-sfa.llnl.gov/

Genome-scale models (GSM) used with constraint-based optimization approaches such as Flux Balance Analysis (FBA)[1] allow system-level characterization of metabolic traits such as needs for specific nutrients or the ability to produce compounds of interest such as biofuels. FBA models are constrained by experimental measurements and fundamental physio-chemical laws and solve for a feasible metabolic flux pattern that would result in the optimum value for one biological objective (e.g., growth or ATP production). To assess metabolic changes in the microalga Chlamydomonas reinhardtii as it transitions through its daily lifecycle, we have used a comprehensive omics dataset and a curated GSM[2], along with a suite of LLNL-developed constraint-based modeling approaches. We used transcriptomic and proteomic data within our GX-FBA[3] modeling approach to identify critical active pathways as the cell changes its metabolism from energy-rich daytime phototrophic metabolism to nighttime conditions where chemotrophic metabolism dominate. The GX-FBA approach expands the utility of FBA models by using omics data such as gene-expression measurements to constrain GSM of metabolism and optimize (within constraints of mass balance) metabolic flux variations associated with changes in enzyme concentration. Thus, GX-FBA calculates metabolic changes as a system adapts to new environments. Our results indicate that significant changes associated with production of energy, biomass, redox potential and antioxidants are needed to neutralize harmful reactive oxidative species. Furthermore, by mapping the measured nutrient flux rates within the *n*-dimensional Pareto solution space generated by our high-dimensional Multi-Objective Flux Analysis (MOFA) modeling approach, we quantified the metabolic trade-offs among critical biological objectives at various stages of the C. reinhardtii lifecycle. Our results show that in addition to previously identified dominant biological objectives (growth, ATP production, optimum resource allocation)[4], during energy abundant periods, the cell diverts some energy toward other biological activities such as production of byproducts. However, our simulations show that significant production of any energy rich compound such as lipids or hydrogen gas has deleterious effects on cellular growth. Future plans include examination of environmental factors that determine the choice of which metabolic byproducts are produced and their biological roles.

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 the LLNL Biofuels SFA, FWP SCW1039 and a cooperative agreement with the Office of Biological and Environmental Research program under Award Number DE-FC02-02ER63421. Work at LLNL is performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344.

- 1. Orth, J.D., I. Thiele, and B.O. Palsson, *What is flux balance analysis?* Nat Biotechnol, 2010. **28**(3): p. 245-8.
- 2. Chang, R.L., et al., *Metabolic network reconstruction of Chlamydomonas offers insight into lightdriven algal metabolism.* Molecular systems biology, 2011. **7**(1).
- 3. Navid, A. and E. Almaas, *Genome-level transcription data of Yersinia pestis analyzed with a New metabolic constraint-based approach.* BMC Syst Biol, 2012. **6**(1): p. 150.
- 4. Schuetz, R., et al., *Multidimensional optimality of microbial metabolism.* Science, 2012. **336**(6081): p. 601-4.

## Hyper-accumulated Mn, co-localized in *Chlamydomonas reinhardtii* acidocalcisomes with Ca and P, can be mobilized in Mn-deficient situations and protects against oxidative stress

Munkhtsetseg Tsednee,<sup>*a*</sup> Madeli Castruita,<sup>*a*</sup> Brianne Elizabeth Lewis,<sup>b</sup> Ajay Sharma,<sup>c</sup> Christina E. Ramon,<sup>*d*</sup> Sean D. Gallaher,<sup>*a*</sup> Patrice Salomé,<sup>*a*</sup> Marisa Otegui<sup>*e*</sup>, Si Chen,<sup>*f*</sup> Martina Ralle,<sup>*g*</sup> Timothy L. Stemmler,<sup>*b*</sup> Brian M. Hoffman,<sup>*c*</sup> Peter K. Weber,<sup>*d*\*</sup> (weber21@llnl.gov), Jennifer Pett-Ridge,<sup>*d*</sup> and Sabeeha S. Merchant<sup>*a*</sup> (merchant@chem.ucla.edu)

<sup>a</sup>Department of Chemistry and Biochemistry and Institute for Genomics and Proteomics, University of California, Los Angeles, CA, <sup>b</sup>Department of Pharmaceutical Sciences, Wayne State University, Detroit, MI, <sup>c</sup>Department of Chemistry, Northwestern University, Evanston, IL, <sup>d</sup>Nuclear and Chemical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA, <sup>e</sup>Department of Botany, University of Wisconsin, Madison, WI, <sup>f</sup>Advanced Photon Source, Argonne National Laboratory, Argonne, IL, <sup>g</sup>Department of Molecular and Medical Genetics, Oregon Health and Sciences University, Portland, OR

Project Goals: The LLNL Biofuels SFA is developing advanced methods to support robust and sustainable microalgae fuel production through a systems biology understanding of algal-bacterial interactions. We hypothesize that by understanding the factors that control cellular physiology and biogeochemical fluxes in and out of algal cells, particularly through the phycosphere, we can advance the efficiency and reliability of algal biofuel production. Our research includes studies of probiotic traits of phycosphere-associated bacteria, systems biology studies of model algae, and genome-enabled metabolic modeling to promote bioenergy production and healthy co-cultures. Our overall goal is to develop the comprehensive understanding of complex microbial communities needed to advance the use of biological properties for practical energy production.

URL: <u>http://bio-sfa.llnl.gov/</u> http://www.chem.ucla.edu/dept/Faculty/merchant/#research

Acidocalcisomes are 100-200nm diameter electron-dense organelles, rich in calcium and polyphosphate. They have emerged as an important intracellular compartment for handling metal homeostasis in eukaryotes, both in the context of nutrition and toxicity. Using a reference organism, *Chlamydomonas reinhardtii*, for which we have genome-level information on the occurrence and expression of metal transporters and homeostasis factors, we are using state of the art imaging methodologies to visualize the function of and interactions between individual pathways for Mn, Fe, Cu and Zn utilization, and Ag, Hg, Cd sequestration. Because metals occur at parts per million and lower levels in some cells, they can be difficult to image by standard methods. Here we use high spatial resolution secondary ion mass spectrometry with the LLNL Cameca NanoSIMS 50 to image metals in embedded and sectioned cells. The NanoSIMS data are correlated with other methods to provide a more complete understanding of metal metabolism in these microalgae.

In this study we examined manganese (Mn) metabolism in microalgae. Mn is an essential element for most forms of life. In photosynthetic organisms like *Chlamydomonas reinhardtii*, a eukaryotic alga, photosystem II and Mn-superoxide dismutases are major sites of Mn utilization.  $Mn^{2+}$  is assimilated via divalent cation transporters of the NRAMP family or via phosphate

transporters as a counterion with inorganic phosphate. In contrast to Fe and Cu assimilation, Mn can be hyper-accumulated in Chlamydomonas cells in proportion to extracellular Mn supply. Hyper-accumulation is independent of phototrophic vs. heterotrophic physiology. While hyperaccumulation does require aerobic conditions, X-ray absorption near edge structure (XANES) suggests predominantly  $Mn^{2+}$  species. We used multiple imaging approaches to localize intracellular Mn, including X-ray fluorescence microscopy, transmission electron microscopy energy dispersive X-ray spectroscopy (TEM-EDS) and NanoSIMS, which indicate that the primary foci of intracellular Mn accumulation are co-localized with Ca and P, suggesting this accumulation is within acidocalcisomes, which are known to be acidic vacuoles containing polyphosphate. Mutants in components of the VTC polyphosphate polymerase which have reduced intracellular total P and dramatically lower Ca content, cannot hyper-accumulate Mn. Growth on low inorganic phosphate, but not low Ca, recapitulates this phenotype, suggesting that Mn hyper-accumulation requires an interaction with phosphate/polyphosphate. Electron nuclear double resonance (ENDOR) spectroscopy confirms association of Mn ions with inorganic phosphate and phytate, but in cells which hyper-accumulated Mn, the bulk of the Mn ions are associated with other species. Hyper-accumulated Mn is bio-available because it can be mobilized in situations of Mn-deficiency for synthesis of Mn-SOD, consistent with a role of the acidocalcisome in metal homeostasis.



Fig. 1. Correlated STEM and NanoSIMS ion images. The location of the cells can be visualized in the C and Ca ion images. The color addition image (lower left) combines the NanoSIMS C<sup>+</sup>, Ca<sup>+</sup> and Mn<sup>+</sup> images. The arrows indicate the same location in all images. Inspection shows that Mn foci tend to correlate with Ca and P foci. Count rates for P are relatively low because of low P<sup>+</sup> yield.

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. Use of the Advanced Photon Source at Argonne National Laboratory was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number: DE-AC02-06CH11357. BMH and TLS acknowledge support by the NIH (GM 111097 and DK068139).

## Exploring the functional aspects of microalgal interactions with associated microbes, predators and pathogens

Christopher S. Ward<sup>1</sup>, Haifeng Geng<sup>2</sup>, **Todd W. Lane<sup>2</sup>**, **Michael P. Thelen<sup>1</sup>**, **Xavier Mayali<sup>1</sup>**, **Rhona K. Stuart<sup>1\*</sup>** (stuart25@llnl.gov)

<sup>1</sup>Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, CA, <sup>2</sup> Department of Systems Biology, Sandia National Laboratories, Livermore, CA

Project Goals: The LLNL Biofuels SFA seeks to support robust and sustainable microalgae fuel production through a systems biology understanding of algalbacterial interactions. We hypothesize that by understanding the factors that control cellular physiology and biogeochemical fluxes in and out of algal cells, particularly through the phycosphere, we can advance the efficiency and reliability of algal biofuel production. Our research includes studies of beneficial traits of phycosphere-associated bacteria, systems biology studies of model algae, and genome-enabled metabolic modeling to predict the interspecies exchanges that promote algal growth, lipid production and healthy co-cultures. Our overall goal is to develop a comprehensive understanding of complex microbial communities needed to advance the use of biological properties for practical energy production.

### URL: http://bio-sfa.llnl.gov/

Algae biomass production is a promising renewable energy source, and mass algal cultures in open ponds are currently grown worldwide. However, costs are above target levels needed to achieve economic feasibility, in large part due to biotic interactions that lead to unpredictable biomass yield in outdoor ponds. In open ponds, algae come in contact with numerous microbial 'contaminants', chiefly heterotrophic bacteria and predatory eukaryotes. Microalgal grazers and pathogenic fungi can cause rapid declines in algal populations in engineered mass algal cultures, as well as natural aquatic systems. However, we have a limited mechanistic understanding of how these organisms interact with microalgae and what roles bacterial members of the algal microbiome play in mitigating or enhancing these interactions. Understanding the ecological interactions that arise between algal cells and other microbial members in open systems is essential for managing algal health, and, ultimately, optimizing C flow into algal biomass.

Our previous studies of algal microbiomes indicate the bacterial community associated with *Nannochloropsis salina* is relatively stable in open cultivation ponds. To investigate the biochemical and metabolic roles that the microbiome provides for its host, we performed a temporal study in *N. salina* open systems to track profiles of microbial taxa and their metabolic pathways across three multi-generational algal cohorts. Shotgun sequencing of the time series of 26 samples yielded more than 300 million high quality sequence readings, assembling to ~230 million contigs, and 5.5 million ORFs (Open Reading Frames). By combining high-resolution taxonomic analysis, metabolic reconstruction and functional diversity assessment, we found evidence of complex relationships between microbiome composition and function that govern microbial community assembly during microalgal growth.

To better understand how top-down pressures shape algal population dynamics, we have developed simplified model systems (alga-rotifer and alga-chytrid) and are studying how their interactions influence the physiological and metabolic outcomes of both members, and how potentially beneficial phycosphere bacteria may influence these interactions. Results show that chytrids can interact with algae via metabolism of algal-derived organic matter and infection (Figure 1), though differences across chytrid strains and algal growth stages suggest that variable expression and sensing of biochemical factors may be responsible for differential outcomes. Further, grazing assays suggest that presence of certain bacteria can reduce the interaction of the rotifer grazer, *B. plicatilis,* with *N. salina*. Ongoing work will focus on biochemical and multi-omic approaches to examine the mechanistic underpinnings of metabolic and trophic interactions between predators, algae and phycosphere bacteria.



**Figure 1** Infection of algal culture in red cyst stage by *Paraphysoderma* 10 days after inoculation. Red spots: Algal cells (autofluorescence of chlorophyll), yellow spots: fungal sporangia stained with Nile Red

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 at Sandia National Laboratories, under Contract DE-AC04-94AL85000. Funding was provided by Genome Sciences Program of the Office of Biological and Environmental Research under the LLNL Biofuels SFA, FWP SCW1039. Work at Sandia was also partially supported by the Laboratory Directed Research and Development Program.

### Quantifying the contribution of viruses and phage to nutrient cycling with NanoSIMS

Xavier Mayali<sup>1</sup>, Erin E. Nuccio<sup>1</sup>, Steven J. Blazewicz<sup>1</sup>, Sean D. Gates<sup>1</sup>, Richard Condit<sup>2</sup>, Nissin Moussatche<sup>2</sup>, Jennifer Pett-Ridge<sup>1</sup>, Alex J. Malkin<sup>1</sup>, Joaquin Martinez<sup>3</sup>, Peter K. Weber<sup>1\*</sup> (weber21@llnl.gov)

<sup>1</sup>Lawrence Livermore National Laboratory; <sup>2</sup>University of Florida; <sup>3</sup>Bigelow Laboratory for Ocean Sciences

Project goals: The LLNL Biofuels SFA is developing advanced methods to support biofuel research in particular, and the study of complex microbial communities in general. We are currently working on algal biofuel ponds, and we are motivated to extend our methods to microbial activity associated with biofuel crops. We are intersted in the nutrient cycles that underpin the productivity of these systems, as well as pathogens that undermine them. Our approach is to combine genomic methods with stable isotope probing and high spatial resolution secondary ion mass spectrometry imaging. This approach enables sequence identity to be linked to function in complex systems so that we can test genome-derived and other hypotheses. Here we extend our approach to the characterization of viruses.

### URL: http://bio-sfa.llnl.gov/

Viruses and phage are thought to have significant roles in nutrient cycling in virtually all ecosystems through their effects on microbial and eukaryotic populations. In the environment, most of our ecological knowledge about phages comes from aquatic systems where they are known to kill at least 40% of the standing bacterial stock and sustain up to 55% of bacterial production through liberation of C. Its also known that algal viruses significantly affect C cycling in natural algal blooms, and in soils, we hypothesize that phage are disproportionately active in the rhizosphere, where microbial activity is higher compared to the bulk soil. Viruses and phage hijack the protein and nucleic acid synthesis systems of their host cells to replicate themselves, resulting in the eventual rupture of the cells. The nutrients from the lysed cells and newly produced viral or phage particles then are presumably available to the microbial community for reuse, but this remains untested. Our goal is to develop quantitative methods for characterizing the viral and phage role in these systems using a combination of methods centered on stable isotope probing and high spatial resolution imaging mass spectrometry ("nanoSIP").

To develop this new approach, our first goal was to assess feasibility of viral particle analysis in the LLNL Cameca NanoSIMS 50, a high spatial resolution imaging mass spectrometer we use for nanoSIP. In a series of pure cultures experiments, we determined that individual viral particles yield sufficient data to not only extract single particle isotopic information, but also to characterize nucleic acid isotopic enrichment (Fig. 1). We found that for larger virions such as *Vaccinia*, analytical precision can be as good as 10% for individual virions. For phage and smaller viruses, isotopic enrichment of individual particles is detectable at approximately twice natural isotopic abundance. These results have established the analytical methods for data extraction from viral particles and the analytical bounds for virus and phage quantitation.

In a second set of experiments, we examined the potential to quantify the degradation and uptake of viral particles by consumers in its native community. We also sought to determine viral C and N partitioning among consumers. For these experiments, we used EhV, a virus that infects the

coccolithophore *Emiliania huxleyi*. We isotopically labeled EhV by producing it in <sup>13</sup>C- and <sup>15</sup>N-labeled *E. huxleyi*. Then, the <sup>13</sup>C- and <sup>15</sup>N-labeled EhV was provided to the native microbial community without the algal host. Incubations were carried out for 40 hours and 27 days to characterize the degradation and uptake of the viral particles over time. In NanoSIMS imaging, viral particles and microbes can be differentiated based on size and morphology. The time-zero sample contains a high abundance of isotopically enriched submicron particles (EhV), whereas EhV-sized particles are much less abundant in the 40-hour sample, and in their place, there is a high abundance of isotopically enriched microbes (Fig. 2). Isotopic enrichment levels in the microbes is many times higher than natural isotopic compositions, consistent with incorporation of the EhV or its degradation productions. The nanoSIP data also show that the microbes have a range of relative C and N uptake from EhV, consistent with differences in microbial need and processing capability.

While these nanoSIP results are interesting unto themselves, significant additional work remains before can achieve our goal of quantifying the role of virus and phage in nutrient cycling microbial systems. Towards that end, we are looking to include other methods, such as rapid phage isolation, enrichment, identification, visualization, and quantification, and to utilize nucleic acid sequence data to characterize and monitor phage and host dynamics in natural systems. We intend to apply these new tools in our future SFA work to gain a more mechanistic and holistic view of whether virus and phage-mediated lysis of algal, bacterial and fungal cells is a critical controller of the ultimate fate of algal/root C.



Fig. 1. Correlated AFM (A) and NanoSIMS (B - E) <sup>12</sup>C<sup>14</sup>N<sup>-</sup> images of depth-profiled vaccinia virions. The viral particles were produce with <sup>15</sup>N-labeled thymidine, which labeled their DNA. Isotopic data extracted from this time series shows the isotopic label is collocated with the DNA, where it is expected. Scale bars: 500 nm.



Fig. 2. Color addition mosaic image of nanoSIP experiment on an aquatic microbial community incubated in stable isotope ( $^{13}C \& ^{15}N$ ) labeled *E. huxleyi* viruses for 40 hours. The data show that the viral particles were degraded and incorporated by the microbes, and further show that some microbes incorporated more C or N from the viral particles. RED =  $^{13}C$  enriched; GREEN =  $^{15}N$  enriched; BLUE = unenriched organic matter. Cells that incorporate  $^{13}C$  and  $^{15}N$  tend to appear green with this color scale. The mosaic is 80 x 100 microns.

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 the LLNL Biofuels SFA, FWP SCW1039. Work at LLNL is performed under the auspices of the U.S. Department of Energy under Contract DE-AC52-07NA27344.

How Redox Fluctuation Shapes Microbial Community Structure and Mineral-Organic Matter Relationships in a Humid Tropical Forest Soil

Authors: Ashley Campbell<sup>1\*</sup>, Amrita Bhattacharyya<sup>1,2</sup>, Yang Lin<sup>3</sup>, Malak M Tfaily<sup>4</sup>, Ljiljana Pasa-Tolic<sup>4</sup>, Rosalie Chu<sup>4</sup>, Whendee Silver<sup>3</sup>, Peter Nico<sup>2</sup>, **Jennifer Pett-Ridge<sup>1</sup>** 

<sup>1</sup>Lawrence Livermore National Laboratory, Livermore CA, <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley CA, <sup>3</sup>University of California, Berkeley, CA <sup>4</sup>Pacific Northwest National Laboratory, Richland, WA

Project Goals: This Early Career research examines the genomic potential and expression of tropical soil microorganisms as they experience shifts in soil temperature, moisture, and depth and oxygen availability. By also tracking the degradation and fate of organic carbon compounds, this work will increase the accuracy of predictions about how microbial processes affect whether organic carbon is retained or lost from tropical systems. The mechanistic understanding produced by this research will directly benefit attempts to improve the predictive capacity of mathematical models that forecast future tropical soil carbon balance.

Wet tropical soils can alternate frequently between fully oxygenated and anaerobic conditions, constraining both the metabolism of tropical soil microorganisms and the mineral-organic matter relationships that regulate many aspects of soil C cycling. Tropical forests are predicted to experience a 2–5°C temperature increase and substantial differences in the amount and timing of rainfall in the coming half century. Yet we have a poor understanding of how soil microbial activity and C cycling in these systems will respond to changes in environmental variability caused by climate change. Using a 44 day redox manipulation and isotope tracing experiment with soils from the Luquillo Experimental Forest, Puerto Rico, we examined patterns of tropical soil microorganisms, metabolites and soil chemistry when soils were exposed to different redox regimes - static oxic, static anoxic, high frequency redox fluctuation (4 days oxic, 4 days anoxic). or low frequency redox fluctuation (8 days oxic, 4 days anoxic). Replicate microcosms were harvested throughout the incubation to understand how changes in redox oscillation frequency altered microbial community structure and activity, organic matter turnover and fate, and soil chemistry. While gross soil respiration was highest in static oxic soils, respiration derived from added fresh litter was highest in static anoxic soils, suggesting that decomposition of preexisting SOM was limited by O<sub>2</sub> availability in the anoxic treatment. Microbial communities responded to shifting O<sub>2</sub> availability in the different treatments. Specifically, community composition in the anoxic treatment enriched for many members in Proteobacteria and Firmicutes relative to the initial community. Additionally, we measured significant differences in DOC concentration and molecular composition (measured by FTICR-MS) corresponding to O<sub>2</sub> availability. DOC and Fe<sup>2+</sup> concentrations were positively correlated for all four redox treatments, and rapidly increased following oscillation from oxic to anoxic conditions. Prolonged anoxia led to the reductive dissolution of Fe oxides, thereby increasing DOC availability. Fe reduction coupled to OM decomposition may help to explain the rapid turnover of complex C molecules in these soils. These results, along with parallel studies of biogeochemical responses (pH, Fe speciation, P availability), suggest a highly responsive microbial and geochemical system, where the frequency of low-redox events controls exchanges of C between mineral-sorbed and aqueous

pools. Our findings highlight the need for a more explicit representation of soil redox dynamics in our understanding of C cycling in dynamic tropical forest ecosystems.

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

# Omics-Driven Analysis of *Clostridium ljundhalii* Disentangles the Complexity of Energy Conservation in Autotrophic and Heterotrophic Growth Conditions

Al-Bassam, M.M. (malbassam@ucsd.edu)<sup>1\*</sup>, Kim, J.<sup>1</sup>, Zaramela, L.<sup>1</sup>, Liu<sup>1</sup>, J.K.<sup>1</sup>and **Zengler, K**<sup>1</sup>

<sup>1</sup>University of California, San Diego, 9500 Gilman Drive 0419 La Jolla, CA 92093

Project goals: We aim to enrich the knowledge base for *C. ljungdahlii* to use it as a model for studying carbon fixation and biofuel production. We use genome-wide translational, transcriptional, transcription start sites, RNA polymerase binding sites, proteomics and other analysis to advance our understanding of energy metabolism and regulatory networks in order to improve the predictability of the ME model. Ultimately, the optimized ME-model will guide future strain design for production of valuable biocommodities and biofuels.

The availability of an elaborate metabolic model has already placed *C. ljungdahlii* as "the organism of choice" for the study of carbon fixation, acidogenesis and biofuel production. Here we carried out genome-wide ribosome profiling and RNA-seq experiments under two autotrophic (H2:CO2 and CO) and one heterotrophic (fructose) growth conditions to understand the energy requirements for growth. In all conditions studied, the correlation between ribosome profiling and RNA-seq was surprisingly high with Spearman's  $\rho$  ranging from 0.86-0.88, suggesting that gene expression is mostly regulated at the transcriptional level.

To study genes that are significantly regulated at the translational level, we filtered outlier genes that had either high or low translational efficiency (TE), such that the Spearman's correlation between ribosome profiling and RNA-seq is increased to 0.9 when they are removed from the total datasets. Among those outliers, we show that the genes encoding the RNF complex have significantly low TE in heterotrophic growth and that they are highly regulated at the translational level. Furthermore, we classify genes into 25 functional categories and based on the growth conditions, we analyzed the differential prioritization of translational and transcriptional resources per metabolic category and the subcategories within. We also report differential distribution of low and high TE genes across conditions with regard to metabolic categories.

Finally, we determined a subset of differentially translated and transcribed regulators that are likely to be involved in the regulatory switch between autotrophic and heterotrophic growth modes. First, we generated a two-way blast map of transcription factors (TFs) from all sequenced acidogenic, solventogenic and pathogenic Clostridia. Then we generated a hierarchical cluster map based on the conservation of each TF. Interestingly, TFs either differentially translated in heterotrophic or autotrophic growth are highly conserved in acetogenes and these are considered to be the growth condition switch regulators. On the other hand, TFs highly conserved across all Clostridia were not differentially expressed in *C. ljungdahlii*, suggesting they are essential for the regulation of housekeeping genes.

*Funded by DOE Grant No. DE-SC0012586 [Next-Gen3: Sequencing, Modeling and Advanced Biofuels]* 

## A New Synthetic Community System for Studying Microbial Interactions Driven by Exometabolites

John Chodkowski<sup>1\*</sup> (chodkows@msu.edu) and Ashley Shade<sup>1,2</sup>

<sup>1</sup>Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing 48824

<sup>2</sup>The DOE Great Lakes Bioenergy Center; Program in Ecology, Evolutionary Biology, and Behavior; and The Plant Resilience Institute, Michigan State University East Lansing 48824

http://ashley17061.wixsite.com/shadelab/synthetic-microbial-communities http://ashley17061.wixsite.com/shadelab/plant-associated-microbiota

**Project Goals: 1) Determine how interactions among microbial community members are underpinned by chemical interactions within their community. 2) Understand the consequences of and feedbacks on member gene regulation for microbial community interactions by linking changes in member transcripts to community exometabolite production.** 

Microorganisms exist in communities of species that can interact chemically, and these chemical interactions underlie a range of relationships from commensalism to antagonism<sup>1-4</sup> Yet there is little known about how microorganisms interact with each other in their native habitats, or how these microbe-microbe interactions scale up to impact community outcomes. Because of the specificity of many known microbe-microbe relationships, it is thought that most microorganisms produce certain chemical products only within a particular community<sup>1,5,6</sup>. It follows that investigating a microorganism in isolation offers a narrow perspective of the full spectrum of its metabolic potential. Therefore, we use a simple and advantageous synthetic microbial community system to interrogate chemical interactions among microorganisms (via exometabolites, signaling molecules and other extracellular compounds), allowing us to observe behaviors that only occur when those microorganisms exist as part of a particular consortium.

In the synthetic community system, microbial members are arrayed randomly into a 96-transwell plate with 0.22-µm filter-bottoms (per well) that physically separate each member from its neighbors but permits resource and metabolite exchange through a shared media reservoir. Community exometabolites from the media reservoir are extracted with a protocol that captures a variety of molecules, which are then analyzed using sensitive mass spectrometry. We use the system to determine how community exometabolite composition and dynamics change given particular member combinations and/or experimental treatments. In addition, we quantify member outcomes using livedead staining with flow cytometry.

We conducted two experiments to demonstrate the synthetic community system's efficacy<sup>7</sup>. The first experiment showed the system's potential for capturing antagonistic

interactions via an antibiotic, and the second experiment showed the system's potential for capturing synergistic interactions via a shared signaling molecule. Next, we demonstrated the system's potential by assessing community exometabolite changes in a three-member community over time in stationary phase<sup>7</sup>. For this demonstration, community members included typical environmental strains with relevance for plant-soilmicrobiome interactions: Pseudomonas syringae DC3000, Chromobacterium violaceum ATCC 31532, and Burkholderia thailandensis E264. We observed directional changes in community exometabolite production over stationary phase, supported by highly reproducible biological and technical replication. We also linked the production of some exometabolites to certain members or member combinations. For example, B. thailandensis was associated with many of the most dynamic and consistent mass spectral features. We found evidence for a previously undescribed antagonistic member interaction, as there consistently were reduced live cell counts of *P. syringae* when grown in the same synthetic community as *B. thailandensis*. Finally, our results suggest that microbial interactions facilitated by exometabolites do not necessarily have an outcome for members' population sizes, suggesting that nuanced microbial interactions may be overlooked if population changes are the only data considered.

Ultimately, this project seeks to integrate –omics approaches to understand dynamics of microbial interactions within their communities, with an overarching goal of understanding how these interactions translate to community function. In doing so, we will understand how specific member combinations determine collective community outcomes. We have established this synthetic community system as a reproducible laboratory model that will provide a tool to uncover interactions among members of engineered or environmental microbial communities. Future directions include collaboration with the Joint Genome Institute through a Community Science Project to link member gene expression to community exometabolite production.

### References

- 1. Little, AEF, Robinson, CJ, Peterson, SB, Raffa, KF & Handelsman, J Annu Rev Microbiol 62 2008
- 2. Damore, JA & Gore, J J Theor Biol 299 2012
- 3. Cornforth, DM & Foster, KR Nat Rev Microbiol 11 2013
- 4. Großkopf, T. & Soyer OS Curr Opin Microbiol 18 2014
- 5. Yang, YL, Xu, Y, Straight, P & Dorrestein, PC Nature Chem Biol 5 2009
- 6. Watrous, J *et al Proc Natl Acad Sci* **109** 2012
- 7. Chodkowski JL and A Shade. A synthetic community system for probing microbial interactions driven by exometabolites. *In preparation*.

This project is supported by Michigan State University, with travel support from the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and the DOE OBP Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830). Ongoing research will be supported in part by the DOE-JGI Community Science Program (Proposal #502921).

## Using a Metabolic and Gene-Expression Model to Predict and Analyze the Phenotypic Response of Acetogen *Clostridium ljungdahlii*

J. Liu<sup>1</sup>\* (jkl055@ucsd.edu), C. Lloyd<sup>1</sup>, A. Ebrahim<sup>1</sup>, J. Kim<sup>1</sup>, M.M. Al-Bassam<sup>1</sup>, C.A. Olson<sup>1</sup>, K. Zengler<sup>1</sup>

<sup>1</sup>University of California, San Diego, 9500 Gilman Drive 0419 La Jolla, CA 92093

http://www.zenglerlab.com

Project Goals: We have reconstructed a metabolic and gene-expression model (ME-model) for the acetogen *Clostridium ljungdahlii*. This model details the organism's interconnectivity of metabolism, energy conservation, and macromolecular synthesis in a computable format, which substantially enhances our knowledge about acetogens. We are now using the model to explore the potential for biocommodity production from inexpensive sources through alterations to media composition, genetic manipulation, and more.

The acetogen *Clostridium ljungdahlii* has emerged as a potential chassis for strain designed chemical production for not only can it grow heterotrophically on a diverse set of sugars, but it can also grow autotrophically on carbon monoxide (CO), carbon dioxide (CO<sub>2</sub>) and hydrogen (H<sub>2</sub>), or a mixture of all three gases (i.e. syngas). When grown autotrophically, *C. ljungdahlii* metabolizes the gases into multicarbon organics, an ability that can be redirected and engineered to produce biocommodities from low cost substrates.

To advance towards this goal, a constraint-based modelling method was used to systematize the biochemical, genetic, and genomic knowledge of *C. ljungdahlii* into a computable mathematical framework. This <u>metabolic</u> and gene <u>expression</u> model (ME-model) accounts for 944 ORFs that are responsible for the production of transcriptional units, functional RNAs (e.g., tRNAs, rRNAs), prosthetic groups, cofactors, and protein complexes that are necessary for all of the major central metabolic, amino acid, nucleotide, and lipid biosynthesis pathways. This ME-model is able to compute the molecular constitution (i.e. transcriptome, proteome, and fluxome) of *C. ljungdahlii* as a function of genetic and environmental parameters, and is able to do so accurately, as the ME-model's *in silico* transcriptome reflects *in vivo* subsystem expression under CO and fructose growth ( $r_{CO} = 0.887$ ,  $r_{fructose} = 0.906$ , p<0.001).

Not only does the ME-model recapitulate results from standard laboratory growth conditions, but it can also calculate *C. ljundahlii*'s phenotypic responses to gene knockouts, alternate carbon sources, and even changes in metal availability. For example, a simulated carbon monoxide dehydrogenase; acetyl-CoA synthesis knockout in the ME-model predicts that *C. ljungdahlii* will stop acetate production, increase ethanol production, reduce CO uptake, and reduce CO<sub>2</sub> production, similar to the findings in Liew *et al.* 2016. Additionally, the ME-model, unlike a metabolic model, can predict electron overflow resulting in ethanol secretion and, for certain carbon sources, glycerol production, which was validated through HPLC. Finally, the ME-model provides a systems biology approach to analyze unmetabolized media components in a constraint-based manner, which was validated with the effects of nickel and manganese availability on both heterotrophic and autotrophic growth rates and secretion profiles.

Thus, with this ME-model, we have a foundation for predicting and understanding the phenotype of *C. ljungdahlii* for multiple situations, which is vital for effective strain design.

Supported by DOE-DE-SC0012586, Next-Gen<sup>3</sup>: Sequencing, Modeling and Advanced Biofuels

### Title: Resistance to Stalk Pathogens for Bioenergy Sorghum

**Deanna Funnell-Harris,**<sup>1,2</sup> (<u>Deanna.Funnell-Harris@ars.usda.gov</u>), Scott Sattler,<sup>1,2</sup> and Erin Scully<sup>3,4</sup>

<sup>1</sup>United States Department of Agriculture, Agricultural Research Service (USDA-ARS), Wheat, Sorghum and Forage Research Unit, Lincoln, NE; <sup>2</sup>University of Nebraska, Lincoln; <sup>3</sup>USDA-ARS Center for Grain and Animal Health Research, Stored Product Insect and Engineering Research Unit; and <sup>4</sup>Kansas State University, Manhattan

Project Goals: This research is focused on discovering molecular and metabolic networks that drive sorghum resistance or tolerance to stalk rot pathogens. We will use advanced molecular tools to identify key factors associated with resistance and tolerance to stalk pathogens in sorghum lines optimized for key bioenergy traits (modified phenylpropanoid metabolism), with enhanced drought tolerance (nonsenescent) or with increased tolerance to stalk pathogens. Some lignin altered lines and drought tolerant lines already have demonstrated increased tolerance to these pathogens but the mechanisms that mitigate pathogenic growth have yet to be identified. Our goal is to develop sorghum lines that can withstand increased pathogen loads under reduced water conditions based on knowledge gained through this research.

Sorghum is a promising bioenergy crop with high yield potentials and significant tolerance to both drought and heat. However, sorghum is prone to stalk rots, which can significantly limit sorghum biomass production through yield reductions and lodging. Stalk rot-causing fungi normally grow endophytically within sorghum plants. When sorghum plants experience water stress, host changes often trigger a developmental switch causing the fungi to become pathogenic. The underlying plant molecular circuits that can limit or exacerbate this fungal transition from endophytic to pathogenic growth are not known and are the focus of this proposal. Several publicly available lines have previously demonstrated resistance or tolerance to sorghum stalk pathogens, including lines with postflowering drought tolerance (nonsenscence), which appears to suppress pathogenic growth. In addition, we have developed several nearisogenic sorghum brown-midrib (bmr) 6 and 12 lines with altered lignin content and composition, which were previously demonstrated to have increased resistance or tolerance to sorghum stalk rot pathogens. Lignin, a component of plant cell walls, has been a focus for development of bioenergy sorghums because its presence increases recalcitrance of biomass to cellusolic conversion, but its presence also increases total energy content of biomass, which is important for thermal conversion technologies. To increase energy content, we have engineered sorghum plants overexpressing a Myb transcription factor that induces lignin synthesis, and a gene encoding caffeoyl-CoA O-methyltransferase, a monolignol pathway enzyme. Both the transgenic and *bmr* plants accumulate phenolic intermediates from lignin synthesis that inhibit stalk pathogens in vitro. We recently have developed the ability to apply water-stress in a controlled environment, which reliably induces the developmental switch from endophytic to pathogenic growth of sorghum stalk rot fungi. Using this unique collection of plant lines described above, in combination with genomics and metabolomics tools, the following hypothesis will be directly evaluated: "based on collected metabolite and transcriptome data, we will identify metabolic networks that will lead to host resistance." To address this goal, we will

inoculate sorghum with fungi that are the most common causes of stalk disease in the U.S. (Macrophomina phaseolina that causes charcoal rot and Fusarium thapsinum that causes Fusarium stalk rot) using our newly-developed greenhouse drought bioassay under well-watered and water-deficient conditions. The specific objectives of this proposal are: 1) to determine pathogenic and endophytic growth of stalk pathogens in sorghum lines under water deficit conditions; 2) to identify host metabolites and metabolic pathways involved in resistance or tolerance to fungal stalk rot pathogens under water deficit conditions in lignin modified, nonsenescent and stalk rot resistant/tolerant lines; 3) to identify host genes from the phenylpropanoid and defense-related pathways with altered expression levels during pathogenic or endophytic growth of fungal stalk pathogens under water deficit conditions; and 4) to identify genes, gene networks, and metabolic pathways whose expression is altered in stalk rot tolerant sorghum lines under water sufficient versus deficit conditions. Our aim is to discover host metabolic and signaling pathways that enhance endophytic growth of stalk fungi and inhibit the developmental switch to pathogenic growth that frequently occurs under periods of prolonged abiotic stress. From this valuable research, we will identify biomolecular markers for resistance that will significantly enhance efforts to develop superior bioenergy sorghum with resistance to increasing disease and environmental stresses.

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2016-67009-25420.

## Identification of a U/Zn/Cu responsive global regulatory two-component system in *Caulobacter crescentus*

Dan Park, Wes Overton, Grant Bowman, and Yongqin Jiao (Jiao1@llnl.gov)

Lawrence Livermore National Laboratory, Livermore CA

Project Goals: Depleted uranium is a widespread environmental contaminant that poses a major threat to human health. In contrast to humans and animals where a trace amount of uranium can cause damage to kidneys, liver and heart, it is well known that some bacteria can tolerate high levels of uranium and influence its mobility and bioavailability in the environment. As a non-pathogenic bacterium, *Caulobacter crescentus* is an attractive bioremediation candidate due to its high tolerance to heavy metals, and its ability to mineralize uranium. Our goal is to decipher the physiological basis for U response and tolerance in *C. crescentus*, and provide insight into the effect of aerobic bacteria on U biogeochemistry and assess the utility of them in biomineralization applications.

Despite the well-known toxicity of uranium (U) to bacteria, little is known about how cells sense and respond to U. The recent finding of a U-specific stress response in Caulobacter crescentus has provided a foundation for studying the mechanisms of Uperception in bacteria. To gain insight into this process, we used a forward genetic screen to identify the regulatory components governing expression of the *urcA* promoter  $(P_{urcA})$  that is strongly induced by U. This approach unearthed a previously uncharacterized two-component system, UzcRS, which is responsible for U-dependent activation of  $P_{urcA}$ . UzcRS is also highly responsive to zinc and copper, revealing a broader specificity than previously thought. Using ChIP-seq, we found that UzcR binds extensively throughout the genome in a metal-dependent manner and recognizes a noncanonical DNA binding site. Coupling the genome-wide occupancy data with RNA-seq analysis revealed that UzcR is a global regulator of transcription, predominately activating genes encoding proteins that are localized to the cell envelope; these include metallopeptidases, multidrug resistant efflux (MDR) pumps, TonB-dependent receptors and many proteins of unknown function. Collectively, our data suggest that UzcRS couples detection of U, Zn and Cu with a novel extracytoplasmic stress response.

This study was supported by a Department of Energy Early Career Research Program award from the Office of Biological and Environmental Sciences (to Y.J.). This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-ABS -703325).

## Exploiting Natural Diversity to Identify Alleles and Mechanisms of Cold Adaptation in Switchgrass

Hari Poudel<sup>1</sup>, Joseph Dabringer<sup>1</sup>, Andrew Hummel<sup>1</sup>, Lois Bernhardson<sup>2</sup>, Millicent Sanciangco<sup>3</sup>, Nathan A. Palmer<sup>2</sup>, Gautam Sarath<sup>2</sup>, Michael Casler<sup>1</sup>, C. Robin Buell<sup>3\*</sup> (<u>buell@msu.edu</u>)

<sup>1</sup>USDA-ARS U.S. Dairy Forage Research Center, Madison, WI; <sup>2</sup>USDA-ARS Wheat, Sorghum and Forage Research Unit, Lincoln, NE & <sup>3</sup>Michigan State University, East Lansing MI

### **Project Goals**

One of the proposed mechanisms to increase switchgrass biomass is to grow lowland cultivars in more northern latitudes where they have higher yield potential, due in part, to their significantly later flowering time at these latitudes. However, lowland genotypes are not adapted to the colder winter conditions in northern climes and exhibit low survival rates when grown at latitudes outside their adapted range. Many unadapted populations possess a small frequency of individuals that can survive the winter, indicating that, within the collective genetic diversity of the population, alleles are present that confer cold tolerance. If these alleles could be catalogued and converted into molecular markers, they would facilitate accelerated breeding and provide a mechanism to improve the efficiency of breeding switchgrass cultivars with high biomass and cold hardiness. Our hypothesis is that alleles favorable to cold hardiness will be enriched in individuals that survive over-wintering conditions outside their native hardiness zone and that these alleles will be under-represented in individuals within the population that fail to survive over-wintering. Thirty-eight populations of lowland switchgrass have been planted at three locations that provide a span of temperature regimes to assess cold tolerance: DeKalb IL; Madison WI; Arlington WI. Genome-wide association analyses of survivors vs. non-survivors from Winter 2015-16 are in progress. In 2017, we will use allele frequencies determined through bulk segregant exome capture sequencing of Winter 2016-17 survivor and non-survivor pools to identify genes, transcripts, and miRNAs important to cold hardiness in these 38 lowland switchgrass populations. Data from this project will be central to furthering our understanding of cold tolerance in switchgrass and to identify alleles that that can be used in breeding programs to increase switchgrass biomass.

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2014-67009-22310.

## Title: Genetic dissection of AM symbiosis to improve the sustainability of feedstock production

### **Maria J. Harrison**\* (<u>mjh78@cornell.edu</u>) and Zhangjun Fei Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, NY 14850

### **Project Goals:**

Fertilizers comprise a significant proportion of biofeedstock production costs and large-scale fertilizer use has many negative environmental consequences. If biofeedstock production is to be economically viable and environmentally sustainable, increasing the efficiency of mineral nutrient acquisition is an important priority. One approach to this is to harness the soil microbiota, in particular, arbuscular mycorrhizal fungi, with which plants form stable mutualistic symbioses. Currently, the genetic basis of arbuscular mycorrhizal symbioses in feedstock species is largely unexplored. The overall goals of the proposed research are to determine which genes control development and functioning of AM symbiosis in feedstocks through studies of a feedstock model species, *Brachypodium distachyon*, as well as sorghum, a feedstock species.

Using *B. distachyon*, we aim to identify genes that are critical for development and functioning of AM symbiosis. This will be achieved through analysis of *B. distachyon* T-DNA insertion mutants and/or transgenic lines in which gene function has been down-regulated by RNAi or disrupted by genome editing using the CRISP/Cas9 system. *B. distachyon* T-DNA insertion mutants will be obtained from the DOE-funded insertion mutagenesis project, <u>http://brachypodium.pw.usda.gov/TDNA/</u>. Candidate genes have been selected based on previous transcriptional analyses of *B. distachyon* during AM symbiosis with three AM fungi, *Glomus versiforme, Glomus intraradices* (now *Rhizophagus irregularis*) and *Gigaspora gigantea*.

With support from the DOE Feedstock Genomics Program, Dr. Stephen Kresovich and colleagues are developing nested association mapping (NAM) populations for diverse sweet and cellulosic (bioenergy) sorghums. These NAM populations may also provide an opportunity to investigate variation in responsiveness to AM symbiosis in biofeedstock sorghum genotypes. Relative to other crops, sorghum shows high nutrient use efficiency; however, in most production environments, fertilizers are still used and therefore improving nutrient capture would be economically and environmentally beneficial. As a first step, we will evaluate the parents of the NAM populations for responses to AM symbiosis including development of symbiosis, symbiotic mineral nutrient acquisition and increases in shoot biomass. For a selection of genotypes, root and shoot transcriptional responses to symbiosis will be analyzed by RNA-seq. The profiles will provide insights into AM symbiosis and the effects of plant genotype on symbiotic gene expression.

## This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0014260.

### A Sorghum NAC Gene Affects Vascular Development and Biomass Properties

Jingnu Xia<sup>1</sup>, Yunjun Zhao<sup>2</sup>, Carrie Thurber<sup>1</sup>, Payne Burks<sup>1</sup>, Markus Pauly<sup>2,3</sup>, **Patrick J. Brown<sup>1</sup>**\* (pjb34@illinois.edu)

<sup>1</sup>Department of Crop Sciences, University of Illinois at Urbana-Champaign, IL; <sup>2</sup>Department of Plant and Microbial Biology, University of California, Berkeley, CA; <sup>3</sup>Department of Plant Cell Biology, University of Düsseldorf, Germany

Project Goals: This project, "Coordinated Genetic Improvement of Bioenergy Sorghum for Compositional and Agronomic Traits", is focused on the identification of natural variants that increase net energy production from bioenergy sorghum. Since increased digestibility is often associated with decreased agronomic performance, these traits need to be considered together. Our team includes biochemists at Berkeley performing highthroughput compositional analysis of sorghum biomass, and geneticists at Illinois running field experiments on large panels of diverse photoperiod-sensitive sorghum accessions. Detailed characterization of natural variants that modulate biomass yield and composition, such as the NAC074 variant presented here, will help guide the engineering of nextgeneration grass feedstocks.

Sorghum bicolor is a C4 grass widely cultivated for grain, forage, sugar and biomass. The sorghum Dry Stalk (D) locus controls a qualitative difference between juicy green (dd) and dry white (D-) stalks and midribs, and co-localizes with a quantitative trait locus for sugar yield. Here, we apply fine-mapping and GWAS to identify a premature stop codon in a NAC gene as the candidate polymorphism underlying the sorghum D locus. Allelic variation at D affects grain and sugar yield, biomass composition, and vascular anatomy in nearly-isogenic lines. Green midrib (*dd*) NILs show reductions in lignin and the number of vascular bundles in stalk tissue, but produce higher sugar and grain yields under well-watered field conditions. Increased yield potential in *dd* NILs is associated with increased stalk mass and moisture, higher biomass digestibility, and an extended period of grain filling. Transcriptome profiling of midrib tissue at the 4-6 leaf stages, when NILs first become phenotypically distinct, reveals that dd NILs have increased expression of a miniature zinc finger (MIF) gene. MIF genes dimerize with and suppress zinc finger homeodomain (ZF-HD) transcription factors, and a ZF-HD gene is associated with midrib color variation in a GWAS analysis across 1694 diverse sorghum inbreds. Together, these results suggest that interaction between NAC, MIF, and MIF-targeted ZF-HD transcription factors regulates vascular development and agronomic potential in cereals.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0012400

## Metabolomics as a Molecular Tool to Understand the Mechanism of Nitrogen Use Efficiency and Interactions with the Soil Microbiome in Energy Sorghum.

**Jessica E. Prenni (jprenni@colostate.edu)**<sup>1</sup>, Amy Sheflin<sup>1</sup>, Daniel P. Schachtman<sup>2</sup>, Rebecca Bart<sup>3</sup>, Thomas Brutnell<sup>3</sup>, Daniel Chitwood<sup>3</sup>, Asaph Cousins<sup>4</sup>, Jeffrey Dangl<sup>5</sup>, Ismail Dweikat<sup>2</sup>, Andrea Eveland<sup>3</sup>, Maria Harrison <sup>6</sup>, Stephen Kresovich<sup>7</sup>, Peng Liu<sup>8</sup>, <sup>3</sup>Todd Mockler, Susannah Tringe<sup>9</sup>, and Arthur Zygielbaum<sup>2</sup>

<sup>1</sup>Colorado State University; <sup>2</sup>University of Nebraska-Lincoln; <sup>3</sup>Donald Danforth Plant Science Center; <sup>4</sup>Washington State University; <sup>5</sup>UNC-Chapel Hill; <sup>6</sup>Boyce Thompson Institute; <sup>7</sup>Clemson University; <sup>8</sup>Iowa State University; <sup>9</sup>Joint Genome Institute-DOE

Project URL: <a href="http://sorghumsysbio.org/">http://sorghumsysbio.org/</a>

**The** *overall goal of the project* is to establish a foundational, systems-level understanding of plant, microbial, and environmental interactions that will lead to strategies for enhancing growth and sustainability of sorghum through genetic and microbial adaptations to water and nitrogen limited environments.

Metabolomics is a powerful analytical approach that can be utilized to explore changes at the molecular scale occurring in plants as a response to stress (e.g. nitrogen availability) or interaction with the soil microbiome. In the first year of our project, a preliminary field study was performed in which 10 sorghum genotypes were grown under conditions of both low and high nitrogen. Root samples (collected at 2 time points: July and September 2015) were harvested for metabolomics analysis. Molecular profiles were characterized using a combination of mass spectrometry platforms to increase coverage of both primary and secondary metabolites including key energy substrates in carbon and nitrogen metabolic pathways and phytohormones important in the plant stress response. Results of this analysis reflect alterations within pathways related to the metabolic storage of nitrogen in root cells, an expected result given that plants receiving high nitrogen treatment would likely be storing available reserves.

Metabolomics results also suggest incomplete flux through the Shikimate pathway in plants grown under low nitrogen conditions. Complementary investigation of the rhizosphere microbiome revealed a bloom of Pseudomonas in samples from the July harvest that was absent in samples collected in September. Integration of metabolite and microbiome data revealed interesting correlations that also point to incomplete Shikimate pathway metabolism, co-occurring with the Pseudomonas bloom. Since metabolic end products of this pathway are important in lignification and plant defense, reduced availability of these compounds could compromise plant defenses. Levels of phytohormones (in roots) with known roles in plant defense also varied in July compared to September, further supporting that reduced flux through the Shikimate pathway may be altering the plant defense response. However, it is unclear if this effect is in response to low nitrogen conditions or the interaction with Pseudomonas.

In summary, the molecular results from this pilot experiment indicate the value of metabolite analysis in understanding mechanistic responses to stress and changes in the soil microbiome. The 2016 field study was designed with increased biological replication (to improve statistical power) as well as the inclusion of a water stress environment. Analysis of these samples is underway and will enable a closer evaluation of these pathways and integration of the molecular results with both microbiome and phenotypic field data.

*This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant no DE-SC0014395*
## Systems Biology Guided by Global Isotope Metabolomics

Erica M. Forsberg<sup>1,\*</sup> (<u>forsberg@scripps.edu</u>), Tao Huan,<sup>1</sup> Erica L. Majumder,<sup>2</sup> Judy W. Wall,<sup>2</sup> Michael P. Thorgersen,<sup>3</sup> Matthew Fields,<sup>4</sup> Michael W. Adams,<sup>3</sup> and **Gary Siuzdak<sup>1</sup>** 

<sup>1</sup>The Scripps Research Institute, La Jolla, California; <sup>2</sup>University of Missouri, Columbia; <sup>3</sup>University of Georgia, Athens; Montana State University, Bozeman

https://xcmsonline.scripps.edu

### http://enigma.lbl.gov

Project Goals: The goal of our ENIGMA research is to employ global metabolomics on microbial systems to better understand how they function, what molecular species they consume/produce and identifying metabolic pathways that are effected by various stressors. Examples of stressors include metal contamination and nitrate stress. Here we apply numerous bioinformatic tools to process raw liquid chromatography mass spectrometry data obtained from cellular extracts to identify statistically significant dysregulated metabolites and the pathways they are involved in. We also utilize an autonomous approach to validate metabolite identities through data dependent tandem mass spectral acquisition. Additionally, we employ a systems biology approach to cross reference metabolic data with genomic and proteomic data to look for changes that occur systems-wide. For pathways that are difficult to identify, our goal is to elucidate them with additional cell growth experimental protocols using stable isotope labeled substrates in both stress and non-stress conditions. Global isotope metabolomics is employed to identify these pathways in an unbiased manner. We have used these methods to elucidate nitrate assimilation pathways in *Psuedomonas* strains RCH2, N2A2 and N2E2.<sup>1</sup> We are currently performing similar experiments in a dual labeled system with *Bacillus cereus* ATCC 14579 to identify altered assimilation pathways using fully labeled <sup>13</sup>C lactate and <sup>15</sup>N nitrate under metal stress, and sulfur metabolomics using *Desulfovibrio vulgaris* Hildenburough (DvH) with <sup>34</sup>S labeled sulfate.

Abstract. Bioinformatics has become essential part of analyzing large global metabolomics datasets. The XCMS Online<sup>2</sup> platform significantly decreases the time required to process raw liquid chromatography mass spectrometry (LC-MS) data for retention time alignment, feature detection and statistical analysis of dysregulated features. In a typical metabolomics workflow, the accurate mass of each feature is matched with potential candidates from a database of metabolites. Data dependent tandem mass spectra are often used to validate these identities, which can be done using an autonomous workflow<sup>3</sup> or by manual interpretation of pooled quality control samples. This is followed by analysis of the metabolic pathways they are involved in to determine how they are affecting the whole system as a whole. We have recently developed a streamlined method to easily identify these aberrant pathways directly from the raw metabolomic data using a predictive pathway analysis time. Dysregulated pathways can be further understood with respect to upstream gene and protein expression processes by correlating genomic and

proteomic data, also in an automated approach<sup>5</sup>. Resulting overlaps can be easily visualized using the newly developed Pathway Cloud Plot, where the statistical significance (*p*-value) of the perturbed pathways are plotted versus the percent overlap of the identified dysregulated metabolites in the total identified metabolites of a given pathway. Additional information about the pathway size is indicated by the radius of the bubble. This novel cloud plot allows for an easy visual interpretation of perturbed metabolic pathways in the entire system. *Desulfovibrio vulgaris* Hildenborough was subjected to both nitrate stress and exposure to mercuric chloride and preliminary results indicate alterations in nitrate assimilation and sulfate reduction processes respectfully. Further analysis would be useful to gain more insight into how these processes are changing the system.

In some instances, metabolic pathways can be obscured by multiple enzymatic reactions that utilize the same substrates and/or yield the same end products. To elucidate these pathways, global isotope metabolomics can be employed using substrates that contain stable isotopes. Labeled starting materials are used to culture microbial strains, allowing them to be metabolized into the system until it reaches a steady state. Cell cultures must be quickly pelleted by centrifugation and flash frozen with liquid nitrogen to maintain the metabolic profile of the system without applying additional stress to prevent alterations from fast-acting enzymatic and signaling processes. The global isotope metabolome is tracked throughout a biological system by performing LC-MS on metabolite extracts and by isotope pattern ratio analysis of metabolite features between the labeled and unlabeled samples and between the stressed and unstressed samples. This comparative analysis provides information on energy consumption, biosynthesis and salvage processes that are not easily identified without looking at how the stable isotope is incorporated into the system. The major advantage of this approach is that it is unbiased and therefore able to detect novel processes that could not be achieved by looking at a targeted set of metabolites.<sup>1</sup> Here we demonstrate isotope analysis on *Bacillus cereus* exposed to a metal mixture using <sup>13</sup>C-lactate and <sup>15</sup>N-nitrate and preliminary results on DvH cultured with <sup>34</sup>Ssulfate in the presence of mercuric chloride.

#### References

- 1. Kurczy, M.E. et al. Global Isotope Metabolomics Reveals Adaptive Strategies for Nitrogen Assimilation. *ACS Chemical Biology* **11**, 1677-1685 (2016).
- 2. Tautenhahn, R., Patti, G.J., Rinehart, D. & Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **84**, 5035-5039 (2012).
- 3. Benton, H.P. et al. Autonomous metabolomics for rapid metabolite identification in global profiling. *Analytical chemistry* **87**, 884-891 (2014).
- 4. Li, S.Z. et al. Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput. Biol.* 9, 11 (2013).
- Huan, T.F., E.M.; Rinehart, D.; Johnson, C.H.; Ivanisevic, J.; Benton, H.P.; Fang, M.; Aisporna, A.; Hilmers, B.; Poole, F.L.; Thorgersen, M.P.; Adams, M.W.W.; Krantz, G.; Fields, M.W.; Robbins, P.D.; Niedernhofer, L.J.; Ideker, T.; Majumder, E.L.; Wall, J.D.; Rattray, N.J.W.; Goodacre, R.; Lairson, L.; Siuzdak, G. A Cloud-Based Systems Biology Platform Guided by Metabolomics. *Nature Methods* In Review (2017).

Funding: ENIGMA, Scientific Focus Area Program at Lawrence Berkeley National Laboratory for the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under contract number DE-AC02-05CH11231, and the National Institutes of Health.

# Advancing Field Pennycress as a New Oilseed Biofuels Feedstock that does not Require New Land Commitments

John Sedbrook<sup>1</sup>\* (jcsedbr@ilstu.edu), Winthrop Phippen,<sup>2</sup> and David Marks<sup>3</sup>

<sup>1</sup>Illinois State University, Normal, Illinois; <sup>2</sup>Western Illinois University, Macomb, Illinois; <sup>3</sup>University of Minnesota, Minneapolis

http://www.wiu.edu/pennycress/ http://cbs.umn.edu/marks-lab/home https://about.illinoisstate.edu/jcsedbr/Pages/Research.aspx

Project Goals: This is a collaborative project to generate and employ genetic and germplasm resources in improving the agronomic traits of Field Pennycress (*Thlaspi arvense* L.; pennycress) for its use as a winter annual oilseed/meal/cover crop grown throughout the U.S. Midwest Corn Belt. While wild strains can produce up to 2,200 kg of seeds per hectare (840 L/ha oils e.g. for biofuels and 1,470 kg/ha press-cake e.g. for animal feeds), they suffer from inconsistent seed germination and stand establishment, sub-optimal oil and meal quality, and yield loss due to pod shatter. Our aims are to 1) generate large EMS mutant populations for internal and community use; 2) employ forward genetic screens to identify mutants having improved agronomic traits; 3) identify the genetic lesions underlying those traits, focusing on those having the highest scientific and agronomic value; and 4) employ reverse genetic tools including CRISPR-Cas9 to target mutations in genes shown to improve traits in other species.

Pennycress (*Thlaspi arvense*; Field pennycress) is an oilseed plant of the Brassicaceae family that is closely related to Arabidopsis, camelina, and rapeseed canola. Pennycress is native to Eurasia and naturalized to North America, growing widespread throughout temperate regions of the world. Pennycress can be grown as a winter annual oilseed-producing cover crop, for example, planted in the fall in standing corn and harvested in the spring in time to plant full-season soybeans throughout the 80 million-acre U.S. Midwest Corn Belt. Once commercialized, elite pennycress varieties will provide additional income to farmers and agribusinesses thereby strengthening rural communities. Pennycress will also provide ecosystem services as a cover crop, for example, reducing soil and nutrients runoff and providing habitat and pollinator support on otherwise vacant farmland.

Field trials with current isolates have demonstrated that pennycress can be seeded in upper Midwest cornfields in the late summer and fall, at which time the plants begin to grow then overwinter, producing mature seed in the spring that can be harvested without disrupting soybean planting or yields. As an energy crop adopted throughout the U.S. Midwest, pennycress varieties could annually produce 1.3 billion gallons of liquid transportation fuels and 15 million tons of high-protein seed meal, once modest breeding improvements are made (facilitated by resources and germplasm from this project). Many other products could also be produced from this oilseed feedstock. Longer-term agronomic and genetic improvements have the potential to more than double this impact.

While pennycress holds much agronomic promise, economically-viable varieties remain to be developed. Current varieties are hampered by inconsistent germination and stand establishment,

un-optimized maturity for a given growth zone, suboptimal oils quality for biodiesel and jet fuel production, high seed glucosinolate content, and significant harvest loss due to pod shatter. This project is employing modern forward and reverse genetic strategies to rapidly generate and identify lines of pennycress that harbor mutations/natural gene variants conferring superior agronomic traits. These lines are being incorporated into breeding programs located in the Midwest. We envision pennycress adoption will occur throughout the U.S. and the world.

# Some project highlights:

A) We have generated large EMS mutant populations (over 20,000 lines) and performed medium and high-throughput screens in identifying nearly 100 mutant lines exhibiting agronomically beneficial phenotypes including reduced pod shatter, early flowering/senescing, and larger flowers/pods/seeds. In one screen, seeds from over 13,000 individual M<sub>3</sub>-generation lines have been subjected to NIR spectroscopic analysis. Reanalysis of M<sub>4</sub>-generation family members confirmed possible oil content alterations. For example, M<sub>4</sub>-generation seeds from one M<sub>3</sub> plant segregated for higher oil content (37.5% vs. 33.5% co grown wild type), while seeds from others segregated for fatty acid profiles more suitable than wild type for various biofuel and industrial applications. We are performing targeted DNA sequencing of these lines as well as nextgeneration sequencing to identify causative mutations in candidate genes. The most promising mutants are being phenotypically and genotypically characterized in detail as well as crossed with breeding lines and field tested.

B) We have used the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) genome editing approach to generate and identify pennycress plants homozygous for heritable frameshift mutations in agronomically relevant genes. These pennycress mutants are being phenotypically analyzed as well as crossed with breeding lines. Many other pennycress genes are being targeted for knockout using our streamlined genome-editing pipeline so as to rapidly improve pennycress traits including reduced glucosinolate, reduced seed coat fiber, improved oil quality and quantity, and reduced pod shatter.

C) We identified a natural single base-pair deletion in the *DELAY OF GERMINATION1 (DOG1)* gene of pennycress variety Spring32. *DOG1* mutations confer reduced seed dormancy in canola and Arabidopsis without causing adverse phenotypes. We are assessing how widespread loss-of-function *DOG1* mutations are in pennycress varieties and are introgressing this natural gene variant into breeding lines.

D) Both our in-field and in-lab experiments have shown that the Elizabeth variety has superior seed germination, stand establishment, and yield compared to most other pennycress varieties we have tested in Central Illinois. Elizabeth was isolated by Terry Isbell at the USDA (Peoria, IL) as a natural variant within the Beecher strain. We have generated a segregating population of Elizabeth x Beecher crossed plants and will be scoring in Year 3 the reduced seed dormancy phenotype; bulk segregant analysis and next generation sequencing will be performed in order to identify the genetic basis for these agronomically superior traits.

This project is supported by a Plant Feedstock Genomics for Bioenergy grant funded by the U.S. Department of Agriculture National Institute of Food and Agriculture.

Microbial community proteogenomic analyses indicate extensive depth-dependent CO oxidation and C1 metabolism in soil and increased capacity for N<sub>2</sub>O reduction with increased rainfall

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

<sup>1</sup>University of California, Berkeley; <sup>2</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee; <sup>3</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>4</sup>DOE Joint Genome Institute, Walnut Creek, CA

Global climate change will alter the patterns of distribution of rainfall over the continents, leading to shifts in the rates of production and consumption of greenhouse gases in soil. Microbial metabolism in soil plays a large role in these processes, yet some organisms, C and N currencies and transformation pathways are incompletely accounted for. Previous studies quantified species and/or functional gene abundances, but lack of information linking the organisms present to their metabolic capacities precluded comprehensive identification of biogeochemical processes ongoing in soil. Here, we quantified organism abundance, metabolic potential and in situ function in grassland soil at three depths using genomes reconstructed from 60 deeply sampled soil metagenomes, linked to proteomic analyses. Included in the study were samples from a rainfall manipulation experiment designed to simulate predicted climate change. We detect highly abundant organisms with genomically encoded capacities for C and N turnover whose abundances did not vary with depth and rainfall input. Among these is a novel Eurvarchaeote that is inferred to play important roles in methane and/or ammonia oxidation. Many organisms have the capacity to oxidize CO, and the required proteins were highly represented in the proteome. CO is produced by plant roots, and generated through breakdown of heme and other porphyrins in plant and microbial biomass. The abundances of organisms predicted to produce greenhouse gases CO, from CO and NO from nitric oxide showed a statistically significant increase in abundance between 20 cm and 40 cm soil depths. Increased rainfall caused a statistically significant increase in the abundances of two Sphingobacteria capable of conversion of  $N_2O$  to  $N_3$ . One of these, represented by a near-complete genome, also has an extensive capacity for complex carbohydrate degradation. Overall, changed water inputs are predicted to affect the capacity for denitrification and N<sub>0</sub>O production and the importance of C1 relative to complex carbon metabolism in soil.

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

# An O-Glycosylated Archaeal Flagellin

Rachel Ogorzalek Loo<sup>1,2\*</sup> (<u>rloo@mednet.ucla.edu</u>), Hong Hanh Nguyen,<sup>2</sup> Nicole Poweleit,<sup>2</sup> Deborah R. Leon,<sup>3</sup> **Sabeeha Merchant**,<sup>1,2</sup> **Joseph Loo**,<sup>1,2</sup> **and Robert Gunsalus**<sup>1,2</sup>

<sup>1</sup>UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA; <sup>2</sup>University of California-Los Angeles, Los Angeles, CA, USA; <sup>3</sup>Boston University School of Medicine, Mass Spectrometry Resource, 670 Albany Street, Rm 511, Boston, MA 02118-2526 USA

Project Goals: To elucidate the biological pathways of microbes relevant to microbial biofuel production and to global carbon cycling. These studies employ proteomics and mass spectrometry to characterize protein post-translational modifications.

# http://www.doe-mbi.ucla.edu/

*Methanospirillum hungatei*, an archaeon isolated from sewage sludge, plays an important role in waste treatment and bioenergy by degrading organic wastes to methane. In association with syntrophic organisms, it facilitates chemical conversions that are essential to recycling carbon throughout the environment. Understanding these syntrophic consortia requires not only knowledge of organisms' genetic potentials, but also how the organisms' gene products vary in response to environmental changes (quantitatively and via post-translational modification). Little is known about the modifications borne by proteins within environmental microbes. Glycosylation is particularly important, impacting cell-cell adhesion and recognition, as well as cell stability. We examine flagella from *M. hungatei* and characterize its constituent protein.

In high phosphate (45 mM) medium, *M. hungatei* grows in non-motile, linear chains (up to 9 to 12 cells long) surrounded by a protein sheath within which protein plugs separate individual cells. In low phosphate medium (3-30 mM), however, the organism presents mostly as sheath-enveloped single cells or as short chains made motile by two polar tufts of flagella transecting the multilayered terminal cell plug. High resolution cryo-electron microscopy (cryoEM) structures of flagella sheared from *M. hungatei* JF1 (ATCC# 27890) suggest that the single gene product assembling into this appendage, FlaB3 (Mhun\_3140), bears several post-translational modifications.

Enzymatic digestion (trypsin, chymotrypsin, elastase, pronase, and/or pepsin) of flagellar protein and tandem mass spectrometry identified 6 *O*-linked glycosylation sites, each trailing a 508 Da disaccharide, comprising 190 (dimethyl hexose) and 318 Da units (Thr-acetamido-deoxy-hexuronic acid), with the latter sugar linked to peptide. *M. hungatei* presents the first archaeal flagellum found bearing *O*-glycosylation.

This work was supported by the Department of Energy Office of Science (BER) through DE-FC02-02ER63421 to the UCLA-DOE Institute.

## Germplasm Development for Sustainable Production of Camelina sativa Oilseed

**John K. McKay**,<sup>1, \*</sup> (jkmckay@colostate.edu), Jack Mulen,<sup>1</sup>Ed Cahoon,<sup>3</sup> Luca Comai,<sup>4</sup> and Tim Durrett<sup>4</sup>

<sup>1</sup>Colorado State University; <sup>2</sup>University of Nebraska; <sup>3</sup>UC Davis; and <sup>4</sup>Kansas State University

## **Project Goals:**

The semi-arid west is capable of making significant contributions to the production of bioenergy if appropriate feedstocks are utilized. *Camelina sativa* is poised to become an important biofuel oilseed crop for the extensive dryland farming regions of the west . We focus on development of *Camelina* as a crop adapted for growth on marginal farmland with relatively low inputs under dry conditions, as part of the DOE effort to create "regionally adapted oilseed feedstocks with enhanced yields and desirable oil qualities for biofuels". The Great Plains and intermountain West have historically lacked rotation crops that perform well and have good market value. *Camelina* is suited for rotation with wheat, and could also reduce erosion and increase amounts of soil carbon. In our current project we have: 1) created 3 mutant populations and demonstrated their utility using a subset of the lines for forward and reverse genetics, 2) identified genomic regions controlling natural variation in oil profile and abiotic adaptation 3) engineered oil profile using a combination of mutants and transformation to create genotypes optimized for a liquid fuel supply chain.

#### Abstract:

Objective 1: Development of functional genomic resources of Camelina

A TILLING population of Camelina var. Ames 1043 was developed in the Comai lab. The population DNA has been re-standardized pooled and arrayed. As a result 2,048 individual DNAs are being TILLED for loci involved in abiotic stress tolerance and lipid biosynthesis. In addition we have created 2 additional mutant populations. These two genotypes, Lindo and Licalla, are the parents of a mapping population that we have used to map drought adaptation QTL. One M2 seed per family was planted to produce M2 tissue and M3 seed. Together, these populations will ensure that we reach the target of 1000 mutant lines per accession. Excess M2 seed will be banked and used for forward genetics screens (see below). In a population of this size we expect a minimum of 90 mutations in the average 1.5kb target gene with a corresponding probability of over 97% of finding a predicted Knock Out.

Objective 2: Development of germplasm for improved drought and heat tolerance and oil

Replicated field trials were performed in Fort Collins and Greeley, Colorado, under differential irrigation treatments at each site to collect phenotypic data on a variety of traits. Sixteen new QTL were discovered from this data, along with nine QTL using data from Colorado trials of the same population in previous years. Seven QTL were discovered for yield. We are moving forward with fine mapping these QTL. We also have identified QTL affecting oil composition, of interest for objective 3 below. To identify QTL associated with heat tolerance, we have also grown this Lindo x Licalla mapping population in growth chambers under a 35°C daytime temperature regime. We have observed substantial variation in growth

under heat stress. While measurements of yield are still being collected, we have identified a QTL for biomass under heat-stress conditions.

QTL and candidate gene validation. QTLs of major importance were chosen for fine mapping. We are developing near isogenic lines (NILs) in which small genomic regions containing the QTL of interest are introgressed between genotypes. This will include fine mapping of major QTLs for oil content. NIL creation is currently in progress, with the completion of backcrossing several existing RIL lines to the parents (Lindo and Licalla) as the introgression recipient. F1 plants have been backcrossed again, and now the resulting crosses are currently growing for selfing. Recombinants will be identified by genotyping at least 384 BC1S1 progeny using SNP markers flanking the QTL interval. All of these recombinants will be selfed and the progeny genotyped, and again we will select the rare recombination events so that small homozygous Lindo introgressions are captured in a homogeneous Licalla background, and vice versa. Once QTL are fine mapped to small intervals, we will use the TILLNG population to identify knockouts of candidate genes in those intervals. These knockouts will be tested for effects on the relevant phenotype and then used for complementation studies with the parental alleles.

Mutant screens for drought and heat tolerance. To identify important loci not captured by the variation in the mapping population, we have begun screening the existing mutant population to identify lines with improved tolerance to heat and drought. In 2015 we evaluated 1008 M3 families in field plots under both well-irrigated and drought stress conditions at the CSU Agricultural Research, Development and Education Center. This identified 36 mutants with differential sensitivity to water stress.

Objective 3: Genetic improvement of Camelina oil quality

Seed from mutagenized lines was screened using electrospray ionisation mass spectrometry (ESI-MS) for alterations in the molecular species composition of neutral lipids and phospholipids. Seed from interesting lines was grown to the next generation, with three replicate plants per line. The seed from these lines has been harvested and is currently being analyzed to confirm the mutant phenotypes.

In addition, 993 mutagenized lines from the Ames 1043 TILLING population were analyzed by GC (gas chromatography) to identify fatty acid compositions that are ideal for biofuel production. We have identified 8 lines with possible FAD2 mutations, 19 with FAD3, 4 with FAE1 and 24 with FATB. The 993 mutagenized lines were also analyzed by HPLC (high performance liquid chromatography) for increased tocopherol levels. 16 lines had increased tocopherol levels compared to wild type.

This research was supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0000213141

# Expression of S-adenosylmethionine Hydrolase in Tissues Synthesizing Secondary Cell Walls Alters Specific Methylated Cell Wall Fractions and Improves Biomass Digestibility

Aymerick Eudes<sup>1,2\*</sup>(ageudes@lbl.gov), Nanxia Zhao<sup>1,2,3</sup>, Noppadon Sathitsuksanoh<sup>1,3,4</sup>, Edward E. K. Baidoo<sup>1,2</sup>, Jeemeng Lao<sup>1,2</sup>, George Wang<sup>1,2</sup>, Sasha Yogiswara<sup>1,2,3</sup>, Taek Soon Lee<sup>1,2</sup>, Seema Singh<sup>1,5</sup>, Jenny C. Mortimer<sup>1,2</sup>, Jay D. Keasling<sup>1,2,3</sup>, Blake A. Simmons<sup>1,2,5</sup>, and **Dominique Loqué**<sup>1,2,6</sup>

https//www.jbei.org/

<sup>1</sup>Joint BioEnergy Institute, EmeryStation East, 5885 Hollis St, 4<sup>th</sup> Floor, Emeryville, CA 94608, USA

<sup>2</sup>Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

<sup>3</sup>Department of Bioengineering, Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720, USA

<sup>4</sup>Department of Chemical Engineering and Conn Center for Renewable Energy, University of Louisville, Louisville, KY, USA

<sup>5</sup>Sandia National Laboratory, P.O. Box 969, Livermore, CA 94551, USA

<sup>6</sup>Université Lyon 1, INSA de Lyon, CNRS, UMR5240, Microbiologie, Adaptation et Pathogénie, 10 rue Raphaël Dubois, F-69622, Villeurbanne, France

Project Goals: The most abundant organic material on earth is lignocellulosic biomass or non-food plant material. JBEI's mission is to convert biomass to biofuels. The goal is to provide the nation with clean, renewable transportation fuels identical to gasoline, diesel and jet fuel. Building a successful lignocellulosic biofuels industry depends, in part, on developing specialized biofuel crops or feedstocks that are optimized for deconstruction into sugars and fermentation into biofuels and bioproducts.

Absttract: Plant biomass is a large source of fermentable sugars for the synthesis of bioproducts using engineered microbes. 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). In this study, we demonstrate in Arabidopsis stems that targeted expression of S-adenosylmethionine hydrolase (AdoMetase, E.C. 3.3.1.2) in secondary cell-wall synthesizing tissues reduces the AdoMet pool and impacts lignin content and composition. In particular, both NMR analysis and pyrolysis gas chromatography mass spectrometry of lignin in engineered biomass showed relative enrichment of non-methylated p-hydroxycinnamyl (H) units and a reduction of dimethylated syringyl (S) units. This indicates a lower degree of methylation compared to that in wild-type lignin. Quantification of cell wall-

bound hydroxycinnamates revealed a reduction of ferulate in AdoMetase transgenic lines. Biomass from transgenic lines, in contrast to that in control plants, exhibits an enrichment of glucose content and a reduction in the degree of hemicellulose glucuronoxylan methylation. We also show that these modifications resulted in a reduction of cell wall recalcitrance, because sugar yield generated by enzymatic biomass saccharification was greater than that of wild type plants. Considering that transgenic plants show no important diminution of biomass yields, and that heterologous expression of AdoMetase protein can be spatiotemporally optimized, this novel approach provides a valuable option for the improvement of lignocellulosic biomass feedstock.

#### Funding statement.

This work was part of the DOE Joint BioEnergy Institute (http:// www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

## Transcript switches for fine-tuning of transgene expression

Yan Liang,<sup>1,2</sup>\*(<u>vliang@lbl.gov</u>), Tania L. Gonzalez, <sup>3</sup> Sarah Richardson,<sup>1,2</sup> Veronica T Benites, <sup>1,2</sup> Clarabelle Cheng-Yue,<sup>1,2</sup> Jay Keasling,<sup>1,2,3</sup> Ming C Hammond,<sup>2,3</sup> and **Dominique Loqué**<sup>1,2,3,4</sup>

<sup>1</sup>Joint BioEnergy Institute, Emeryville, CA; <sup>2</sup>Lawrence Berkeley National Laboratory, Berkeley, CA; <sup>3</sup>University of California, Berkeley; and <sup>4</sup>Université Claude Bernard Lyon, Villeurbanne, France

# Project Goals: The project is aimed at developing novel tools to achieve stringent and precise control of transgene expression.

Since its first establishment around 30 years ago, plant transgenic techniques have been widely used for basic biological research as well as for improving agronomic traits in crop plants. To maximize the benefit and minimize the side effect of genetic manipulation of plants, stringent and precise control of transgene expression is desired. In the current report, we will present two two-component systems to support tight transgene expression: a transgene activation switch (TAS) and a transgene repression switch (TRS) were developed. The TAS is based on an alternative splicing mechanism and involved a synthetic splicing cassette and a splicing factor in pair. By employing TAS, effector proteins in effector-triggered immunity were tightly regulated and background hypersensitive response was prevented. The strategy allowed for the first time the generation of healthy transgenic Arabidopsis that initiated hypersensitive response after supply of an artificial inducer. In contrast to TAS that allows transgene to be expressed when the activator protein is present, TRS will eliminate expression of the transgene when the repressor protein is present. Using both GFP and DsRed reporters and different promoter combinations, over two order of magnitude of transgene repression were achieved with TRS as well as simultaneous repression of multiple transgenes. TRS was validated in monocotyledonous and dicotyledonous species using transient and stable expression approaches. To further increase expression flexibility, several TRS systems are in current development. TAS and TRS demonstrate the possibility of fine tuning transgene expression by layering regulatory elements. The tools are likely applicable across species boundaries and may be used as standard elements for plant synthetic biology in the future.

Y.L., S.R., V.T.B., CC, J.D.K. and D.L. were funded by the DOE Joint BioEnergy Institute (http://www.jbei.org) which is supported bythe 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. T.L.G. and M.C.H. were funded by National Institutes of Health New Innovator Award (1DP2-OD008677 to M.C.H.); Career Award at the Scientific Interface from the Burroughs Wellcome Fund (CASI1007224 to M.C.H.); UC Berkeley Chancellor's Opportunity Fellowship (to T.L.G); NIGMS Center for RNA Systems Biology at UC Berkeley (P50-GM102706).

#### Characterizing the Defense Hierarchy of Populus trichocarpa

Posy E Busby<sup>1</sup>, Shawn Brown<sup>1</sup>, Mary Ridout<sup>2</sup>, Wellington Muchero<sup>3</sup>, Brian Stanton<sup>4</sup>, David Weston<sup>3</sup>, Dale Pelletier<sup>3</sup>, Gerald Tuskan<sup>3</sup>, George Newcombe<sup>2</sup>

<sup>1</sup>Oregon State University <sup>2</sup>University of Idaho <sup>3</sup>Oak Ridge National Laboratory <sup>4</sup>GreenWood Resources

Plants host a diverse array of non-pathogenic microorganisms that can affect plant disease severity either by interacting directly with pathogens or by modulating the plant defense response. Our ongoing project seeks to characterize the contribution of fungal and bacterial endophytes, and an endophytic, eriophyid mite, to rust disease severity in the model tree, Populus trichocarpa. Fungal endophytes antagonize the leaf rust pathogen (*Melampsora*) whereas the mite competes with the rust pathogen; both interactions result in lower rust disease severity. To date, we have identified a QTL and individual genes (using GWAS) that are associated with genetic resistance to the mite. In particular, a gene involved in cutin/suberin production in response to wounding is found in resistant tree hosts. We are in the process of validating the GWAS result for mite resistance, while simultaneously using GWAS to identify genes associated with endophyte antagonism of the rust pathogen. Next, using a combination of field and greenhouse inoculation experiments, we have found that rust pathogen antagonism by fungal endophytes can be preempted by host genes for resistance that suppress pathogen development in the first place, and 2) rust pathogen antagonism by endophytes can secondarily be preempted by competitive exclusion of the rust by the mite. Overall, these results point to a *Populus* defense hierarchy with resistance genes on top, followed by pathogen competition, and finally pathogen antagonism. Our ongoing efforts seek to test the role of bacterial endophytes in the *Populus* defense hierarchy.